Features and Projections

Studies in Generative Grammar

The goal of this series is to publish those texts that are representative of recent advances in the theory of formal grammar. Too many studies do not reach the public they deserve because of the depth and detail that make them unsuitable for publication in article form.
We hope that the present series will make these studies available to a wider audience than has hitherto been possible.

Editors:
Jan Koster
Henk van Riemsdijk

Other books in this series:

Pieter Muysken and
Henk van Riemsdijk

Features and Projections

Contents

Preface

Syntactic features have played a somewhat marginal role in the develop-
ment of the theory of grammar over the past fifteen or twenty years.
Even basic questions such as 'how many are there?', 'what are they?',
'how do they distribute over syntactic structures?' were hardly addressed,
let alone answered. Nevertheless, it is clear that syntactic features do play
an important role in syntax. Few, if any, grammarians today hold, that
syntactic categories are unanalyzable atomic primitives, and any addi-
tional intrinsic properties of syntactic categories are expressed in the form
of features. It would appear to be high time, therefore, to examine the
theory of syntactic features in a more systematic way. With this goal in
mind, we organized a workshop on this topic at Tilburg University from
October 19–22, 1983.

The present collection of articles contains a number of salient contri-
butions from the workshop as well as a few articles that were not pre-
sented but nevertheless seemed relevant in this context. The book is
organized as follows. In an introductory chapter, the editors present an
overview of the state of the art in this domain. This grew out of a much
shorter document which served as a basis for the discussions during the
workshop. The next two articles present alternative feature systems for
the major syntactic categories. The contribution by Hale and Platero,
written in 1978 but never published, was thought to fit in well here,
and Reuland's talk at the workshop consisted of an ambitious revision
of the feature system which we felt had to be included. Hellan's con-
tribution, which comes next, deals with the distribution of features
within NPs and the relation of such distributions for the notion of
"head of phrase". Among recent developments in the theory of grammar,
the so-called Case Theory is one of the modules which interact, presum-
ably, in a most intricate way with systems of morpho–syntactic features.
The articles by Haegeman and Taraldsen are both detail studies on nomina-
tive case marking which illustrate this interaction. Finally, the last two
contributions address the relationship between syntactic features and
semantic categories. Silverstein discusses some fundamental issues under-
lying his conception of features as presented in his 1976 article, which
we reprint here, since it is on the one hand a 'classic' but on the other

hand appeared previously in a fairly inaccessible book. Hale, who un-
fortunately was prevented by illness from presenting his talk, discusses in
his contribution a fundamental semantic distinction which is seen to affect
a large number of apparently quite unrelated domains of the grammar
of Warlpiri.

It was a gratifying experience to be able to discuss the somewhat
reviled topic of syntactic features in a pleasant and constructive atmos-
phere. This we owe to the contributors and participants, who by their
open-mindedness and frank discussions helped create it. These partici-
pants were, in addition to the editors, Reineke Bok-Bennema, Hans
Bennis, Hans den Besten, Anneke Groos, Liliane Haegeman, Lars Hellan,
Teun Hoekstra, Richard Kayne, Mary-Louise Kean, Jan Koster, W.U.S.
van Lessen-Kloeke, Marina Nespor, Jan Odijk, Eric Reuland, Michael
Silverstein, Rik Smits, Tim Stowell, and Tarald Taraldsen. We thank
them all.

<div align="right">

Amsterdam/Tilburg
November 1984

Pieter Muysken
Henk van Riemsdijk

</div>

Projecting Features and Featuring Projections

Pieter Muysken and Henk van Riemsdijk

0. INTRODUCTION

In this introduction we will attempt to present a state of the art report on syntactic features. Our overview will take Jackendoff's (1977) sketch of a theory of phrase structure as a point of reference and focus on a number of recent developments in syntactic theory which are directly or indirectly related to the issue of syntactic features. We have organized these developments into four main sections which will address: the ontological status of syntactic features (1), the simplification and modularization of the phrase structure component (2), the emergence of new categories (3), and the distribution of features in trees (4).

The reader should not hope for a full-fledged introductory text on features in the pages to follow. On the one hand, too little is known about them for there to even be such a text. On the other hand the issue of features branches out into such a wide variety of aspects of linguistic theory that such an undertaking would require much more space. We therefore essentially limit ourselves to pointing out what the main issues and connections are as we see them, backing these pointers up with bibliographical references.

1. THE ONTOLOGICAL STATUS OF SYNTACTIC FEATURES

Why have features at all? Early versions of generative grammar worked with essentially atomic category systems, although they made fairly liberal use of additional diacritics (such as V_I vs. V_T) which might be argued to have the same formal status as features. Such diacritics were mainly used as 'distinguishers' while the category symbol as such served to express the essence of a category. Features were introduced into category theory in Chomsky's 'Remarks on Nominalizations' (1970) for two basic reasons: first the need to provide a substantive foundation for a theory of categories, and second, the need to express cross-categorial similarities among syntactic categories. Let us address these in turn.

In phonology, there is little doubt that the categories have a substan-

tive foundation. It does not come as a surprise that nasals function as a natural class for phonological rules because there is a physiological correlate of the notion of nasality. The universal system of phonological features is thought to be substantively related to the physiological properties of the speech organs. This relation is probably not a trivial one-to-one relationship, but there is little doubt that such a relationship does exist. The universal set of features is supplemented with a markedness theory (cf. Kean 1975) to provide us with a substantive theory of phonological categories.

It would be very desirable if the theory of syntactic categories could also be argued to have a substantive basis, but, unfortunately, the going is a lot rougher. The main reason why this would be desirable is that it would provide us with a clue as to how the child selects the primitives for grammar at the initial stage of the language acquisition process. As in phonology, the most plausible scenario would be one in which categories such as noun, verb, etc. can be related to some extralinguistic concept which the child may be assumed to be equipped with. In certain domains of morpho-syntax such a connection is not that hard to imagine. Take for example the localist notions involved in the oblique case system. These will undoubtedly incorporate such notions as "in" vs. "at", "near" vs. "far", "horizontal" vs. "vertical", etc. And such notions may well turn out to be fundamental in other domains of cognition, such as visual perception, as well. But can such considerations also be carried over to the system of "grammatical" categories? Chomsky, taking up essentially the Port Royal view on the issue, proposes to define the main syntactic categories in terms of the two notions 'substantive' and 'predicative'. While an obvious link to some extralinguistic concept is far from easy to establish, such a link is at least imaginable.

The idea, then, is that 'substantive' and 'predicative' are the epistemologically basic concepts in the definition of grammatical categories. And since there are four main categories, viz. N, V, A, and P and their projections, it is attractive to regard these as features:

(1)   [± substantive] (or, in Chomsky's notation, [±N])
       [± predicative] (or, in Chomsky's notation, [±V])

This yields the well-known category matrix (2)

(2)

		N	
		+	−
V	+	A	V
	−	N	P

The second motivation for the use of features rather than atomic categories concerns the existence of certain cross-categorial generalizations. As an example, consider the fact that verbs and prepositions are considered to be the canonical case assigners in the recently developed case theory (cf. Rouveret & Vergnaud 1980, Chomsky 1981). In terms of (2), this generalization can be expressed as [-N]. This type of argument is not without problems, however. Van Riemsdijk (1978) points out several shortcomings. First, the set of four categories is too small to sustain a non-trivial notion of natural class. If one tries to do so anyway, it is hard to see, for example, how {N, V} is less of a natural class than {A,V}, if we look at the facts in a variety of languages. Furthermore, it could be plausibly argued that there are natural classes comprising three out of four categories, which could not be expressed in a feature system such as (2). An obvious candidate is {N, A, P} for having a QP-type specifier system. In a more radical attack, Williams (1981a) argues that the only real generalizations across categories are those that apply to all four of them.

The prospect for arguments from natural class considerations has not improved since the emergence of the government-binding theory (Chomsky 1981). The argument from case theory mentioned above is far from unproblematic if one takes into account a wider variety of languages. It has been argued, for example, that ergative languages are characterized by the fact that verbs are *not* case assigners (cf. Bok & Groos 1984, Burzio 1981). On the other hand, adjectives are case assigners in German. More generally, whenever we have a statement to the effect that a certain subset of categories S has a property P, it turns out that class membership in S is parametrized. Take the notion of (proper) government. For each type of government a different subset is involved, and very often certain categories belong in the set for some languages, but not others. Roughly, the picture is as follows, where parentheses indicate that class membership of the category in question is parametrized.

(3)   – theta government: N, V, A, P
      – case government: P, (V), (and perhaps (A), (N)), INFL [+tense]
      – proper government: V, A, N?, (P), (INFL [+tense])

It would appear next to impossible, in other words, to base any firm conclusions about natural classes of categories on this type of consideration at the present stage. The situation becomes infinitely worse if we allow ourselves, as Jackendoff does, the use of various notational devices borrowed from phonology, such as the α-notation and angled brackets. These notational devices make it possible to state just about any kind of dependency between arbitrary subsets of categories, thereby providing us with an excessively powerful descriptive tool whose explanatory value is minimal.

These considerations about arguments from natural classes carry over *a fortiori* to the question of the choice of the features as such. Jackendoff (1977) proposes a system which differs from that given in (1) and (2). His features are based on the (im)possibility of a category to have a subject (±S) and an (NP-)direct object (±0). This yields the following system.

(4)

$$
\begin{array}{c|cc}
 & \multicolumn{2}{c}{S} \\
 & + & - \\
\hline
+ & V & P \\
- & N & A \\
\end{array}
$$

(with 0 labelling the left rows)

If cross–categorial generalizations do not provide us with a means to choose among (2) and (4), then how else can we decide? The only way, it appears, is to rely on admittedly speculative considerations relating to the epistemological status of the features. If the notions of subject and object are relational notions, as is generally taken to be the case in generative grammar (cf. Chomsky 1965), then they are not primitive since they are defined in terms of categories and structural notions such as dominance. Categories must therefore be epistemologically prior to grammatical relations. Hence, grammatical relations cannot be taken to be the basic notions involved in the definition of categories. In other words, Chomsky's arguments against taking grammatical relations as primitives of the theory (cf. e.g. Chomsky 1982) carry over to Jackendoff's feature system. This is in addition to the empirical problems of the system, described e.g. in Stowell (1981).

In the recent literature a new potential source of insight into the system of categorial features has emerged, viz. the conception of neutralization. A partial feature matrix like, say, [+V] may be interpreted in two ways. First it may be a cross–categorial statement at the level of the rule system, referring indiscriminately to either [+V, +N] or [+V, -N] in the actual structure of sentences. Second, it may also refer to a new type of category in the structure of a sentence, a category which is neither A nor V, but in some sense both. Arguments to the effect that neutralized categories of this type exist have been presented in Aoun (1981) and Van Riemsdijk (1983). Both authors argue, in fact, that the distinction between A and V may be neutralized to [+V] in some languages. If these arguments are correct, they select Chomsky's feature system (2) over Jackendoff's (4), since the latter makes it impossible to neutralize V and A. Similarly, Stowell (1981) treats English gerunds as [+N].

A somewhat different, though related, conception of neutralization is elaborated in Muysken & Lefebvre (1984). They argue that in an X–bar

projection a positively specified feature may become negatively specified – the positive specification may get lost as it were:

(5) $[-F]^{i+1}$
$|$
$[+F]^i$

The reverse is excluded, since it is always the head that is more completely specified. The evidence includes cases where nouns ([+N, -V]) and verbs ([-N, +V] may head PPs ([-N, -V]) but not vice versa, and where nominalized verbs ([+N, +V]) can head an NP ([+N, -V]), or a clause ([-N, +V]). The neutralization between A and V would be interpreted in this system as a construction with an adjectival head ([+N, +V]) which has a verbal phrasal projection ([-N, +V]). Again these considerations, if valid, provide empirical evidence against Jackendoff's feature system which we earlier rejected on the basis of epistemological considerations.

This latter conception of neutralization has a bearing on the ontological status of the feature values "+" and "-". Both for Chomsky and for Jackendoff these have the same status. The category P, which is negatively specified for both features, is no less lexical a category than A, which is [+N, +V]. The Muysken and Lefebvre proposal mentioned above implies, however, that there is a difference, in that (5) is possible but (6) is excluded:

(6) $[+F]^{i+1}$
$|$
$[-F]^i$

For another proposal to treat 'transcategorial' phenomena of this kind, see Hale & Platero (in this volume).

In a similar vein, Reuland (1983) has proposed that, in order for a category to be a proper governor (cf. (3) above), it must have at least one positively specified feature (where INFL is taken to be specified as [+N]). And Kayne's N/V contrast (cf. Kayne forthcoming)) may be taken to be the contrast between [+N] and [+V], with the negative specifications being irrelevant (cf. also Van Riemsdijk (forthcoming) for some discussion).

Finally, notice that it is interesting that no one, to our knowledge, has yet proposed a feature system based on multivalued rather than binary features, even though the "squishy" properties of categories (cf. Ross 1973) would at first sight seem to invite such a view. The tendency, instead, is to introduce additional features, if necessary. And here we venture into another murky area. Proposals to augment the two–feature

systems of Chomsky or Jackendoff in order to characterize "minor" syntactic categories such as AUX, COMP, Particle, Adverb, QP, etc. have so far been scarce and unsystematic. Correspondingly, they have had little or no follow-up. Jackendoff (1977) introduces the feature [± complement taking] ([± C]) to define the following eight categories:

(7)   [− C]   [+ C]

[+S, −0]	N	~	quantifier
[+S, +0]	V	~	auxiliary
[−S, −0]	A	~	adverb
[−S, +0]	P	~	particle

Along somewhat different lines, Reuland (in this volume) proposes an extension of the [±N, ±V] feature system, including a zero specification.

Other extensions, though at first sight less obvious ones, are found when we look at some of the attributes which syntactic categories have acquired over the years. Here are some examples:

(8)   a. $\begin{bmatrix} NP \\ +wh \end{bmatrix}$, $\begin{bmatrix} INFL \\ +Tense \end{bmatrix}$, $\begin{bmatrix} AGR \\ \alpha\ person \\ \beta\ gender \\ \gamma\ number \end{bmatrix}$, ...

   b. $A^0, N^1, V^2, ... (= A, \bar{N}, \bar{V}, ...)$

   c. $NP_i, PP^i, {}^kNP, {}_lNP_m, ...$

Let us start with (8b/c), bars and the various kinds of sub- and superscripts which have been proposed in the literature (cf. Hellan (forthcoming) for discussion of indices of this sort). Observe that formally they must be considered to be on a par with features, though at least the indices could hardly be reduced to binary ones. We will have nothing more to say about indices here.

As far as the bars are concerned, two major problems have emerged since their introduction. First, how many bars does the maximal projection of each category have? Do they all have the same number, as Jackendoff (1977) proposed? Few linguists today believe that such a uniform bar level hypothesis can be maintained. But if it is rejected, then how do we generalize over the maximal projection categories? Using a notation such as "X^{max}" amounts to glossing over the formal problem that 'max' may have a different value for each instantiation of X. The second pro-

blem has to do with how we can motivate the difference between the various non-maximal projection levels.

Taking the interpretation of bars as features at face value, Muysken (1983a) proposes to introduce two features to replace the bars: [± maximal] and [±projection]. This yields the following classification:

(9) a. X^0 $= \begin{bmatrix} X \\ -\text{proj.} \\ -\text{max.} \end{bmatrix}$

  b. X^i (0<i<max) $= \begin{bmatrix} X \\ +\text{proj.} \\ -\text{max.} \end{bmatrix}$

  c. X^{max} $= \begin{bmatrix} X \\ +\text{proj.} \\ +\text{max.} \end{bmatrix}$

This proposal embodies the empirical claim that no rule can refer to the specific number of bars of a projection, contrary to many earlier proposals about the internal structure, e.g. of NP. It has the advantage of offering a straightforward solution to one of the most serious problems for the uniform bar level hypothesis, viz. the status of such categories as particles which cannot plausibly be argued to have any projection at all, but are otherwise identical to prepositions. These fit nicely into the fourth slot provided by the feature matrix:

(10) Prt = [-N, -V, -proj., +max.]

Turning now to (8a), various kinds of morpho-syntactic features have been assumed to crop up in syntactic representations. For example, it has been argued that non-argument positions ($\overline{A}$-positions) are typically characterized in terms of such features (cf. Emonds 1976, Den Besten 1977, Van Riemsdijk 1978). We will return to this issue below in section 3. As more and more central modules of the grammar make use of such features, the need for a more general theory about morpho-syntactic features becomes more pressing. By the same token, the question of the trade-off between syntax and morphology can no longer be ignored.

In the present context we will simply point to two areas where morpho-syntactic features are thought to play a role. We return to some other aspects of these in later sections.

First, consider the work on the so-called pro-drop or null-subject parameter (cf. Rizzi 1982). Taraldsen (1978) has revived the traditional

idea that there is a correlation between the richness of the morphological expression of the distinctions in the verbal paradigm and the possibility for overt subject pronouns to be absent. There is much discussion about the correct syntactic representation of verbal inflection: the existence of an AGR(eement) node, its categorial feature make-up, the functioning of the 'rule R' which attaches AGR to the verb. What is clear, however, is that the features for person, number and gender must be involved in AGR. Given the fact that the system of (pro)noun types and their correlates in the verbal paradigm are sometimes extremely complex (think of duals, inclusive vs. exclusive, nominal classifiers, etc.), the corresponding set of morpho-syntactic features must be of considerable complexity. For some proposals as to the actual instantiation of such feature systems, see Hale (1973). The cross-classification of (pro)nominal classes in terms of features has been studied in Silverstein (1976, reprinted in this volume).

Subject-verb agreement is, of course, only one type of agreement. There are various other types to some of which we return below in section 4. But the features involved in these other types may plausibly be thought to belong to the same system.

The second area is that of case theory (cf. Rouveret & Vergnaud 1980, Chomsky 1980). What appears to matter for syntax proper is only that a lexical noun phrase have case. Particular proposals to tie syntactic principles to specific cases have so far not been successful, witness the fate of the Nominative Island Condition (NIC) of Chomsky (1980) or of the oblique case filter of Weinberg and Hornstein (1981). Nevertheless, it would seem redundant to introduce a special feature [± case marked] for syntax, just to keep out the actual case features which play a role at some other level. Furthermore, it is far from obvious that those processes which are sensitive to the specific cases, viz. case assignment and case agreement, fall outside the domain of syntax proper. From this perspective, the distinction between abstract and morphological case is a dangerous one, since it has often been used as a pretext for ignoring the question as to what the set of features is that actual case systems are rooted in. The only proper interpretation of the abstract vs. morphological distinction is that some impoverished case feature system covering at least the major grammatical cases must be taken to be present even in those languages in which case is not morphologically expressed at all.

Research on languages with overt case systems has led to some scattered proposals for case feature systems. For the grammatical cases, see Jakobson (1936), Van Riemsdijk (1983). The study of oblique case systems is, if anything, even less far advanced.

To the extent that anything systematic can be said about such feature systems, it is that the features in question are likely to be quite abstract. They can be argued to be operative in very different parts of the grammar

(cf. Hale (in this volume)) and to account for various fundamental case distinctions in widely divergent languages). With this in mind, we return to the question of the epistemological status of features. The few indications that we have appear to encourage the view that there is a finite, universal, substantively rooted list of (morpho-)syntactic features from which languages may select those features that they actually use. In this respect, then, this aspect of syntax is much like phonology, except that research on cognition in general does not even provide us with any clues (yet) as to what the features might be. For the time being, then, all proposals must be based on classificatory considerations. But the working hypothesis that the features constitute a finite, universal set imposes interesting limitations on whatever proposals are advanced.

2. THE SIMPLIFICATION AND MODULARIZATION OF THE PHRASE STRUCTURE COMPONENT

Until the late sixties, the phrase structure component was a virtually un-restricted rule system which did not even incorporate some fundamental traditional insights such as the endocentricty of the major phrasal cate-gories. Chomsky's (1970) proposals for X-bar theory and subsequent elaborations by Emonds (1976) and Jackendoff (1977) improved the situation somewhat. Let us briefly list the major properties of the X-bar system of the mid-seventies, limiting ourselves to the general picture:

(11) a. there are four major categories: N, A, V, P;
      b. a. system of categorial features cross-classifies these categories;
      c. in addition, there are a number of "minor" categories;
      d. the phrase structure rules constitute an elaborate descriptive system, but conform, in principle, to X-bar theory;
      e. X-bar theory itself involves the following primitives:
         i. left-right relations;
         ii. numerical specifications of the projection level;
         iii. 'meta-categories' such as specifiers and complements (these may not be primitive, though);
      f. there is no limit to the number of 'non-head' daughters that a node can have.

Within this general format, the phrase structure rules of a language retain a large number of stipulative elements. In particular, they must specify which categories, and how many, occur on which side of the head, and in what order. Such a situation had long been recognized as unsatisfactory. In particular it was noticed that many of these stipulations are also part

of the lexicon. Hence, there is an unwanted redundancy which gave rise
to the idea that some of these stipulations could be factored out from the
phrase structure component and accounted for in a modular fashion by
other components of the grammar (cf. Heny 1979, 1981).

The ensuing research program was taken up in Stowell's (1981) investi-
gation of the 'origins of phrase structure'. Stowell argues that the follow-
ing ingredients will carry us quite far along this road.

(i)   X-bar theory must be so restricted as to disallow stipulations
       concerning the number and order of complements, both with
       respect to the head and with respect to each other;

(ii) lexical entries uniquely specify the number of complements, and
       the theta criterion will prevent the generation of non-strictly-
       subcategorized phrases;

(iii) the order of the complements follows from a directional theory of
       government (cf. also Hoekstra 1984, Kayne 1983), where the
       direction is specified at the level of grammar and may (in the
       marked case) be different (a) for different lexical heads (e.g.
       P-NP vs. NP-V), and/or (b) for theta government vs. case govern-
       ment (cf. (Koopman 1984);

(iv) case assignment is subject to a strict adjacency requirement to
       account for the fact that NP-complements are generally closer to
       the case assigning head than PP-complements.

With regard to (iii) we should note, however, that directionality of govern-
ment, as it is interpreted in several recent papers, does more than just
specify the ordering of X' complements, particularly in an X^{max} govern-
ment theory (Aoun & Sportiche 1983). This is simply because government
then extends beyond the immediate X'-level. Hence directionality of
government is a more 'global' device than a phrase structure system:
it could determine the distribution of specifiers at the same time as that
of complements.

This type of approach has as a corollary that the attention is focused
on differences rather than similarities among categories. While Jackendoff
stressed parallelisms between categories, much recent work is dedicated
to showing how categories differ. Recall, first, the observation that in
many languages A and N do not, but that V and P do assign Case to an
adjacent noun phrase, an idea already encoded in Jackendoff's categorial
feature [± Object]. Second, the differences between NP and S, stressed
by Stowell (1981) and Aoun & Sportiche (1983). These differences are
taken by them to result from the fact that S contains two projections
(of V and INFL), and NP only one. Third, differences in directionality
of government for different categories and/or types of government (cf.

(iii) above). Fourth, the fact that not all categories (cf. P) are proper governors. Fifth, the polarizing properties of N and V referred to in Kayne (forthcoming) and Hoekstra (1984). Sixth, the suggestion in recent work on theta-assignment (e.g. Williams (1982) and Higginbotham (1983)) that nouns have no argument structure associated with them, while verbs do. Note, of course, that the assumption that the categories are different in essential ways underlies the very possibility of a theory of neutralization between categories discussed in section 1.

Further steps in the direction of the dismantling of the phrase structure component can be taken when more progress is made in the elaboration of the theories of argument structure and morphology. Work on the precise definition of the lexical entry and its relation to the head projection may lead to explanatory accounts as to which argument (e.g. agent, theme, goal) occurs in which syntactic position; cf. Williams' (1981b) work on argument structure, and Hale's (1983) proposals for lexical structure. As for morphology, Stowell (1981) proposes that 'extended word information rules' might be taken to characterize many of the configurations which were previously assumed to be generated by minor phrase structure rules.

We can construct a continuum in present theorizing that runs from (A) through (D):

(A) The lexical entry contains the specification of the cases assigned, and of the thematic roles involved, but no linking between them, and no specification of which thematic role is assigned where;

(B) The lexical entry contains a list of the thematic roles, and for one of the roles it is specified that it is assigned to the prominent or 'external' argument (Williams 1981b); schematically this position can be presented as:

$$\theta_1 \quad \overbrace{}^{\alpha} \quad \theta_2 \ldots \theta_n$$
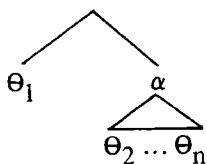
where α is taken to be X^{max} by Williams and most other authors, though not all;

(C) The lexical entry contains a syntactic representation of a number of case positions, and there is a set of association rules linking thematic roles to these positions (Hale 1983):
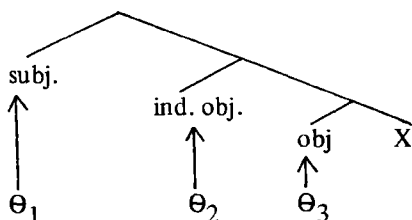
(D) The lexical entry is somewhat like the one in option (C), and in
     addition is subject to all kinds of lexical operations (Bresnan 1982).

Many of these proposals remain tentative, of course, and some, such
as the extended word formation hypothesis (D), do not really reduce the
amount of stipulation in the grammar but merely shift the stipulation
from one component to another. However, such shifts have on occasion
turned out to be fruitful in the long run. On the other hand, there are
quite a few remaining elements in the realm of phrase structure which
have so far largely defied a modular treatment along similar lines. We will
mention two of these.

   First, the subject. The obligatory presence of the subject does not
follow from any of the modules discussed so far (and neither, we might
add, does the obligatory presence of INFL, unless it is taken to be the
head of S). In order to keep the X–bar component clean, Chomsky (1981)
proposes to shift the stipulation from the phrase structure component to
the projection principle. The 'extended projection principle' now simply
states that 'you have to have a subject and it has to be there at all levels
of representation'. For proposals to derive this extension of the projection
principle, see Groos (1982) and Borer (1984).

   Second, Stowell's story has little or nothing to say about the distribu-
tion and categorial make–up of the minor syntactic positions, in particular
$\bar{\text{A}}$-positions such as COMP, INFL, CLITIC, etc. In some cases, it is not
too hard to find a line of reasoning which will at least in part predict
their positions. The fact that COMP must hang from $\bar{\text{S}}$ might conceivably
follow from the theory of scope. But such a theory, if at all based on
c–command, cannot predict whether COMP precedes or follows S, except
if one assumes directionality of proper government. The position of INFL
might follow from the conjunction of (a) the obligatoriness of the subject
NP, and (b) the directionality of nominative case assignment, and (c)
the adjacency condition on case assignment. But these assumptions would
exclude structures of the type $[\text{NP VP INFL}]_S$, which may very well exist.
Note also that any 'directional' theory has to make opposite stipulations
for COMP (rightward) and INFL (leftward), on standard assumptions.
Finally, the position of CLITIC may be partly predicted on the basis of

the role that the CLITIC node(s) play in the assignment (and absorption) of case.

At this point, however, it remains quite possible that there will be a residue of phrase structural characterizations of $\bar{A}$-positions both regarding their position and regarding their characterization in terms of morphosyntactic features (cf. section 3). If there is such a residue it may remain desirable to restrict the use of morphosyntactic features in phrase structure by making this use subject to a theory of markedness as proposed in Van Riemsdijk (1978).

3. THE STATUS OF OTHER CATEGORIES

As may have become clear in the course of our overview so far, most of X-bar theory and the features incorporated in it center around the four basic categories N, V, A, P and their projections. Correspondingly there are a number of categories which do not fit in any obvious way into the system. Most intriguing among these are INFL and COMP.

These categories as such are not new, of course. COMP was first introduced systematically in Bresnan's (1970, 1972) work, while INFL is essentially just a new name for (some aspects of) the AUX node introduced in Chomsky (1957). What makes them important is that in recent years several proposals have appeared to the effect that INFL and/or COMP are heads in some or all languages. We will not review these proposals in detail but focus instead on some of the major considerations that may play a role in reaching firmer conclusions about this issue than has hitherto been possible.

(i) Both INFL and COMP may be said to characterize the essential nature of the phrase, i.e. the clause, that they occur in. COMP determines what type of complement a clause is, declaration, interrogative, etc. And to the extent that other clause introducers like *while, because, before* are also complementizers like *that* and *for*, COMP also determines a wide variety of adverbial clause types. Similarly, INFL determines the mood of a clause. For the distinction between [+ tense] and [- tense] this is usually not very spectacular, but as soon as we consider languages with a richer mood system, including for example subjunctive, optative, conditionalis, dubitative and what have you, the semantic import of INFL becomes apparent. In English, where INFL is expanded as it were by the auxiliary system, these mood functions are largely taken over by the modal verbs. In addition, the choice of mood may crucially affect the properties of the containing phrase for other modules of the grammar. A declarative clause may be a governing category for the binding theory while a subjunctive one need not be (cf. Anderson (1982) on Icelandic; Yang 1983). All these

properties are head-like properties, but it is important to exercise proper caution, since by similar reasoning e.g. the determiner position of NP may be argued to be head-like as well (cf. Hellan (in this volume) for more discussion).

(ii) A related property is the obligatoriness of INFL and COMP. In most current analyses at least every true sentence has both a COMP and an INFL, though both may on occasion be empty. Furthermore some features of INFL and/or COMP may be so fundamental that they serve to characterize other categories as well. Stowell (1981), for example, proposes to incorporate the feature [± tense] into the category system in the following way:

(12)

	N	V	tense
NP	+	−	−
VP	−	+	−
S	+	−	+
AP	+	+	
PP	−	−	

We will return to the issue of the feature composition of INFL and COMP in (v) below.

(iii) While there is no intrinsic connection between the notions fundamental and head-like, it remains a suggestive connection. In light of this it seems fair to say that recent proposals to the effect that AUX might be a universal category of natural language have also contributed to the popularity of the view that INFL, the nucleus of AUX, is a head. Part of the inspiration for ideas along these lines derives from the work of Ken Hale, Susan Steele and others on auxiliary systems in a wide variety of languages (cf. Steele 1981).

(iv) Assuming that a phrasal category cannot have more than one head, it cannot be the case that both V and INFL are the head of S. In other words, if Jackendoff's assumption that V is the head of S is maintained, then INFL cannot be the head of S. Conversely, if INFL is the head of S, then VP must be a maximal projection. Strangely enough, the latter issue, which in many ways would appear easy to solve, is still far from settled. The standard phenomena of VP-deletion, VP-preposing, etc. have not yielded any new insights, and more theoretical considerations haven't either. To cite just one example, weak crossover phenomena certainly force the presence of a VP node. But this node must be maximal only under some definitions of c-command and government, but not under others. It is not unusual, in recent work, for such uncertainties

to be resolved by not resolving them. The strategy is to say that both options are available in natural language and that the choice between them represents a very fundamental parameter of universal grammar. This strategy has also been adopted in the present case – and admittedly not without success – by Koopman (1984), Taraldsen (1983), and others.

By similar reasoning, INFL and COMP cannot both be the head of $\overline{S}$. Thus, if COMP is a head, then S is a maximal projection. Again, evidence is scarce, and again, there may be some parametric variation involved. Particularly suggestive in this connection is the fact that COMP appears to assume (some of the) functions of INFL in some languages. COMP, for example, may sometimes assign nominative case to the subject, according to some proposals (e.g. Koopman 1984). Implementing the idea, Platzack (1983) has introduced the term CONFL, which more generally would be assumed to be the basic category and which might materialize as a discontinuous category in some (or many?) languages. This is an attractive view given the many other close connections between COMP and INFL. They must agree (*that* – [+ tense] vs. *for* – [- tense] etc.), INFL may move to COMP, V may move either to INFL or to COMP, and so on. Attractive though such an approach may seem, however, it is important to keep in mind that, like those mentioned above, it dramatically increases the number of options available for the grammars of natural languages.

(v) If categories like INFL and COMP are headlike, then that ought to be reflected in their feature make-up. Note, for example, that INFL is taken to consist of two parts, both pertaining to aspects of verbal inflection:

(13)   INFL

        [± tense]   AGR
                               [α person]
                               [β number]
   [γ gender]
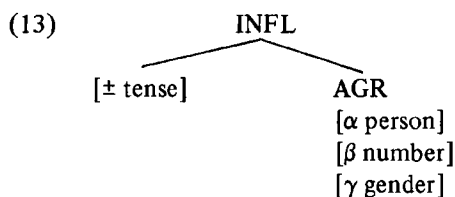
Concentrating on AGR, note that these features are characteristic both of verbal inflection and of the (pro-)nominal system. Accordingly, INFL might be thought to be either nominal or verbal in nature. As expected, avid use has been made in the literature of these new options for parametric variation. For example, it has been suggested that the empty subject of pro-drop languages is licensed by a nominal INFL which serves as a proper governor, while in non-pro-drop languages INFL is taken to be non-nominal (cf. Rizzi 1982).

Similarly, it has been argued by Aoun (1981) that INFL can be characterized by the same features that cross-classify the nominal system,

viz. [± anaphoric] and [± pronominal] (cf. also Finer 1983). On the other hand, it is not obvious that such properties are characteristic of heads rather than phrases. Furthermore it has been argued, albeit back in the pre–government–and binding days, that Ā-positions such as COMP and perhaps INFL are typically characterized by morpho–syntactic features which are taken to constitute a subset of 'minor' syntactic features clearly distinct from the major categorial features. See Van Riemsdijk (1978), den Besten (1977), as well as Bok–Bennema (1981) for a proposal to distinguish M–binding (morpho–syntactic binding) from argument–binding. The issue as such remains fully relevant. If it is true, say, that only *wh*-phrases may move into COMP, then this fact must be expressed somewhere in the grammar. Given that *wh*-movement is subsumed under 'move α', it must be stated elsewhere. Two options come to mind. Either we formulate a filter, or we use morpho–syntactic features such as [± wh] to characterize Ā-positions in the phrase structure component. The latter option may appear to be unattractive in view of the dismantling of the phrase structure component discussed in section 2 above. But a filter is not particularly attractive either, and furthermore, as noted above, the dismantling process has only marginally affected the distribution and categorial make–up of Ā-positions. In addition, such an approach would make it possible to maintain the markedness theory for morpho–syntactic phrase structure positions developed in Van Riemsdijk (1978).

   The dilemma may well be a spurious one, however. We already noted above that INFL has internal structure, cf. (13). Similarly, COMP is generally taken to be branching:

(14)   COMP

        landing site   position of the
        for *wh*-phrases   lexical complementizer
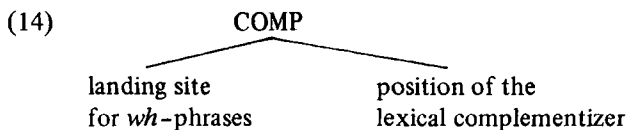
While the above considerations may apply to the landing site in COMP, they do not affect the position of the lexical complementizer. In fact, it has long been noted that there are close connections between complementizers and prepositions. It can hardly be an accident that such words as *for, because, before* can be classified both as complementizers and as prepositions. Pursuing this idea, Emonds (forthcoming) argues that COMP = P and, consequently, that $\bar{S}$ = PP.

   (vi) In terms of the theory of government, heads are assumed to be accessible to government by external governors (cf. Belletti & Rizzi 1981). In this light, the fact that matrix verbs appear to subcategorize for the complementizer of the complement clause (*that - for -*) and for [± tense] suggests that we treat COMP and INFL as heads. In addition,

it has been proposed that *wh*-phrases in COMP are sometimes accessible to case marking from the outside (cf. Kayne 1980, Groos & Van Riemsdijk 1981).

Conversely, if it can be argued that COMP and/or INFL act as governors, we have another indication that they may have head status. Both COMP and INFL have been argued to be case governors for nominative case. Also, COMP appears to act as a proper governor under certain circumstances, depending on one's assumptions about index percolation and the like (cf. Pesetsky 1981, Bennis 1980), and so does INFL, at least in pro-drop languages. More considerations arise under the assumption that government is directional (cf. Kayne 1983, Hoekstra 1984, Koopman 1984). But again, too much remains uncertain. For example, if COMP governs S and if INFL governs VP, then their distribution in English would follow from the consistent head-initial nature of that language. But then INFL would in addition have to govern leftward in order to assign nominative case to the subject NP.

(vii) In Chomsky (1973) and in subsequent work (cf. in particular Van Riemsdijk 1978 and Koster 1978a/b), COMP was taken to be a peripheral position which serves as an 'escape hatch' for movement. Such a view would appear to be difficult to reconcile with the idea that COMP is a head. But to the extent that the (im)possibility of long movement is determined by the bounding theory, it does not really matter whether COMP is a head or not, as long as the right choice of bounding nodes is made. The latter consideration might, however, play a certain role if the set of bounding nodes were to be a subset of the set of maximal projection nodes. Under that assumption, S would have to be maximal projection in English, being headed either by INFL or V, but not in Italian.

A related consideration has to do with Ross's (1967) left branch condition (LBC). It is curious that no one adopting successive cyclic *wh*-movement has been bothered by the fact that movement out of COMP constitutes an LBC-violation. Of course, there is already an impressive list of well-established LBC-violations including *combien*-e-extraction in French (cf. Obenaur 1976), *r*-movement in Dutch (cf. Van Riemsdijk 1978), *was für/wat voor* extraction in German and Dutch (cf. Den Besten 1981, 1982). One possible line one could take on these would be to say that extraction from a left branch is possible just in case the ensuing empty category is licensed by proper government from the outside. Under the aforementioned proposal by Belletti & Rizzi, this would imply that such empty categories are in a head position. Alternatively, one might return to the concept of periphery and extend the Belletti & Rizzi approach by assuming that both head and periphery of a phrase are accessible to outside government, as opposed to non-peripheral non-heads, where the latter notion would seem to correspond roughly to the positions

which are internally governed. This would imply a return to the notion of minimal government i.e. government which in turn is similar to the head constraint of Van Riemsdijk (1978). Either approach would have extensive repercussions, however.

It should be clear from the above discussion that very little can be said with any certainty at this point. Furthermore, complicated theoretical considerations should not make us blind to the more down-to-earth facts. For if INFL and/or COMP are heads, they certainly are pretty strange specimens of the species. In particular they would be branching (cf. (13) and (14)) and they would not in any obvious way be lexical. Furthermore, they would head projections which are quite dissimilar from those headed by N, A or P. They would also not be able to occur without complements (VP and S respectively). So the issue remains completely open.
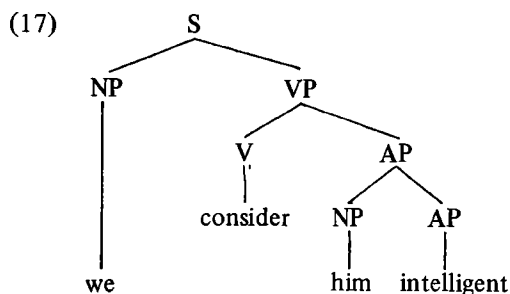
In conclusion, after this long discussion of INFL and COMP, let us turn to a second major new issue in the realm of category theory, viz. small clauses. Since the issue does not have any obvious implications for syntactic features, we will limit ourselves to a few brief remarks. It is well known that certain NPs may simultaneously be subject-like and object-like. Consider *consider*, for example:

(15)   We consider him intelligent

By position and case, *him* looks like the direct object of *consider*. Semantically, however, *him* is the subject of *intelligent*. There are two major ways to approach situations of this type. Either the NP in question is syntactically an object but semantically a subject, or it is a subject throughout. The former approach is the predication analysis of Williams (1980), which expresses the semantic relation between the NP and the predicate by coindexation, as in (16).

(16)
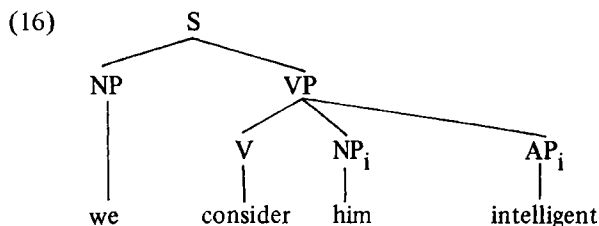
The second approach, due to Stowell (1983), assumes that all major phrases can have a subject NP, even such categories as AP and PP. On this view, (15) would be rendered as (17).

(17)

```
                    S
              ╱─────────
           NP              VP
            │            ╱────────
            │           V        AP
            │           │'      ╱────
            │        consider NP    AP
            │                  │      │
           we                him  intelligent
```

The subject of the AP is taken to receive objective case from the matrix verb, just as in the case of exceptional case marking. Stowell's proposal has far-reaching consequences for the $\overline{X}$-theory, of course, which is why we mention the issue. The choice between the two alternatives remains problematic, however (cf. Williams 1983).

4. THE DISTRIBUTION OF FEATURES IN TREES

(Morpho-)syntactic features are subject both to paradigmatic and syntagmatic constraints. We have very little to say here about the paradigmatic relations that they participate in. It is likely that the total set of features subdivides into a number of relatively independent subsets. Certain features are closely connected, such as those for gender, person and number, for example. On the other hand, it would seem that the gender feature has little interaction with the features for the tense system, say. Eventually one will expect there to be a full–fledged theory in which the features are grouped into hierarchically–ordered classes and subclasses, much as in phonology.

The syntagmatic relations among features are close in nature to those among syntactic categories, which is what the theory of syntax is all about. For example, the relation between two NP-positions in a chain is more precisely a relation between two syntactic positions which is subject to, among others, the condition that both be identical in feature content. (see also the discussion about the status of indices above.) Below, we will address some of the ways in which such relations of identity can be implemented by rules or principles of grammar.

Before doing so, however, let us dwell briefly on the intrinsic properties of trees on the one hand and features systems on the other. Trees incorporate essentially three types of information about relations between categories: dominance, left-right order and adjacency. Inlight of the modular decomposition of phrase structure, the latter two belong to the realm of (different subtypes of) government. But observe now that the