PARAMETRIZING THE NOTION 'HEAD'

Pieter Muysken

Universiteit van Amsterdam

Versions of the X-bar theory of constituent structure differ on the number, depth, and cross-classification of the projections from heads, but they concur in a general notion of 'head' as the lexical nucleus of a projection, θ-marking other constituents of the projection and determining its categorial status. Here I will explore some alternatives to this traditional view, arguing that the definitions of head valid in other components of the grammar, notably morphology, may play a role in phrase structure as well. Grammars may differ in the way heads are defined for different projections. The generalization along a few parameters of the notion head leads to a more general conception of the structure of projections, of which X-bar theory will be a specific variant.

The aim of this paper is to sketch a general theory of hierarchical structure in natural language. This theory should be able to characterize the features common to hierarchical structures in different components of the grammar, as well as provide a set of parameters along which structures may differ. In addition, such a theory should specify what features are invariant for different languages, and along which dimensions the hierarchical structures present in different languages can differ. The X theory and other theories of hierarchical structure (e.g., theories of word formation and theories of syllable structure) are argued to be not comprehensive theories in their own right, but rather specific instantiations of a more comprehensive theory of projections. These instantiations are the result of: (a) the selection of specific parameters of the general theory; (b) the interaction of the general theory of projections with specific subtheories, such as θ-theory, feature theory, and the theory of phonetic realization.

The results of this paper, then, are the sketch of a general theory of headed projections; a theory of homogeneous projections; and an indication of some types of parametric variation.

This paper is organized as follows. In section 1, various versions of the X theory are explored within a historical perspective, and a set of desiderata is outlined for such a theory. In section 2, I discuss the structures occurring in morphology, and I discuss various attempts to relate these structures to the ones found in the syntax, attempts which have used components of the X theory. In section 3, then, the various hierarchical structures which have been proposed in metrical phonology are discussed, and again related to the notions presented in sections 1 and 2. In section 4 I present a more general conception of hierarchical structure, summarizing the previous discussion. Section 5 is dedicated to a discussion of various extensions and variations which have been proposed in the literature, and an incorporation of these in the theory presented in 4.
Before going on, I should say that the present paper can be no more than a beginning. Limitations include: (a) a reliance on existing analyses and proposals, which in some cases are based on incompatible premises; (b) insufficient formalization; (c) a necessarily incomplete review of the literature, particularly the phonological literature. I have written it mostly because I think it is worthwhile to step back and survey accomplishments of the recent years, to try to integrate them into a comprehensive framework, and to note specific points of agreement and disagreement. Only in this way can specific questions for further research be isolated.

1. X Theory and Phrase Structure

While for some researchers in the last ten years X theory has been the focus of attention, for others it has been at most of marginal interest and often it has functioned as a sort of protective belt (in Lakatos' sense). That is to say, they have relaxed otherwise plausible conditions on phrase structure in order to gain coherent results elsewhere in the grammar. At the same time, it is remarkable that there is a broad consensus that something like X theory is needed, given the widely held conviction that natural language is in large part characterized by endocentric n-tuple articulation. The survey of the literature in (1), which is necessarily incomplete, indicates both where there is consensus and where divergence, with respect to the categorial content of the theory and the number of bars:

(1)

<table>
<thead>
<tr>
<th></th>
<th>MP=p^n</th>
<th>AP=Ad^p</th>
<th>VP=VPn</th>
<th>S=V^n</th>
<th>PP=p^n</th>
<th>QP=Q^n</th>
<th>Adv=Adv^n</th>
<th>AUI= juxtaposition</th>
<th>Det=Det^n</th>
<th>S=INFL</th>
<th>S=COMP</th>
<th>CLTIC</th>
<th>COMP</th>
<th>CASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chomsky 72</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bowers 68</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selkirk 81</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bresnan 73</td>
<td>2/3</td>
<td>2/3</td>
<td></td>
<td>2/3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vergnaud 74</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Siegel 74</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Williams 75</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bresnan 76</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emonds 76</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jackendoff 77</td>
<td>3</td>
<td>3</td>
<td>(2)</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Riemsdijk 78</td>
<td>3</td>
<td>3</td>
<td>(1) (2)</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Koster 78</td>
<td>3</td>
<td>3</td>
<td>(1) (2)</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groos 78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keuland 78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Borer 81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Riemsdijk 80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hale 81</td>
<td>1+n</td>
<td>1+n</td>
<td>1+n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chomsky 81</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stowell 81</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A survey of (1) shows that there is a consensus that the categories N, A, and V have a projection, but that for the rest there is little agreement to speak of. Specific differences in the literature include:

(2) a. the numbers of bars in a projection;
b. the nature of the V projection. Is VP the maximal projection of V? Is S? Is S'?
c. the categorial status of S and S', a question closely related to (b). If V is not the head of S or S', do they then fall outside of the X theory (as was assumed by Emonds, 1976, and others) or are they headed by INFL or AUX (as is assumed in much recent literature)?
d. is there a separate P projection? Perhaps this question has generally been answered in the affirmative since Jackendoff (1977);
e. are there projections from non-lexical categories? This question is among the most controversial, as the above schema indicates;
f. are there different types of projections, or are all projections of the same type?

Such a discussion of differences in content (which for the moment seems unsolvable) leads naturally to a discussion of the form of the X theory. A theory about phrase structure should specify at least some of the following:

(3) a. A specification of the categorial continuity between a head and its projection: must they be identical in categorial specification, partially identical, non-distinct?
b. A specification of possible heads: any category of type ±F, any element specified by the features ±N and/or ±V, any lexical element specified ±N, ±V?
c. A specification of possible non-heads or sisters: any type of element, only maximal projections, maximal projections and grammatical formatives/minor elements.
d. A specification of the type of projection nodes allowed: a numerical index (V^2, V^3) with or without maximum number a different type of specification, no specification besides categorial features.
e. Are heads expanded obligatorily?
(3) f. Are sisters expanded obligatorily?

g. A specification of the level of grammar for which the theory holds: only at deep structure, at deep structure and S-structure, at all levels of representation.

Rather than attempting to settle the points of disagreement visible in (1) and (2), I will redefine the X-bar mechanism in a more general way. The most standard version of X-bar theory, and the one with the greatest empirical content so far, is the one in which there is assumed to be a fixed number of bars (commonly two or three) for each projection, and in which there are specific non-head dependents at different specified levels of the projection, e.g. specifier at the two-bar level, complement at the one-bar level. These non-head dependents are then assumed to have cross-categorially stable semantic interpretations.

One of the original motivations of the X theory, the expression of cross-categorial generalizations at the 1-bar or 2-bar level concerning possible right-sisters, has disappeared with the developments of Θ-theory and Case-theory. These theories specify quite independently, using the notion of c-command, which elements must occur on the X' level, and which may not (cf. Chomsky (1981)). Thus there is no more need for specific indices to make, e.g., V'' comparable to N'. This is a desirable result, since these indices stand in the way of a more general notion of projection, common to syntax and morphology, and perhaps phonology.

I will argue here for a more modular view of X-bar, in which the notion projection from a head is generalized to include three types of nodes:

(4) head (-projection, -maximal)
projection (+projection, -maximal)
maximal projection (+projection, +maximal)

Assuming that precise definitions can be given for the features involved, the more familiar specific characteristics of X-bar theory for constituent structure will follow from an interaction of these general principles and configurational (e.g., c-command) restrictions on Θ-marking and Case assignment. Schematically:

(5) a. abolish the rule $X^j \rightarrow \ldots X^{j-1} \ldots$ (where $X$ is a feature matrix, and $j$ an integer $\geq 1$);

b. introduce two node features: $\pm$projection, $\pm$maximal, which define four types of nodes:\n
$X^0 = \text{projection}, \text{maximal}$

$X', X'' = \text{projection}, \text{maximal}$

$X''' = \text{projection}, \text{maximal}$

non-projecting minor elements = projection, +maximal;
(5) c. introduce a phrase structure rule of the following type:

\[ X \rightarrow \ldots X \ldots \]
\[ [+\text{proj}] \quad [-\text{max}] \]

This rule allows for the following configurations:

1. \[ [+\text{proj}] \quad [+\text{max}] \]
2. \[ [+\text{proj}] \quad [+\text{max}] \]
3. \[ [+\text{proj}] \quad [-\text{max}] \]
4. \[ [+\text{proj}] \quad [-\text{max}] \]

An advantage of the system given is that it combines the insight of the X-bar theory that projections have minimally two bars or levels with the possibility that for some projections (e.g. V-proj, when V' is VP, V'' is S, V''' is S' is rightly thought to be a bit shallow) there are more nodes of the type +projection, -maximal while others have minimal depth. The standard model allows flexibility of depth as well, of course, but not without loss of generality, e.g. allowing \( P_{\text{max}} \) to be two bar, and \( V_{\text{max}} \) to be four bar. Here both are simply maximal, the number of intermediate nodes being left unspecified by the theory. If the locality principles of the binding theory (boundedness, subjacency, etc.) are defined in terms of maximal projections, it is perhaps possible to either reduce the node types to two (+ and -projection), maximal being defined independently, or at least to explain why in morphology only heads and projections, but not maximal projections play a role.

d. It is perhaps not necessary to specify that all sisters must be maximal projections, since this can be made to follow from \( \theta \)-role assignment. Otherwise, we will need a filter of the following type:

\[ \ast \quad [\pm F] \]
\[ \quad [-\pm F] \quad [-\text{max}] \]

where \( \pm F \) is \( \pm N \) or \( \pm V \).

Suppose that minor categories are inherently maximal; then this would give as a result that all sisters of a projection were maximal.
(5) e. Assume for phrase structure that X is defined in terms of ±N, ±V.

f. (c) and (d) hold for all levels of representation: only maximal projections can move.²

The theory of phrase structure as outlined in (5) interacts with a theory of θ-marking, as in (6):

(6) a. all lexical items have an argument structure associated with them, which we will term "θ-grid";

b. the domain for θ-marking is the maximal projection of a lexical item (i.e. an element of the class defined by ±N, ±V);

c. the θ-criterion (cf. Chomsky (1981)).

This interaction adds specific characteristics to it, as will be apparent immediately.

There have been several proposals in the literature for trees as in (7):

(7) CL
     /   \
    CL   CL CL
  AUX  AUX  AUX
   /   /   /
  COMP COMP COMP

Suppose such trees fall under a separate sub-theory of minor homogeneous projections. These projections are categorially homogeneous and inherently maximal. They are defined by (8):

(8) a. all constituent elements defined in terms of the same minor feature;

b. a filter *[^proj,-max], ensuring one level depth;

c. formed by rule (5c), but the filter in (8b) blocks all but configuration (5c2);

d. the elements of θ-theory in (6) block elements of the category [±N,±V] in homogeneous projections. Suppose that major categories [±N,±V] can never participate in homogeneous projections since they involve θ-marking, and hence the θ-criterion would be violated.

Before concluding this section on phrase structure, I would like to remark that there are several possible relations between X theory and phrase structure, or indeed any other type of hierarchical structure:
(a) the theory defines the structure completely; the strongest claim, which is explored here;

(b) the theory defines only part of the phrase structure configurations, e.g. the projections of the major lexical categories, but not other parts, e.g. the internal structure of Aux and Comp, and the internal structure of the S'' (or 'E'' node);

(c) structure is freely formed, but the theory is part of the evaluation matrix, essentially defining the unmarked option;

(d) parts of the theory bear relation (a) or (b) to phrase structure, parts relation (c). E.g. the categorial determination principle may be definitorial, the theory of sisters evaluative.

2. **Morphology**

The generalized version of X' theory sketched in section 1 not only gives a more adequate and economical representation of what needs to be stated about phrase structure, it also makes it possible to compare phrase structure in a way both more precise and systematic to word structure. There are comparisons between morphology and syntax in the literature, but these are never specific.

Before stating the definition of morphological structure proposed here, I will briefly outline my assumptions about morphology with respect to national and self-hate, as in (9) (with a big debt to Williams (1981)).

\[(9)\]
\[
\begin{array}{c}
\text{a.} & \text{proj} \\
\text{proj} & \text{b.} \\
\text{max} & \text{max} \\
N & N \\
A & A \\
nation & -al \\
self & hate
\end{array}
\]

Compound heads have an independent θ-grid, and assign a θ-role to their complement within their (maximal) projection. The well-formedness of compounds depends crucially on the possibility of the assignment of a θ-role. Affixal heads have no independent θ-grid, but modify (optionally) the θ-grid of their complement. In Williams' analysis derivational affixes determine the argument structure of complex words, e.g. random (R), randomize (A, Th), but this claim is not tenable, I think. But even if it is, derivational heads differ from compound heads and phrase structure heads in that they do not θ-mark within their own projection.
With these assumptions (to which we will return throughout section 2), we can sketch a theory of word structure which parallels point by point (5) and (8).

(10) Defining characteristics of morphological structure:

a. the nodes in morphology are defined as ±projection, ±maximal (=(5b));

b. categorial continuity is defined by rule (5c);

c. complements are maximal (cf. (5d));

d. morphological heads are defined in terms of \([\pm N, \pm V]\) (cf. (5e));

e. complements are optional, as in phrase structure. (We must assume that the fact that affixes never occur as words is due to their not having an independent θ-grid, while all lexical items do.) Otherwise affixal complements would have to be obligatory;

f. Morphological projections are only one level deep, in the same way as minor homogeneous projections. They share the filter \(\neg [\pm \text{proj}, -\text{max}]\) (8b); (Note that it will not do to collapse \([\pm \text{max}]\) and \([\pm \text{proj}]\), since we assume that complements may be maximal, without being projections, and since the categorial continuity itself is defined in terms of projection.)

g. morphological trees are binary branching;

h. morphological heads are positionally defined in particular languages. (These last two characteristics may be related. Only within a binary branching structure can we define positionality.)

While this list of characteristics does not imply more than an interesting parallel at first sight, it has a number of interesting empirical consequences:

Locality. Suppose locality principles in syntax and morphology both involve the feature \([\pm \text{maximal}]\) as defining bounding nodes; then (5) and (9) would give the right results. Siegel's Adjacency and Williams' Atom Condition could look across exactly one boundary of the right type, just as in syntax. The difference would be that in morphology every boundary would count, given (10f);

θ-marking and compounds. Above we have assumed that the head of a compound θ-marks the other member. Assume that the rules for compound
formation are two-fold: a structure rule such as $Z \rightarrow Y X$, and a well-formedness condition which states that compounds fall within the $\Theta$-criterion essentially. The non-head is $\Theta$-marked by the head. This analysis captures the essential insight of Roeper & Siegel (1978) that compounds are (transformationally in their view) related to phrases (in which there is a $\Theta$-marking relation), but avoids the introduction of transformational power into the lexicon and the formally undesirable restrictions they impose on 'lexical transformations'. The OV character of English compounds in this view is due neither to a historical remnant nor to a transformational procedure, but to the fact that English morphology has right-most heads.

Morphologically complex prepositions. There may appear to be a gap in derivational morphology: while there are phrase structure projections from N, A, V, and P, so far all discussion on affixes has focussed on the categories N, A, and V. (10d) predicts, on the other hand, that there should be morphological heads (affixes or compound heads) of the type P. It can be argued that in Dutch and German, we find affixes of the category P, which can be used to derive complex Ps. Consider in Dutch:

(10) over boven
    'over' 'above, upstairs'

in binnen
    'in' 'inside'

neder beneden
    'down' 'below, downstairs'

uit buiten
    'out' 'outside'

Zuid bezuiden
    'South' 'to the South of'

Noord benoorden
    'North' 'to the North of'

While the relation between the prefix be- and the suffix -en remains to be studied here, it is clear that they are used to derive elements of the category P, and must have the status P themselves.

Particles. Verbal particles (call up) are to be analyzed as complements in verbal compounds, where they will have no projection of their own but be inherently maximal. The complement to a head in a compound is inherently maximal, since otherwise it would $\Theta$-mark the head.
Major Category output. Since all trees in morphology are crucially headed, and since heads which are not [±N, ±V] are excluded, we explain why affixes always result in lexical categories, and not in elements of the type Det, Case, Tense (if these latter elements are not also defined in terms of the major category features).

Case marking in compounds. There is no Case marking in compounds, only Θ-marking, because Case marking results from the interaction of lexical and structural properties, and crucially involves [-maximal, +projection] (=X', X'') nodes. The Case assigning properties of affixes need more detailed investigation.

Major Category Restriction. The framework as presented here has the Major Category Restriction due to Aronoff (1976) as an automatic consequence:

(i) affixes have no independent Θ-structure, but modify existing Θ-grids;

(ii) X+aff has the features [±N, ±V] since affixes have that feature specification, and hence the constituents headed by them also;

(iii) only [±N, ±V] elements have independent Θ-structure;

(iv) since X in (ii) contributes to the Θ-grid of the node dominating it, it must have a Θ-grid of its own, i.e. be of the category [±N, ±V].

We cannot derive the Unified Base Hypothesis, however, unless:

(v) Since affixes define functions over a specific type of Θ-grid, they are limited to such a type.

This is problematic, however, since the Unified Base Hypothesis is defined in categorial terms, and Θ-grids are defined independently of syntactic category. The precise status of the Unified Base Hypothesis is not clear at present, however. Note that the fact that affixes subcategorize necessarily, and cannot occur separately, is not postulated (as in earlier accounts, e.g. Lieber (1981)) but made to follow from Θ-theory.

Thus it can be seen that adopting the theory of morphology sketched in (10) leads to desirable empirical consequences in a number of domains. I will conclude the discussion of morphology by discussing a few more general problems about the relation between morphology and syntax in this and other models.

Williams (1981) suggests that heads are defined positionally in morphology (cf. (10h)), the Righthand Head Rule. This conflicts with the
definition of affixes as heads in the case of prefixes, since Williams allows specific leftmost heads only in the case of affixes. Suppose that there is a relation between the binary branching option of morphology and the positionality option that holds in this component. Notice that only in a binary structure can we unambiguously define positionality, and therefore heads are not defined positionally in syntax. (I am disregarding here Kayne's (1981) claim that binary branching plays a crucial role in syntax as well, although the existence of some n-ary branching is not incompatible with Kayne's proposal. There, binarity is not definitional for phrase structure, but rather necessary to achieve a certain number of unambiguous paths for those configurations which involve government.)

Following an idea of Lieber (1981), I assume that specific restrictions on affixation (sensitivity to 'roots' vs. 'stems', or whatever labels one chooses for specific types of elements, sensitivity to phonological and stratal characteristics of the base, etc.) are marked in the subcategorization frames of affixes, and hence locally. As is obvious from the discussion above, the parallel which Selkirk (in preparation) draws between word structure and phrase structure in terms of levels of labelled nodes (V' and N'' vs. Vroot and Nstem, etc.) but not categorial continuity and headedness, is not expressed in the framework here.

Finally, it should remain clear that the 'unified' system sketched here does not result in a blurring of the divisions between word structure and phrase structure. There is an essential discontinuity between them. A maximal projection in morphology (the word) corresponds to a non-projection in phrase structure (the lexical head).

3. Phonology

Several types of hierarchical structures which have been proposed by phonological sequences seem amenable to discussion in the framework presented here. These include the syllable, the foot, and the segmental projection. We will discuss these very briefly in turn.

In (11) we present the traditional structure assigned to the syllable together with its reinterpretation in the present framework:

\[
(11) \quad \begin{array}{c}
\text{syllable} \\
\text{onset} \quad \text{rhyme} \\
\quad \text{nucleus} \quad \text{coda}
\end{array} \quad \begin{array}{c}
\text{[+syll]}
\text{proj} \\
\text{[-syll]}
\text{proj}
\end{array}
\]

The traditional labels for the nodes of the syllable are replaced by feature bundles:
Besides the conceptual advantages of incorporating the syllable tree within the class of headed projections and being able to define the notion 'branching node', crucial in many phonological accounts, in terms of the feature [+projection] there are some disadvantages to the present system as well. While the existence of both the nucleus and of the syllable node itself are predicted, there is no explanation of why there is only one level of rhyme, and not several intermediate nodes. Additional structure which the syllable may possess would need to follow from other considerations.

Similarly, the status of the onset and the coda seems very different from that of complements in phrase structure, but this is due to the fact that in syntax the theory of projections interacts with θ-theory. Perhaps in the case of the syllable, the theory of projections interacts with a theory of pronounceability.

Just as the syllable may be perceived as the projection of +syllabic elements, the foot can be seen as the projection of the syllable, as in (13):

\[
\text{(13)} \quad \text{[foot]}
\]

\[
\begin{array}{c}
\text{[syllable]} \\
\text{[syllable]} \\
\text{[syllable]}
\end{array}
\]

\[
\begin{array}{c}
\text{[syll]} \\
\text{[syll]} \\
\text{[syll]}
\end{array}
\]

Again we have the conceptual advantage of simplicity and economy, but there is still too much uncertainty about the precise status and internal structure of feet to determine whether the notion of projection is adequate for this type of structure.

Just as the syllable and the foot can be seen as headed projections, the segmental trees (vowel trees, nasal trees, etc.) proposed by Vergnaud (1978) can be seen as minor homogeneous projections. They have a structure as in (14), where F stands for a phonological feature:

\[
\text{(14)} \quad \text{[}\alpha F\text{]}\text{proj}
\]

\[
\begin{array}{c}
\alpha F \\
\alpha F \\
\ldots \\
\alpha F \\
\alpha F
\end{array}
\]
Syllables and feet have the defining characteristics given in (15), and segmental trees those in (16). These parallel again the characteristics given for phrase structure in (5) and (8) and for morphology in (10):

(15) a. nodes in phonology are defined by distinctive features and as [+projection, -maximal];
    b. categorial continuity is defined by rule (5c);
    c. complements need still be specified in detail;
    d. heads are specified by distinctive features from a fixed inventory;
    e. complements are optional;
    f. syllabic and foot projections are more than one level deep (though not indefinitely deep);
    g. headed phonological trees (cf. (11) and (13)) are binary-branching;
    h. phonological heads are positionally specified as rightmost or leftmost; this is not the case for syllables, however, where the head is invariant in position.

(16) a. homogeneous segmental projections share the filter in (8b) [+projection, -maximal], creating one-level deep projections;
    b. are formed by rule (5c);
    c. all sisters are non-distinct with respect to the specification of one feature or more.

As was the case with morphological projections, the foot projections have the combination of binarity and positionality. Syllabic projections show binary branching, but are like phrase structure projections in that the head is not defined positionally but categorically, I assume.

4. Towards a General Conception of Hierarchical Structure

The survey of projections in syntax, morphology, and phonology in sections 1, 2, and 3, respectively, and the tentative specification of these projections along the lines of a number of parameters allows us to compare the projections more generally, as in (17), which is simply a summary of the discussion so far:
Of these parameters, (17a), (17e), and (17f) are fixed for a single value. At least the first two can be taken as definitional for projections.

In all other cases, we find different options realized, and we may wonder whether these options result from the intersection of the theory of projections with other theories. This is certainly the case with (17c).

First of all, θ-theory would block the occurrence of full lexical elements of the class [±N, ±V] in minor homogeneous projections, if the projection were specified in this way. It may be the case that lexical elements could be part of a minor projection (e.g. verbs in AUX) where the projection is not defined in terms of the features of the verb, [−N, +V]. It is tempting to think that a similar distinction holds between syllabic and segmental projections. Consider a phonological theory on a par with θ-theory, crucially involving a feature such as [±syllabic] (which would then play a role similar to [±N, ±V]). Then segmental projections could not be defined in terms of this crucial feature either. In this way, (17c) would have three values:

<table>
<thead>
<tr>
<th>(a) base rule (5c)</th>
<th>(b) feature relation head-projection:</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) identical</td>
<td>(b) non-distinct</td>
</tr>
<tr>
<td></td>
<td>(c) specification heads:</td>
</tr>
<tr>
<td></td>
<td>(a) ±N, ±V</td>
</tr>
<tr>
<td></td>
<td>(b) other features</td>
</tr>
<tr>
<td></td>
<td>(d) specification sisters:</td>
</tr>
<tr>
<td></td>
<td>(a) maximal projections</td>
</tr>
<tr>
<td></td>
<td>(b) any element</td>
</tr>
<tr>
<td></td>
<td>(e) obligatory head</td>
</tr>
<tr>
<td></td>
<td>(f) optional sisters</td>
</tr>
<tr>
<td></td>
<td>(g) head assigns θ-role</td>
</tr>
<tr>
<td></td>
<td>(h) head assigns Case</td>
</tr>
<tr>
<td></td>
<td>(i) filter [+proj, -max]</td>
</tr>
<tr>
<td></td>
<td>(j) branching:</td>
</tr>
<tr>
<td></td>
<td>(a) n-ary</td>
</tr>
<tr>
<td></td>
<td>(b) binary</td>
</tr>
<tr>
<td>(k) head defined positionally</td>
<td>+</td>
</tr>
</tbody>
</table>
(17) c'. (a) ±N, ±V
   (b) ±syllabic
   (c) other features

The choice of a particular value would be dictated, naturally, by the component of the grammar and, with respect to (c) versus (a), (b), by the interaction of the theory of projections with θ-theory and the theory of articulation (or another phonological theory). The same interactions would be responsible for the specification of sisters parameter (17d), and the Case and θ-role assignment parameters (17g) and (17h).

This leaves us with one level depth filter (17i), the binarity parameter, and the positionality parameter. Besides the remark made earlier that the latter two seem related, I have nothing to say on this issue at present.

5. Parametric Variation Between Languages

In the previous section we have explored the parameters along which heads can vary in the different components of the grammar, assuming for each component the unmarked specifications for heads in that component. If we consider (17), however, we see that quite a number of potential combinations of parametric specifications remain unrealized, in the discussion so far. Here we will explore the idea that some of these combinations do occur in specific languages or types of languages, so that the list of parameters for projections in (17) also holds for variation between languages.

A first parameter is the identity/non-distinctness specification for categorial continuity. Non-distinctness may play a role in affixal morphology in the case of the Spanish diminutives and similar phenomena (cf. Lieber (1981), and the discussion summarized there). The affix -ito/-ita can diminutivize both nouns and adjectives, a situation incompatible with the analysis of affixes as in (9a), since there affixes (as heads) are assumed to have a unique categorial specification. Suppose that -ito is specified as [+N, +V]; then its projection can be either [+N, +V] or [+N, -V], depending on the nature of the base for -ito affixation. This means though that here non-distinctness holds in the projection rather than identity.

A somewhat similar situation holds for German adjective-participle constructions (Van Riemsdijk (1980)), which can be analyzed as [aN, +V]proj (in my notation), and are dominated either by [+N, +V] or by [-N, +V]. Again, non-distinctness must be assumed to allow for this type of projection.

A more complicated case of non-distinctness within a headed phrase structure projection would be that of Quechua nominalizations and postpositions, as in Lefebvre & Muysken (1981). These are analyzed as
involving a phrase structure rule as in (18a), which defines the configurations in (18b):

\[(18) \begin{align*}
a. & \quad \left[\alpha F_1\right]^{\text{proj}}_{-F_2} \rightarrow \left[\alpha F_1\right]_{+F_2}, \text{where } \pm F \text{ is } \pm N, \pm V \\
b. & \quad \left[+N\right]^{\text{proj}}_{-V} \left[-N\right]^{\text{proj}}_{+V} \left[+N\right]^{\text{proj}}_{+V} \left[-N\right]^{\text{proj}}_{-V} \\
& \quad \left[+N\right]^{\text{proj}}_{+V} \left[-N\right]^{\text{proj}}_{-V} \left[+N\right]^{\text{proj}}_{-V} \left[-N\right]^{\text{proj}}_{+V}
\end{align*}\]

Again, non-distinctness is crucially needed as a marked option.

A second parameter involved in interlinguistic variation may be filter \((8b) \ast [+\text{proj}, -\text{max}]\) ensuring one-level depth of projections. Hale (1981, postscript) has suggested that \(W^*\) languages are subject to something equivalent to \((8b)\) in their major headed projections as well as elsewhere.

This concludes the discussion of heads in this paper. The account has been given at a level of generality at which it is difficult to survey all empirical consequences. I hope the level of precision, at least, of the paper will make it possible to study proposals for the theory of projections more fruitfully in the future.

**NOTES**

\(^\dagger\)This is substantially the text of talk with the same title given at the GLOW colloquium on dependencies, Paris 1982. The major difference is that the section on parameters in percolation systems has been taken out, for further elaboration elsewhere. I am grateful to the participants of the colloquium, particularly Henk van Riemsdijk, for comments on the spoken version.

\(^1\)R. Kayne suggested that it may be possible to abolish non-categorial node-specifications altogether, and to ensure the well-formedness of projections by requiring that two sisters be always of a different category. In \((a)\) I present some projections blocked by the system in \((5)\), in \((b)\) some projections blocked in the proposal made by Kayne:

\[(a) \quad \ast N^{\text{proj}}_{\text{max}} \quad \ast V^{\text{proj}}_{\text{max}}\]
For these cases, the proposals have the same empirical consequences. Consider (c) and (d), however. Here in (c) the proposal of Kayne's makes the correct prediction directly, while in (d) Kayne's proposal is insufficient in its most simple form if one assumes that S' is the maximal projection of V:

(c) \[ \ast N \text{proj} \]
\[ N \quad N_{\text{max}} \]

(d) \[ \ast V \text{proj} \]
\[ V \quad V_{\text{max}} \]

My own proposal, to be sure, does not rule out the ungrammatical (c) but it does predict that (d) is grammatical.

2Y. Aoun has pointed out that the system in (5) has the desirable consequence that after adjunction, a new \([+\text{maximal}]\) node is created, incorporating the adjoined element.

**BIBLIOGRAPHY**


