A Processing Model for Warlpiri Syntax
and Implications for Linguistic Theory

Barbara A. Brunson

Technical Report CSRI-208
July 1988

Computer Systems Research Institute
University of Toronto
Toronto, Canada
M5S 1A1

Preface to Technical Report Version

This report presents my Masters Forum paper, completed in September 1986. Aside from changes in format, the main differences between this version and the original is the omission of the appendix of annotated parser code and screen-image snapshots of the parser in action. It was felt that a higher level discussion of the parser and a detailed trace of a sample parse would be more appropriate for the wider circulation of this report. To that effect, the discussion in Chapter 4 has been modified and extended.

Small additions and refinements have been made in the earlier chapters, retaining the essential content, but hopefully making the presentation somewhat clearer. Also, a Postscript has been added, summarizing some of the directions that this work has taken since 1986.

A Processing Model for Warlpiri Syntax
and Implications for Linguistic Theory

by

Barbara A. Brunson

Submitted to the Department of
Linguistics

Linguistic Forum, 1986

In Partial Fulfillment of the
Requirements of the
Degree of

Master of Arts

at the
University of Toronto
Sept. 1986
(revised and submitted as a Technical Report
to the Department of Computer Science,
University of Toronto
July, 1988)

Advisors:
        Prof. E. A. Cowper, Dept. of Linguistics
        Prof. G. Hirst, Dept. of Computer Science

Abstract

Much of the development of the current Government-Binding (GB) theory of syntax has progressed independently of concerns raised in theories of language processing. Similarly, models of syntactic processing are often proposed that lack any underpinning in syntactic theory. The work described in this report focuses on the language Warlpiri, an Australian aboriginal language with properties that are difficult to reconcile with most theories of Universal Grammar — properties such as free word-order and discontinuity. This language is studied from the two-fold perspective of establishing a linguistically and computationally sound processing model. This forces the linguistic model to be sufficiently precise to satisfy the demands of implementation as well as forcing the implementation to proceed in a linguistically principled way.

This report presents a portion of Warlpiri grammar in a revised GB-based account, addressing the issues of parsability, as well as more theoretical syntactic issues, that together force a reassessment and parametrization of certain linguistic principles. In particular, a revised version of *theta* theory and the notion of thematic identification are readily interpreted into processing strategies that extend naturally to deal with adjuncts and non-subcategorized arguments in a wide range of languages. The complementary nature of the syntax and morpho-syntax in the satisfaction of syntactic principles as well as in the construction of syntactic representations is addressed, as is the crucial relevance of prosodic information for preserving determinism in the parsing algorithm.

Acknowledgements

Table of Contents

CHAPTER 1

Introduction

In what ways can languages differ? How different can languages be? These are questions that are central to any generalized study of grammar. If we propose a universal grammar, we maintain that there are principles common across languages. What is then needed is a method of characterizing and constraining these principles to allow for the variability that exists in the languages of the world.

Extensive research done by Kenneth Hale and others on Warlpiri, an Australian aboriginal language, has produced a rich source of data[1] on a language with properties that initially seem to preclude a standard syntactic analysis. Being so different, Warlpiri is a natural focus for studying and assessing the extent to which languages can vary, at least within their syntactic component.

Adopting the general syntactic framework of Government-Binding (Chomsky, 1981, 1982, 1986), the task at hand is to reassess the Warlpiri data. In some previous analyses (Hale, 1981, 1982), Warlpiri has been viewed as typologically distinct from, say, English. In such approaches the differences that Warlpiri's syntax exhibits are captured through a declaration of typological classes of languages. In other analyses (Hale, 1983) aspects of the syntactic structure of Warlpiri are thought not to differ greatly from those of other languages. These analyses generally require an abstract notion of syntax, since constituency and precedence relations in a syntactic structure are not necessarily preserved in the linear sequence of words that make up the Warlpiri sentence.

My approach takes a somewhat different angle. I assess the substantive differences between Warlpiri and English in an attempt to isolate the specific principles of universal grammar that are involved in these differences. What emerges is a sort of minimal theory of Warlpiri syntax, leaving as much of universal grammar constant as possible. The syntactic model of Warlpiri is not an abstract level of representation differing in significant respects from the surface string, but is merely a representation of the observable syntactic properties of the language. To the extent that Warlpiri syntax is either depleted or

---

[1] I am greatly indebted to the many excellent papers on Warlpiri that have been written by Ken Hale. These have been my main source of data on this language. I have also made extensive use of David Nash's thesis *Topics in Warlpiri Grammar*. I should point out that *all* of my information on Warlpiri comes from the field work of other researchers — I have done no data elicitation of my own on this language. For a complete list of the references and data sources on Warlpiri that I have used, see the Bibliography.

different, our theory of syntax must be revised to allow this.

Once we develop a theory of Warlpiri grammar, we need to translate this theory into a processing model of Warlpiri in order to complete the picture. By translating the principles of grammar into parsing procedures we have a method of checking the theory for explanatory adequacy, internal consistency, and coverage of the data. In addition, a parsing mechanism can indicate those aspects of the data that, as yet, lack an explanation in the theory.

The task then is a two-fold one. First, we need a theory of Warlpiri grammar. Second, we need a translation of this into a processing model that, ideally, would function as a test of the theory itself. By working within an existing and general syntactic framework, any parsing model that emerges should (with modification of language-specific properties) be transparently extendible to other languages.

In the remainder of this chapter, I provide a brief introduction to the theory of Government and Binding, as well as an introduction to the general properties of Warlpiri syntax, focusing on the difficulties that these properties raise for theories of language and parsing.

1. Syntax

1.1. The Theory of Government and Binding (GB)

A general goal of GB research is to extract generalizations from the properties of languages and present these generalizations in a manner that provides an explanation (at least within the theory) of the properties that they endeavour to characterize. If we refer to these generalizations as principles of grammar, then the aim is to minimize the number of principles while maximizing their actual descriptive coverage. We want a theory of grammar with as few rules as possible where the properties of languages can be predicted from a small set of principles and their interactions.

The strategy used to attain this goal is to modularize the theory of grammar. Levels of representation are defined and sub-theories developed that divide the description of language into manageable parts. By isolating levels of representation and sub-theories, linguistic phenomena can be categorized according to level and sub-theory. As new phenomena are discovered they are assessed in the existing theory and the theory is revised, as needed, to accommodate the new data. For this reason, the theory is a limiting theory in a continual state of change and progress. By studying a wide range of linguistic phenomena, it is hoped that eventually the theory will need fewer and fewer revisions and a finite set of principles will have been discovered.

The levels of representation are, to a large extent, tools that help to break down the task of defining a grammar, although a considerable amount of attention has been devoted to making these divisions as realistic and natural as possible. These levels are not claimed to be exhaustive, and for some phenomena it

is hard to decide what level of representation should be involved. Nevertheless, they do provide a useful classification of most of the phenomena that are currently being investigated. The generally accepted levels of representation are illustrated in (1).

(1)

```
                    d-structure
                         |
                    s-structure
                       /    \
                      /      \
           PF                      LF
     (phonetic form)         (logical form)
```
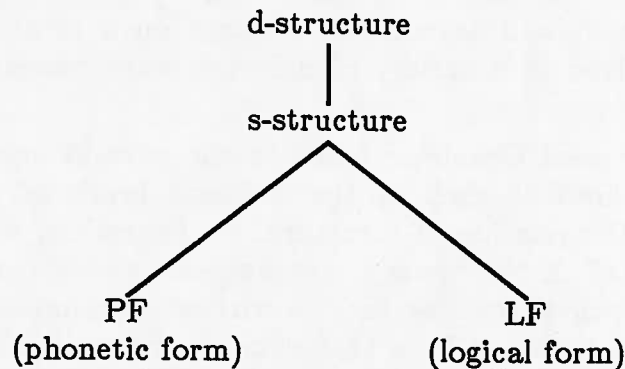
The connecting lines here do not indicate that the one level of representation is derived from the other, but rather they indicate the existence of a mapping defined between the various levels. These mappings define well-formedness conditions on representations. D-structure and s-structure collectively form the syntax in this model. D-structure is a level of pure argument structure. S-structure corresponds to the surface syntactic structure, annotated with traces and empty constituents. Move-α, the mapping between d-structure and s-structure, together with trace theory, constrains the possible binding of elements and empty positions. The remaining two levels of representation, phonetic form and logical form, provide the phonetic realization and partial interpretation (especially with respect to scope and quantification) of the sentence.

The *Projection Principle*, as a constraint on all levels of representation, can be seen as moving much of the descriptive burden of the syntax into the lexicon. It is defined in (2).

(2) *Projection Principle*: Representations at each syntactic level are projected from the lexicon, in that they observe the subcategorization properties of lexical items. (Chomsky, 1981, p. 29)

The function of defining just what the form of an acceptable representation at a particular level should be is taken on by the sub-theories of grammar. In this way, the sub-theories and principles of grammar can be viewed as filters operating on representations. The theory itself has little to say about how a sentence's structure is determined. The focus is on determining when a representation is valid, with the intent that only valid representations will correspond to valid sentences in the language. Since the focus is not on the generation of structures, they may be generated in any way whatsoever — even randomly. The method of generation is unimportant — only the resulting structure is needed to verify that the sentence is well-formed.

Since the early days of generative grammar, the theory of grammar has been viewed as defining a set of relations in terms of which the structure of the human language faculty can be articulated. In this way, the relations defined become a Universal Grammar. More recent developments allow the relations of grammar to be parametrized to account for rich and diverse inter-lingual variability while at the same time capturing similarities between languages. These parametrizations can be viewed as multiple settings on a principle or relation that allow it to be realized in a variety of different ways corresponding to the various language phenomena.

This notion of Universal Grammar holds in the current conception of the GB theory. Relations hold at each of the different levels of representation. Consider, for example, the relation of structural configuration, which is defined in terms of principles of X-bar theory, government and c-command (to be defined later). These principles allow for the virtual elimination of the traditional phrase structure component from the grammar (Stowell, 1981), yet, at the same time, the class of possible phrase structures is reduced in an explanatory fashion that is preferable to the diagnostic descriptive power of the base component rules.

One criticism that can be levied against the GB approach is that the principles of language are crucially defined in the notation of the theory. In this way, the notation becomes a primitive of the theory, rather than just a means of expressing the theory. This is particularly problematic if we attempt to claim psychological reality for the particular principles that we extract through observing linguistic properties. To the extent that we can claim psychological reality, it is only in an abstract sense. The real test is theory-internal: Does the principle simplify the grammar and at the same time describe the observable properties of language? Although introduced as a criticism, the situation described above is unavoidable. Until an adequate account of the 'real' mechanisms behind linguistic function is developed, there is little choice but to propose a theory-dependent model if we wish to pursue linguistic modelling at all.

I will now discuss briefly the various pertinent sub-theories in GB theory. For a more complete introduction to GB theory, see Sells (1986) and van Riemsdijk and Williams (1986); and for a more detailed discussion of GB see Chomsky (1981, 1982, and 1986). Readers already familiar with GB theory may turn to section 2 of chapter 1.

1.2. Subtheories of Grammar

1.2.1. X-bar Theory

X-bar theory provides a method for capturing cross-category regularities in the phrase structure of a language. Phrase structure generally is used to capture hierarchical relationships and structure, and it is commonly represented in tree diagrams (or equivalently, bracketed strings). Traditional phrase structure

rules were stated in the form of rewrite rules
$$(eg.\ \mathbf{S} \rightarrow \mathbf{NP}\ \mathbf{VP}).$$
These rewrite rules are quite unconstrained, and could be used to create structures that never occur in language. For instance, the rewrite rule
$$\mathbf{NP} \rightarrow \mathbf{V}\ \mathbf{PP}$$
would be acceptable to a theory of grammar that allowed unconstrained rewrite rules although it defines an impossible phrase structure for language.

In the attempt to restrict the class of possible phrase structures, defining principles of acceptable phrase structures have been discovered and articulated in a generalized notation that has come to be known as X-bar theory. Central to X-bar theory is the notion of the *head* of a phrase. The head of a phrase is the lexical item whose category defines the function of the phrase. So, for example, the verb is the head of the verb phrase, and the noun is the head of a noun phrase.

There may also be structure intermediate between the phrase and the head. X-bar notation has been developed to make it easier to refer to the various levels of structure within a phrase. If we have a lexical item with a particular category, say $\mathbf{X}$, and this item is the head of a phrase, then that phrase will be an $\mathbf{XP}$. Alternatively, using X-bar notation, the head is an $\mathbf{X^0}$ or a zero-level category; an intermediate constituent is an $\mathbf{X'}$ or one-level category; and the phrase itself is an $\mathbf{X''}$ or two-level category. It is generally assumed in the current literature that a maximally two-leveled X-bar structure is part of Universal Grammar. Using the familiar rewrite rules, the X-bar notation can be summarized as in (3).

(3)

$$\mathbf{X''} \rightarrow ...\mathbf{X'}...$$
$$\mathbf{X'}\ \rightarrow ...\mathbf{X^0}...$$

In a particular phrase, the only non-maximal categories (*ie.* zero or one-level categories) that can be present are the head of the phrase or the intermediate constituent of the phrase. All other categories are optional and must be maximal projections (*ie.* two-level categories). This constraint, as well as the two-level structure, can be captured in the X-bar schema presented in (4).

(4)

```
           X''
          /    \
    specifier    X'
                /   \
              X     complement
```
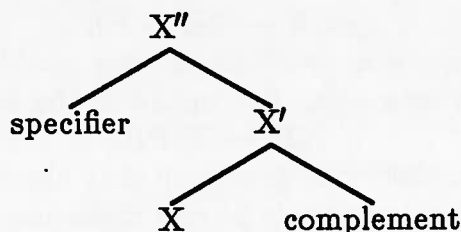
$$X'' \to \text{specifier } X'$$
$$X' \to X^0 \text{ complement}$$

The terms specifier and complement are used to refer to the structural positions sister of X' and sister of X^0, respectively. They are not node labels.

The order of the elements on the right-hand side of the rules presented in (4) is determined by the language, and can vary. The particular choice of this order is referred to as the *headedness* of the language. In the example given in (4), the schema is general yet the headedness is specific to a particular language, in this case English. English has the property that X'' branches in a right-headed fashion. This means that the one-level category is on the right of the specifier position. The one-level category can be thought of as the head of the two-level category. The one-level category is left-headed; that is, the head occurs to the left of the complement position. By allowing recursion of the X' category, we define the class of *modifiers* (sister constituents to a recursive X').

1.2.2. The Structure of S

There has been considerable research recently on just how the uppermost structure of a sentence might fit this schema. Discussion can be found in Stowell (1981),and Chomsky (1986). The current position on the structure of S is that the sentence is actually a maximal projection of the complementizer (Comp)[2] (Chomsky, 1986). This category, although not necessarily possessing an overt lexical item as its head, follows the general schema. Wh-question words occupy the specifier position, and the head, when overt, is a complementizer. The complement of this phrase is a maximal projection referred to as I''. This phrase also obeys the general schema. Its specifier is generally the subject argument, while its complement is V'', the verb phrase. The head is the category I, or **Infl**, which includes tense and agreement elements as well as modals. The categories **Comp** and **Infl** are referred to as non-lexical categories because they can not be defined in terms of the lexical category features $[\pm N]$ and $[\pm V]$[3]. Although it is accepted that all lexical categories project (that is,

---

[2] Not to be confused with the complement of X'.

[3] The lexical category features are discussed in detail in Chomsky (1981). Briefly, the feature $[+N]$ identifies those categories that possess a property that could be called *substantive*. $[+N]$ categories include nouns and adjectives. The feature $[+V]$ identifies those categories that possess the property *predicative*. $[+V]$ categories include verbs and adjec-

that they are the head of a phrase), the claim that non-lexical categories project as well is a fairly recent innovation (Chomsky, 1986).

The upper level structure of a clause, with headedness set as in English, is illustrated in (5).

(5)

```
                        C″
                       /   \
            (wh-phrase)      C′
                            /   \
                           C     I″
                                /   \
                              NP     I′
                           (subject) /  \
                                    I    V″
```
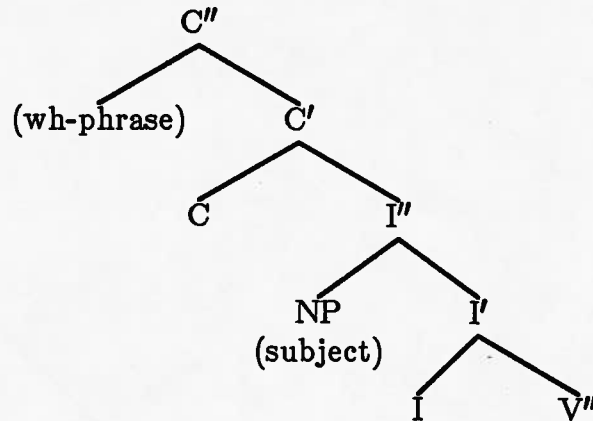
1.2.3. Structural Relations

Several structural relations defined on the hierarchical structure of a sentence figure prominently in later chapters.

The structural relations of precedence and dominance are central to the following definitions. Precedence is defined on the zero-level categories (leaf nodes) of a sentence as simple linear precedence. Dominance is defined on the hierarchical tree structure. A node **A** dominates a node **B** in the tree structure if it occurs on the direct path from **B** to the root of the tree (*ie.* the single initial node).

The structural relation of *c-command* is defined as in (6).

(6) α *c-commands* β if and only if α does not dominate β and every γ that dominates α also dominates β. (Chomsky, 1986, p. 8)

γ is normally restricted to branching nodes. What this means is that the first branching node dominating α also dominates β. In the examples of (7), α c-commands β.

---

tives. Prepositions have the features [−N] and [−V].

(7)
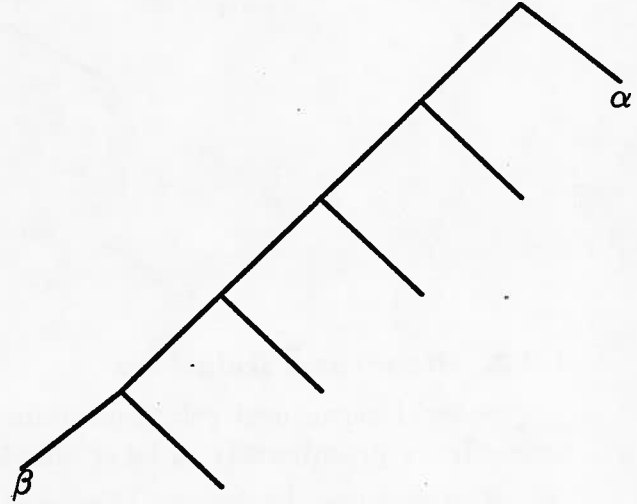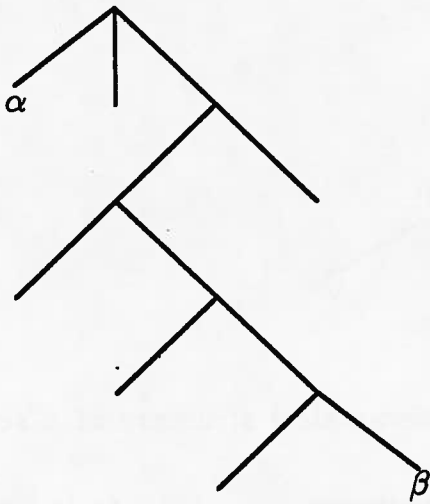
M-command (Chomsky, 1986) is a variant of c-command. Its definition is virtually the same as that of c-command except that the γ is restricted to maximal projections (*ie.* $\mathbf{X''}$).

We say that a node, *A,* is *adjoined* if it occurs in the structural configuration shown in (8).

(8)
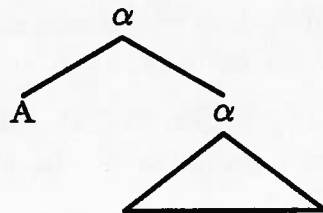
The node **A** is said to be an *adjunct*; the node α, the *adjunction node*; and structures such as these are referred to as *adjunction* structures. Notice that this structure is not produced directly from the X-bar schema, since the two α's are of the same bar-level. In adjunction, the category of the adjunction node is copied and the adjunct is located as the sister of one of the copies and the daughter of the other. Adjunction is generally limited to maximal projections and $\mathbf{X^0}$ categories, and, in addition, the adjunct is generally restricted to be of the same bar-level as the copied node. In the case of adjunction to $\mathbf{X^0}$

categories, the term *head adjunction* is used.

Since the adjunction node is copied, it is necessary to consider whether the uppermost adjunction node dominates the adjunct. As a convention, dominance is assumed not to hold in such situations, dominance being reserved for the situation where all segments of a node dominate another node. To make more specific just how such disjoint nodes as the copied adjunction node will be treated in the theory, we define *exclude* as in (9).

(9) α *excludes* β if no segment of α dominates β (Chomsky, 1986, p.9) (where *dominates* is in the original sense and not the revised sense presented just above).

What we have then is that adjuncts are not excluded by their adjunction node, neither are they dominated by it.

The structural relationship *government* is central to GB theory and is defined as in (10).

(10) α *governs* β if and only if

> i) α m-commands β
> and
>
> ii) there is no γ, γ a barrier[4] for β, such that γ excludes α.
> (Chomsky, 1986, p.9)

The details of this definition may seem a bit mysterious. Recall that m-command made no restriction on how deep in the structure the β could be. That is, α could m-command β and β be embedded arbitrarily deeply into the structure. The second clause of the definition for government essentially imposes a locality condition on the distance between the α and the β. In this way we can view government as a sort of local m-command, where the specific definition of locality is yet to be made explicit.

1.2.4. Trace Theory and Movement

Move-α is commonly defined as a rule that moves[5] anything anywhere in a syntactic representation. To the extent that movement by move-α is constrained at all, it is only through the actions and interactions of other principles of grammar. When movement occurs, the moved constituent occupies a new position and leaves a trace in the gap that was its original position.

There are two types of movement: movement to the specifier position of a phrase and movement into an adjunction structure. This follows from

---

[4] The concept of barrier will be discussed in section 1.2.6, but for now it is sufficient to realize that a barrier is a certain type of maximal projection.

[5] Although I talk about movement here, it is equally acceptable, and possibly even preferable, to consider move-α as a mapping between levels of representation that constrains the identification of traces or gaps and displaced constituents rather as a movement rule *per se*. The use of the term 'movement' here and elsewhere should not be taken as a

interactions of other principles of grammar and need not be stipulated in the theory (see Chomsky, 1986). Movement is even further constrained by *subjacency*, which imposes a locality relationship on the identification and co-indexing of a constituent and its trace. The two major instances of movement to specifier position are movement to the specifier of C'' and movement to the specifier of I''. These positions are fundamentally different. The specifier of I'' is an argument position (recall that it is the subject position) while the specifier of C'' is not an argument position. We call movement to an argument position *A-movement* and movement to a non-argument position *Ā-movement*. Traces of movement are not free to occur anywhere in the structure. There are principles of grammar that constrain their occurrences. These principles are discussed in the next section.

1.2.5. Proper Government and the Empty Category Principle

The *Empty Category Principle* (ECP) is a well-formedness condition that applies to the empty categories that are created by movement (*ie.* traces). The ECP is stated as in (11).

(11) *The Empty Category Principle*: Every trace must be properly governed.
(Chomsky, 1986, p. 16)

Proper government is defined in (12).

(12) α *properly governs* β if and only if

       i) α *θ-governs* β
      or
      ii) α *antecedent-governs* β
      (Chomsky, 1986, p.17)

θ-government and antecedent government are special cases of government and they are discussed in the next section.

1.2.6. Barriers

In Chomsky's 1986 book, *Barriers*, the locality conditions of government and the locality condition of subjacency are brought together in a generalized theory. What emerges is a theory of *bounding* that involves some new definitions. We have already seen certain definitions that involve ideas from this theory. The definition of *government* refers to *barriers*, and the definitions of *θ-government* and *antecedent government* also rely on definitions proposed in Chomsky, 1986.

The definition of *θ-government* requires that three conditions hold. This is presented in (13).

---

stand on this issue.

(13) α *θ-governs* β if and only if

> i) α is an X^0.
> ii) α directly θ-marks β
> and
> iii) α and β are sister nodes in the structure.
> (Chomsky, 1986, p. 18)

θ-marking will be discussed in section 1.2.8. Notice that this relationship is made local by clause iii.

The definition of antecedent government is essentially the same as that of government, with the additional requirement that α and β be coindexed through move-α, with α the antecedent of the trace β.

Locality is assured for proper government. It is either constrained by θ-marking of sisters or by the locality constraint on government. The locality constraint on government is stated in terms of the concept of *barrier*. The definitions in (14) and (15) are preliminary to the definition of barrier and are presented below.

(14) α *L-marks* β if and only if

> i) α is a lexical category
> and
> ii) α θ-governs β (or α θ-governs γ
> where β agrees with the head of γ).
> (Chomsky, 1986, p. 70)

(15) γ is a *blocking category* for β if and only if γ is a maximal projection, γ is not L-marked, and γ dominates β.

The property of being L-marked holds of certain nodes in certain configurations defined by the relationships between constituents in that configuration. This property is relevant in determining whether the node is a blocking category for the constituents and categories that it dominates. The notion of a blocking category figures in the definition of barrier presented in (16).

(16) γ, a maximal projection, is a *barrier* for β if and only if

> i) γ immediately dominates δ, δ a blocking category for β.
> or
> ii) γ is a blocking category for β, $\gamma \neq$ **IP**.
> (Chomsky, 1986, p.14)

The two clauses in the definition of barrier define two ways that a category can become a barrier. Clause i defines inherited barrierhood. If a potential barrier dominates another blocking category, then it is a barrier. This is the only way that **IP** can become a barrier. Clause ii defines inherent barrierhood.

Essentially, all blocking categories except for **IP** are barriers.

The locality condition on government, then, is that government can only occur if there are no intervening barriers for the category being governed.

1.2.7. Case Theory

Case theory provides a partial characterization of well-formedness at s-structure, providing a constraint on the distribution of lexical **N″**s. Within GB, a feature-based account of lexical categories provides a grouping of certain [−N] categories (verbs and prepositions). In certain structurally defined contexts these [−N] categories, together with the non-lexical Infl, assign a property referred to as *case* to another constituent. Case, then, is an abstract property of constituents and corresponds only indirectly to more traditional uses of the term. The relevant structural environments for case assignment are government, adjacency[6], and directionality. Directionality may be variable within a given language. For example, if subject **N″**s in English receive case from **Infl**, we require case assignment to the left, yet object **N″**s receive case from the verb so case assignment must also occur to the right. Directionality may need to be specified on the particular case assigning category in order to account for such variation. At s-structure, all overt[7] **N″**s must have case (the *Case Filter*).

There remain some problems for case theory that, at present, do not have a generally accepted solution. For example, dative verbs take two objects, yet only one of these is adjacent to the case assigner (the verb). Solutions for this include lexically marking such verbs as carrying an extra case assigning feature (ie. marking the verbs as exceptional to case theory). Other solutions propose internal structure within **V″** so that the two objects can both receive case. Proposals of this sort include the possibility that the second object receives case from **V′** in a structure such as that presented in (17).

(17)
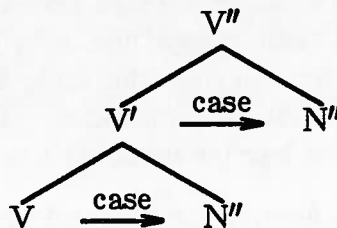
Other proposals maintain that reanalysis takes place giving the verb and first (indirect) object the status of a lexical verb that then assigns case to the second (direct) object (Stowell, 1981). There is currently no general agreement on this and other related problems for case theory as presently stated.

---

[6] It has been proposed that adjacency for case assignment is parametric and may not constrain case assignment in some languages (Stowell, 1981).

[7] Empty categories can occur at s-structure without case.

1.2.8. Theta Theory

In current research within the GB framework, it is generally agreed that the lexical entries of verbs (and probably other categories of predicates as well) contain a semantic-like information template known as the θ-grid. This grid provides information about the roles that various verbal arguments bear to the verb in the event being described. The θ-grid is generally viewed as a list of thematic roles, possibly annotated with information about grammatical realizations of the elements bearing the particular roles. The θ-roles are themselves collections of thematic relations, where thematic relations are drawn from a presumably finite and universal inventory containing such things as *agent, locative, temporal, source, goal, theme,* etc.

The definitions and terminology used here were initially developed by Gruber (1965, 1976) and Fillmore (1968) and were popularized in a somewhat modified sense in mainstream generative linguistics by Jackendoff (1972).

Jackendoff demonstrated that, contrary to earlier beliefs, there is no one-to-one correspondence between syntactic positions and thematic relations. In fact, he showed that a single syntactic position can bear a variety of thematic relations simultaneously.

One of the key insights of Jackendoff's work is the observation that θ-roles consist of a set of thematic relations and not just a single one. Consider the subject noun phrases in each of the following sentences:

(18) I kept the book.
(19) I lost my mind.
(20) I gave the book to Sue.

In sentence (18) the subject N″ bears the thematic relationships of *agent* and *locative* with respect to the predicate, while in sentence (19) the relationships are those of *source* and *experiencer*, and for sentence (20) of *agent* and *source*. If a θ-role consists of a single thematic relation, then the sentences described above prove problematic. A θ-role, then, is a package of thematic relations that are carried by a single element. A θ-role contains at least one thematic relation, and, although not generally made explicit, we can assume that if a θ-role contains more than one such relation, a mechanism of internal consistency would prevent the grouping together of mutually inconsistent thematic relations into a single θ-role.

Most current discussions of θ-theory ignore the original observation that a given element can receive a set of thematic relations. In fact, as far as syntax is concerned, the use of θ-theory has been limited to access to the presence or absence of a θ-role on an element with generally no access to its contents. This is typified by the central condition of θ-theory, the θ-*Criterion*, which is defined in (21).

(21) θ-*Criterion*: Each argument bears one and only one θ-role and each θ-role is assigned to one and only one argument. (Chomsky, 1981, p.36)

·This principle of grammar forces the argument structure of the syntactic representations to be a projection of the lexical properties of the θ-grid. The θ-Criterion is assumed to apply at every level of representation (d-structure, s-structure, and LF) constraining admissible sentence types as well as limiting the move-α relation between d-structure and s-structure.

The assignment of θ-roles to d-structure argument positions involves two mechanisms: direct and indirect assignment. A constituent is said to be θ-*marked*, either directly or indirectly, if it has been assigned a θ-role. Internal verb arguments (ie. objects within V') are assigned θ-roles directly from the verb. External arguments (ie. subjects) are assigned θ-roles indirectly from a percolated grid at the level of V''. At s-structure, θ-roles are borne by θ-chains (Safir, 1985) that correspond to the d-structure argument position through the move-α mapping.

Although very little work has been done to incorporate oblique arguments into this schema, it is generally agreed that with **PPs**, for example, the preposition assigns case and directly assigns a θ-role to its N''. If the verb is involved at all, it only indirectly assigns a θ-role to the entire **PP**. This satisfies the θ-Criterion (which is only concerned that all N''s receive a θ-role) as well as satisfying the Case Filter.

2. An Overview of Warlpiri Syntax

Warlpiri, a central Australian aboriginal language, poses an interesting challenge to any analysis of language that relies crucially on structural configuration. Such so-called 'non-configurational' languages may exhibit a number of properties including rich inflectional morphological systems of case and agreement, and free word order with discontinuous expressions.

In this section I provide an overview of Warlpiri grammar with emphasis on those aspects of morphology and syntax that have set Warlpiri apart in the study of languages.

2.1. Lexical Categories

Warlpiri has a rich case marking system, marking grammatical cases (*eg.* ergative, dative) and semantic cases (*eg.* locative) with suffixes. Dependent infinitival clauses and nominal arguments can both be marked with these suffixes.

The term "nominal" as used in the Warlpiri literature, corresponds to nouns, adjectives, and some verbs (*eg.* want, know). It corresponds only roughly, then, to the English category of noun. With the exception of the nominal 'verbs', the Warlpiri category of verb corresponds to the standard notion of verb.

Clauses can occur in basically two types, tensed and infinitival. Infinitival clauses are suffixed by an infinitival marker, while tensed clauses have a auxiliary. Infinitival clauses also have a complementizer suffixed to the infinitival

marker, while tensed clauses may have a complementizer prefixed to the auxiliary.

There are two ways that pronominals are expressed in this language. One way is through overt pronouns, the other way involves omitting the argument altogether and letting the agreement (person and number markers) on the auxiliary provide the necessary information.

In addition to the major categories discussed above, there are also preverbs (a category that prefixes to verbs, creating complex verbs), modals, and conjunctions.

2.2. Morphology

The mechanisms of Warlpiri morphology are quite straightforward. Vowel harmony is the main active phonological process.

The general process of affixation involves the concatenation of suffixes to the root word. Cliticization is also common in Warlpiri. For example, certain auxiliary words can occur as separate words in their own right, while others require (or at least allow) themselves to be attached (much like a suffix) to the preceding word.

2.3. Ergative/Absolutive Case Systems

Warlpiri has an ergative/absolutive case system. This contrasts with the common nominative/accusative case systems such as that of English in the following way: In a nominative/accusative system, the subject is treated uniformly (as nominative) regardless of the transitivity of the verb. The object (accusative) of transitive verbs also has a uniform treatment in the grammar. In an ergative/absolutive case system, the subject of a transitive verb is ergative, while the object of transitive verbs and the subject of intransitive verbs pattern together and are absolutive.

In Warlpiri the ergative/absolutive distinction is made, with the absolutive case having no overt phonological realization. The Warlpiri auxiliary, however, exhibits a nominative/accusative distinction in the person and number markers that occur on it. Such a situation is apparently fairly common and is known as Split Ergativity.

2.4. Word Order and Constituency

The most outstanding aspect of Warlpiri's syntax is the apparent absence of any significant role for word order. With the exception that AUX usually occurs in second position in the sentence, any linear ordering of the content words in a sentence is permitted. All of (22) through (27) are equivalent in meaning with possibly only a slight difference in focus. (Hale 1981, p.1)

(22)

kurdu-ngku	ka	maliki	wajilipi-nyi
child-ERG	(AUX)pres	dog-(ABS)[8]	chase-NONPAST

'The child is chasing the dog'

(23)

kurdu-ngku	ka	wajilipi-nyi	maliki
child-ERG	(AUX)pres	chase-NONPAST	dog-(ABS)

(24)

wajilipi-nyi	ka	maliki	kurdu-ngku
chase-NONPAST	(AUX)pres	dog-(ABS)	child-ERG

(25)

wajilipi-nyi	ka	kurdu-ngku	maliki
chase-NONPAST	(AUX)pres	child-ERG	dog-(ABS)

(26)

maliki	ka	kurdu-ngku	wajilipi-nyi
dog-(ABS)	(AUX)pres	child-ERG	chase-NONPAST

(27)

maliki	ka	wajilipi-nyi	kurdu-ngku
dog-(ABS)	(AUX)pres	chase-NONPAST	child-ERG

In addition to the free word order, Warlpiri allows discontinuous expressions. As would be expected, the rich morphological system of Warlpiri aids substantially in the identification and interpretation of discontinuous expressions.

(28)

kurdu	wita-jarra-rlu	ka-pala	maliki
child	small-dual-ERG	(AUX)pres-dual(subj)	dog-(ABS)

wajilipi-nyi
chase-NONPAST

'The (two) small children are chasing the dog.'
(Hale, 1981, p. 1 and p. 5)

---

[8] Notice that the absolutive case is unmarked (*ie.* phonologically absent) in Warlpiri.

(29)

kurdu-jarra-rlu   ka-pala   maliki   wajilipi-nyi
child-dual-ERG   (AUX)pres-dual(subj)   dog-(ABS)   chase-NONPAST

wita-jarra-rlu
small-dual-ERG

'*The (two) small children are chasing the dog.*'

(30)

wawirri   yalumpu   kapi-rna   panti-rni
kangaroo   that-(ABS)   (AUX)1sg(subj)   spear-NONPAST

'*I will spear that kangaroo.*'
(Hale, 1983, p. 2)

(31)

wawirri   kapi-rna   panti-rni   yalumpu
kangaroo-(ABS)   (AUX)1sg(subj)   spear-NONPAST   that-(ABS)

'*I will spear that kangaroo.*'

In the discontinuous examples (29) and (31), a prominent interpretation is that the two constituents that agree in markings (*kurdu-jarra-rlu* and *wita-jarra-rlu*; *wawirri* and *jalumpu*) are interpreted as a single constituent. These constituents can occur together in the pre-auxiliary position with the last word being the only one to receive the case and number markings (see examples (28) and (30). Since the auxiliary is found in second position, it is clear that, when contiguous at the beginning of a sentence, the two words, together with markings, form a single constituent.

As mentioned earlier, there are two types of clauses in Warlpiri. We have seen tensed clauses with the auxiliary occurring in second position. Dependent tensed clauses also occur and are found in *adjunction structures* with the main clause. This is illustrated by the examples in (32) and (33).

(32)

njuntulu-lu   katja-npa   wawiri   pantu-nu,
you-ERG   COMP-AUX   kangaroo-(ABS)   spear-PAST,

natjulu-lu   kapi-na   pura-mi[9]
I-ERG   AUX   cook-NONPAST

'*I will cook the kangaroo you speared.*'

---

[9] The notation used here is the notation of Hale, 1976. In later articles Hale uses different transcription conventions. I will adhere to the transcription conventions used in the articles from which the examples are taken.

(33)

    natjulu-lu   lpa-na   kali   tjantu-nu,   kutja-$\emptyset$[10]-npa
     I-ERG   AUX   boomerang trim-NONPAST,   COMP-AUX

       ya-nu-nu   njuntu
    walk-past-hither you-(ABS)

'I was trimming a boomerang when you came up.'
(Hale, 1976, p.79)

The comma indicates the edge of the rising-falling intonation that marks the division between the two clauses. As is illustrated in these examples, the dependent clause can occur on either side of the main clause.

Warlpiri also has dependent infinitival clauses. These are marked by an infinitival suffix followed by a complementizer suffix. The free word order of Warlpiri is not quite so free in infinitival clauses. These are usually quite short, with few syntactic arguments. The verb must occur after all of the arguments of the infinitival clause and immediately before the infinitival suffix. This is illustrated in example (34).

(34)

      karli   jarnti-rninja-rlajinta $\emptyset$-rna-ju
    boomerang   trim-INF-COMP(reflex)   (AUX)perf-1sg(subj)-1sg(obj)

    paju-rnu
    cut-PAST

'I cut myself while trimming the boomerang.' (Hale, 1981, p.5)

Infinitival clauses can be discontinuous. This occurs when the arguments and the verb form separate constituents, and it is characterized by both the argument and the verb+infinitive marker having the complementizer suffixed to them. This is illustrated in example (35).

(35)

      karli-ngkajinta $\emptyset$-rna-ju   paju-rnu
    boomerang-COMP(reflex) (AUX)perf-1sg(subj)-1sg(obj)   cut-PAST

    jarnti-rninja-rlajinta
    trim-INF-COMP(reflex)

'I cut myself while trimming the boomerang.' (Hale, 1981, p.2)

---

[10] The symbol $\emptyset$ is used to indicate a null auxiliary base.

The choice of complementizer for an infinitival clause determines the reference of the omitted argument (the subject). The reference of the subject is selected from the main clause arguments. In the example above, the reflexive complementizer selects the reflexive subject of the main clause as the controller of the subject of the infinitival clause.

2.5. Previous Approaches to Warlpiri Syntax

In the early days of generative grammar, focus was primarily on widely studied European languages. As the theory attempted to balance descriptive adequacy with explanatory adequacy, claims emerged about universal properties of human language. It is not surprising that early biases from European languages found their way into the proposed principles of current linguistic theory.

Many of these assumptions about Universal Grammar have been seriously called into question as evidence from a wider range of languages have come to bear on linguistic theory. Some properties of Language that seemed to be excluded (in principle) from syntactic theory were such properties as total freedom of word order, and discontinuity of semantic expression. These properties, we have seen, are exhibited by Warlpiri.

An early proposal by Hale (1967) to account for these apparently exceptional properties involved a context-free phrase structure grammar that created an arbitrary ordering which was, in turn, "scrambled" by local transformations (possibly of the equivalent of the PF component). Such a view is a natural extension of early Transformational Grammar (TG), however it merely makes an observationally adequate statement that languages that allow scrambling have free word order. The first attempt to address the properties of languages like Warlpiri within a more current framework was also proposed by Ken Hale (1981).

2.5.1. The W* Approach

Hale (1981) contained the first proposal to characterize languages with the apparently exceptional properties of discontinuity and free order by means of a parametrization of Universal Grammar (UG). In this work, the locus of variation was claimed to be the base component. The claim is that there are two types of base components: X-bar and W*. An X-bar type base component yields the familiar hierarchically structured syntactic trees with relatively fixed order and configuration. On the other hand, a W* base component produces a concatenation of words drawn (in any order) from the word formation component. The output of the W* base component is a string of words with no internal structure and no requisite order or configurational constituency — hence the term non-configurational. Since there is no constituency, there are no syntactically defined positions (*eg.* subject, object, specifier, complement) and also no empty categories or transformational rules.

A W* base component produces merely a concatenation of words. Semantic interpretation would require more information than the output of this component would provide. Hale introduces "a system of PARSING PRINCIPLES which determine the constituency and category of expressions present in a given concatenation of words. In effect, these parsing principles impose a labelled bracketing upon the strings of words." (Hale, 1981, p. 16)

Since labelled bracketings are isomorphic to configurational tree structures, such parsing principles mimic the descriptive function of the X-bar type rules, providing the syntax with the constituent structure that it requires but that the W* base component fails to provide. So, for example, infinitival clauses are claimed to be described by the W* schema, with no set word order or constituency. A strategy of *infinitival bracketing* (presented in (36)) would group together the infinitival clause. This strategy refers to the order of elements and hence, in a fashion, imposes an order on the elements if they are to be bracketed together.

(36) *Infinitival Bracketing*: Bracket together with the infinitival verb any immediately preceding contiguous string of words.

Since a W* base component imposes no ordering or constituency on a sentence, to the extent that there is word order or constituent structure, these must follow from other principles. Since they are not captured in the base component (where statements of constituency and order are traditionally made) they have to be stated in this new component of grammar that is only relevant for W* languages.

Consider the consequences of this analysis from an acquisition point of view. Such parsing principles mimic the descriptive function of the X-bar type rules. For a language like Warlpiri, where virtually all permutations of a sentence count as semantic repetitions, the required parsing principles are minimal (especially if we assume that all derivation and inflection is accomplished independently of the syntactic component). So, it seems that the W* theory accounts for the intuitively real differences between, say, Warlpiri and English. However, given that parsing principles form part of a W* grammar, there is no reason, in principle, to prevent all languages from being successfully analyzed as W* with potentially elaborate parsing routines that duplicate the X-bar rules. If we claim the X-bar/W* distinction as a parameter of UG, then we must be able to provide a triggering context to decide between them. However, it seems that whatever would trigger the parsing strategies given a W* analysis should have been sufficient to trigger an X-bar base component.

In addition to the parsing strategies, a punctuation component (essentially, the PF level of representation discussed earlier) is used to capture some ordering facts in Warlpiri; in particular, the second position of the auxiliary is captured through well-formedness conditions of the punctuation

component.

The distinction between X-bar and W* base components creates a typological distinction between *configurational* and *non-configurational* languages. Accepting this, somehow languages choose to differ in this fundamental component of their syntax. But, as Hale points out, W* languages exhibit some properties that are characteristic of X-bar type languages. As a result, parsing strategies must be incorporated into a W* grammar, and these parsing strategies mimic the X-bar type base component rules. Given the need for these X-bar style parsing strategies, the typological distinction between X-bar and W* is not nearly so clearly defined. In fact, it is not at all clear that we have a strong typological distinction between these language types at all.

This is an objection that Hale realized and in the postscript to Hale (1981)[11] he retreated somewhat from the extreme W* approach described above. In the postscript, Hale proposes a minimal X-bar type phrase structure component for languages like Warlpiri. The maximal projection for such a language would be a single bar-level (*ie.* $X^{MAX} = X'$), where the nodes are not category-specific. In this way the nodes become tree labels that contain no category information, rather than variables that range over categories. In this theory of non-configurationality, the base component would have one rule, namely:

$$X' \rightarrow X'^* \, X$$

Lexical insertion occurs freely into the phrase structures defined by this rule, thereby accounting for free order without requiring scrambling rules. Since the syntactic nodes are unspecified for category, they become instantiated as category labels when lexical insertion occurs. For this reason, there can be no empty categories in the phrase structure of non-configurational languages that are defined by rules such as the one above.

In order to account for the position of the auxiliary as well as the free position of V (which presumably is projected to S), Hale proposes a rule that applies only to tensed sentences.

$$S \rightarrow AUX \, X'^* \, V \, X'^*$$

The aux-second phenomena continue to be captured through a local movement rule in the punctuation component.

David Nash (1980) adopts a somewhat revised version of this proposal. Nash proposes a single phrase structure rule for Warlpiri, namely

$$X \rightarrow X^*$$

where the X is not a category label as it is in standard X-bar theory, but, rather, is totally unspecified for category. This rule allows recursion and, hence, intermediate constituency, but there is no inherent X-bar notion of headedness (*ie.* the rule is *not* $X' \rightarrow X^*$). Such a base component allows virtually any well-

---

[11] The paper (less the postscript) was written in 1979.

formed tree. The role of Hale's parsing principles is taken over, in Nash's model, by rules assigning categorial labels to the tree nodes.

In a 1982 paper, Ken Hale once again attributes the parameter to the base component, but, rather than propose a different type of base component, Hale proposes to parameterize within the accepted principles of X-bar theory. The standard X-bar schema has two parts:

$$1) X'' \rightarrow ... X' ...$$
$$2) X' \rightarrow ... X ...$$

The proposed parameter, then, is that configurational languages use both 1 and 2, while non-configurational languages use only 2. The distinction is between relatively flat structures (non-configurational) and relatively hierarchical structures (configurational). In this paper, Hale also recognizes that free word order is not the only issue involved in characterizing non-configurational languages. There are properties that tend to cluster on non-configurational languages. These properties include:

    i)   free word order
    ii) discontinuity
    iii) pronoun drop
    iv) lack of NP movement
    v)   lack of pleonastics
    vi) rich case system
    vii) complex verb or verb/AUX system

In Hale(1982), these properties are claimed to be derivative of the flat structure produced by the base component. The claim is that government is parameterized conditional to the selection of base components (*ie.* government is only relevant in two-bar languages). If government is defined as a relation between a head and its sisters, then, in a two-bar language, there are two levels of government. In a one-bar language, government cannot distinguish between the various sisters of the head, and, hence, fails to apply. In such a system, some of the above properties are derivative.

A synthesis of the approaches of Hale (1982) and Nash (1980) is developed by Farmer (1984) who proposes that Japanese has a phrase structure component of the following form:

$$X' \rightarrow X'^* X$$
$$X \rightarrow$$

(*ie.* context free lexical insertion)

These phrase structure rules create an unlabelled tree with head information, but no category information. Lexical insertion instantiates the category of X which percolates features to X'. Since lexical insertion is claimed to be context free, and since the trees do not receive features until lexical insertion has taken place, free constituent order is predicted in this model without the use of anything such as scrambling rules.

26

A different approach to characterizing non-configurationality, yet one that is still in keeping with Hale (1982), is found in Bouchard (1984). Bouchard proposes a similar flat phrase structure for non-configurational languages, but instead of opting for government not to apply (as in Hale (1982)), Bouchard maintains that government does apply, only ambiguously. So in a structure like:

$$[_{V'} \; NP_1 \; NP_2 \; V \;]$$

both NP_1 and NP_2 are governed by V. The richness of the verbal inflection in non-configurational languages leads Bouchard to propose that V is actually V+Infl in the above. This means that both NP_1 and NP_2 are governed by Infl as well, hence allowing either to be the subject or object. Bouchard claims "that the basic difference between configurational and non-configurational languages is whether the structure is ambiguous or not" (p.157). The degree of flatness can vary from language to language, creating varying degrees of ambiguity, hence gradient non-configurationality. In the extreme case, we find wildly ambiguous structures typical of Warlpiri discontinuity.

2.5.2. Hale's Configurationality Parameter

In (Hale, 1983), Hale proposes a parameter stated in the theory of grammar that is claimed to account for the typological distinction between configurational and non-configurational languages.

In this paper, Hale maintains a minimal X-bar type schema with categorially vacuous nodes, yet proposes a wider-reaching configurationality parameter in terms of the Projection Principle (Chomsky, 1981). Hale distinguishes between two levels of structure: lexical structure (LS) and phrase structure (PS). LS is an abstract structure that Hale maintains is universally configurational and exhibits such configurational properties as subject/object asymmetries. Since this level of structure is universally configurational, languages that are non-configurational differ in the mapping between LS and PS. The proposed difference between configurational and non-configurational languages is the extent to which the universal LS configuration is visible in PS.

In order to formalize this distinction, Hale proposes the *Configurationality Parameter*, presented in (37).

(37) *Configurationality Parameter*:
        a) In configurational languages, the projection principle holds of the pair (LS, PS).
        b) In non-configurational languages, the projection principle holds of LS alone. (Hale, 1983, p. 26)

The Projection Principle requires that all levels of representation be projected from the lexicon and observe the subcategorization properties of lexical items. Given this, the Projection Principle must hold within LS. The proposed Configurationality Parameter amounts to parameterizing the domain of application of the Projection Principle. If part a) is satisfied, we have a bijective

mapping defined between LS and PS, and, hence, LS properties are transparently mirrored in the phrase structure. If PS is not tied to LS by the Projection Principle, then LS may be a highly abstract level of structure that need not correspond to the actual phrasal organization of the sentence. Any connection between LS and PS must follow from other principles of grammar — such a connection cannot be arbitrary or else semantic interpretation would be impossible. The proposed connections are *linking rules* that identify cased arguments, and a principle of coherence. These are the principles responsible for guaranteeing a well-formed PS given the potential for overgeneration that the Configurationality Parameter introduces.

From the Configurationality Parameter, it follows that non-configurational languages do not require empty categories at PS. Since the Projection Principle does not enforce the requirement that nominals fill the subcategorized grammatical relations, null anaphora can be accounted for without recourse to empty categories. Also, since there is no requirement that particular grammatical relations bear a particular structural position (this is required universally at LS, but not at PS in non-configurational languages), free word order and discontinuity are derivable.

As an example of an LS from Warlpiri, (38) shows the LS of the Warlpiri verb *panti-rni* ('*pierce*').

(38)

$$[_{V'} \; erg, [_{V} \; abs, \; panti\text{-}rni]]$$

(Hale, 1983, p.23)

The ergative and absolutive arguments of sentences with the verb *panti-rni* need not correspond to this order or constituency. Assuming exactly one ergative and one absolutive constituent in a tensed clause with *panti-rni*, all 6 possible orderings of the two arguments and the verb are possible. Hence all 6 possible PS structures would have the same LS structure as in (38), with only one of the six corresponding in linear order to this LS.

In many ways, this solution is analogous to the W^* solution, only shifted to another principle. With the W^* analysis, the sacrifice in giving up the X-bar base was too great and forced S-bar style parsing strategies to be incorporated into the model. Here we sacrifice the Projection Principle, but lose descriptive adequacy, and are forced to introduce linking and coherence constraints. These rules guarantee that the PS not contradict the lexical subcategorizations, and determine the appropriate linkings of arguments in LS to overt nominals. This seems, once again, to mimic at least part of the role that the Projection Principle normally plays. Again, we find that the one step parameter of fiat is too powerful and forces the introduction of separate mechanisms to regain some of the descriptive adequacy that was lost. Of course, by being a separate

mechanism, even though we regain descriptive adequacy, it is at the expense of explanatory adequacy.

In a similar vein to Hale (1983), Marantz (1984) maintains that a language is configurational "to the extent that it maps S-Structure relations onto the surface structural relations of structural government and adjacency" (Marantz, 1984, p.88). In Marantz' model, S-Structure is a non-interpreted level between the phonologically interpreted surface-structure and the semantically interpreted logico-semantic structure. Essentially, Marantz proposes that non-configurational languages are characterized by a relaxing of the mapping principles between levels of representation.

In a reply to Hale (1983), Jelinek (1984) argues against the Configurationality Parameter. She maintains that Hale's proposal "threatens to permit languages with uninterpretable surface structures" (Jelinek, 1984, p. 43), and generally argues that it predicts that languages could be too different from one another. She proposes that the real verbal arguments in a language like Warlpiri are the clitic pronouns, and that nominals are actually optional *adjuncts* to these clitic pronouns. In Warlpiri, clitic pronouns are prefixed to the auxiliary, are obligatory, and have fixed ordering — requirments that are reminiscent of standard argument requirements in configurational languages. The free order and discontinuity of nominal phrases can be derived from their adjunct status.

In chapters 2 and 3 I will outline a syntactic analysis of Warlpiri that is more consistent with standard GB theory than previous analyses of Warlpiri. In chapter 4 I will show how this theory can be translated into a processing model for Warlpiri syntax and discuss the organization and design of an implemented Warlpiri parser.

3. Parsing

3.1. GB and Parsing

Initially, the choice of the Government-Binding framework for a study of language processing issues would seem an unlikely one. The general philosophy of overgeneration (even random generation) of sentences followed by subsequent filtering out of those that are ungrammatical stands in blatant contradiction to commonly accepted theories of parsing where the constraints and conditions apply during the generation of the structure and not as a success measure after the parse is completed. The movement within GB away from phrase structure rewrite rules and derivation has left many language processing enthusiasts cold.

Nevertheless, the research strategy adopted by those working within the GB framework has proven to be quite successful. By isolating the method of language generation from the study of syntactic phenomena, a clearer picture of the subsystems that are operating in language has emerged. Governing principles that initially were conceived of as constraints on overgenerated structures

have proven to be essential properties of parsing mechanisms. As evidence of this, consider the claim by Marcus (1980) that such locality constraints as subjacency, developed within the early GB framework, proved essential to the parsing formalism that he developed. In fact, given the structure of the Marcus Parser, it would only be with a considerable degree of difficulty that subjacency could be violated. These results have been further developed in subsequent work by Berwick and Weinberg (1984). The possibility of a totally principle-based parsing mechanism is the focus of this research, as well as others' (Wehrli, 1984; Abney and Cole, 1985).

The task of rendering the principles of GB theory into a parsing mechanism involves the transformation of the constraints and filters of the theory into procedures defined to construct structures that would have conformed to the constraints as originally stated. This is clearly a difficult task, yet one that I am convinced is achievable.

3.2. Determinism and Parsing: The Marcus Parser

There are many different approaches to syntax evidenced in the various parsing algorithms that are in common use. From a syntactic theory point of view, few, if any, of the currently accepted parsing algorithms are satisfactory. One of the more promising types of parsers from the point of view of syntactic theory is the Marcus Parser.

In his 1977 dissertation and 1980 book, Marcus described a working parser, PARSIFAL, that was designed to explore the linguistic and psycholinguistic principles that underlie language processing.

Central to the parsing mechanism that Marcus developed is the *determinism hypothesis*. This hypothesis states that it is possible to parse the syntax of a language without simulating a non-deterministic device through backtracking or simulated parallelism.

Marcus ensured that his parser would adhere to the determinism hypothesis by enforcing the properties in (39) in the implementation of the parser.

(39)

      a) All syntactic substructures created by the grammar are permanent.

      b) All syntactic substructures are incorporated into the final parse.

      c) No temporary syntactic structures can be hidden in the internal state of the processing mechanism.

In order to accomplish this, the parser must be assured that it is making the correct choice at each stage of the parse.

However, there are many sentences in language that seem to require some degree of non-determinism in order to arrive at a successful parse. Consider the following sentences:

      (40) Have the students who missed the exam take the makeup today.

(41) Have the students who missed the exam taken the makeup today?
(Marcus, 1980, p.2)

If determinism is maintained, how does the parser know how to distinguish between these two sentences in the initial stages of the parse. Sentence (40) is a command with the main verb 'have', while sentence (41) is a question with an auxiliary verb 'have'. The crucial information that decides this issue, as far as the written string is concerned[12], occurs after the parser would presumably have already processed the initial words of the sentence.

To handle such situations as this while maintaining determinism, Marcus included a look-ahead buffer in his parser. This buffer is of a small size (usually with only three positions), and the buffer positions are used as a work space in the construction of the parse. Incoming morphological units occupy buffer positions, as do already completed substructures that are waiting for a position in the parse tree. The buffer essentially becomes a sliding window across the sentence giving access to later words of the sentence as the earlier words are processed into substructures or incorporated into the parse tree.

3.3. Parsing Warlpiri

If determinism is a criterion of parsing strategies, Warlpiri provides an interesting challenge. If we accept the Marcus Parser strategy of implementing determinism, the parsing of discontinuous expressions proves to be quite a difficult task. Since there are no apparent locality conditions on discontinuous expressions in Warlpiri, the parser will never be able to close off a constituent since it will never be sure whether there is a discontinuous piece of the constituent beyond the view of the buffer. Essentially the buffer's view will never extend beyond the third piece of structure of the Warlpiri sentence, since to close off a constituent would mean that any subsequent discontinuous pieces of that constituent would not be able to be processed without violating determinism and reopening the already closed constituent.

Since I agree that determinism should be a goal of parsing routines, I will provide an implementation that, although based somewhat on the Marcus Parser, is in many ways considerably different. I provide a implementation that involves division of parse substructures according to governing principles of GB as well as the theory of Warlpiri grammar that will be developed in chapters 2 and 3. These governing principles of grammar, as well as the requirement of determinism, will control the passing of sub-structures through the parse to their position in the output tree. Discontinuities are linked by predicate argument structures that are constructed in parallel with the parse tree. Hence, we have discontinuous syntactic expressions united in a predication structure. The parse is expectation-driven as well as data-driven, with selectional expectations of lexical items as well as structural expectations determining the appropriate flow of control among sub-structures in the parsing routine.

3.3.1. Kashket (1986)

In a paper presented at the 1986 ACL meeting, Michael Kashket presented an approach for parsing Warlpiri. His approach was based on Hale's 1983 analysis and GB Theory, yet determinism was not a goal of this implementation since multiple passes of the sentence are allowed during the parse. In keeping with Hale's proposal of a universal configuration to the LS structure, Kashket takes the free-order string of words that constitute a Warlpiri sentence and associates each argument with the LS structural position of that constituent. In that way, the order of the input sequence is irrelevant to the output of the parse.

A directed process of *selection* is used to account for fixed morpheme order in Warlpiri, while non-directional processes of argument linking allow for the free word order in the language.

The range of Warlpiri sentence types covered in the presented parser are quite limited. Essentially, the sentences are restricted to simple sentences with simple arguments (*ie.* no embedded clauses or complex nominal arguments). The parser does not handle the auxiliary word in Warlpiri, so it is omitted from the input sequence. All of the subcategorized arguments of the predicate must be overtly present in the input string, and hence the range of sentences with non-overt arguments are not dealt with. In addition, no discontinuous expressions are handled in the presented version of the parser[13].

---

[12] Intonational information provides disambiguating cues in the spoken form of these sentences.

[13] Extensions to accommodate discontinuities are the current focus of a revised version of the parser (M. Kashket, p.c.).

CHAPTER 2

Configurational and Non-Configurational Languages

1. GR-Configurationality

The term 'non-configurational' is often used but seldom clearly defined. At times it is equated with a W^* base component (Hale 1981) wherein non-configurationality indicates an almost total absence of hierarchical structure in the syntax (see chapter 1, section 2.5.1). A somewhat less extreme view has non-configurationality equated with an X-bar type base component (Hale 1983) in which maximal projections are only a single bar-level. Another common use of the term is to describe the situation in which grammatical relations (*eg.* subject, object) cannot be identified from the hierarchical structure of a clause (Farmer 1984). In the ensuing discussion I will adopt the definitions proposed in (1) and (2).

(1) *S-Configurational*: The presence of hierarchical structure in the syntax.

(2) *GR-Configurational*: *G*rammatical *R*elations are defined in terms of hierarchical structure.

Notice that if a language is GR-Configurational then it is also S-Configurational, but the converse need not be true. A language may be S-Configurational yet GR-NonConfigurational (*ie.* not GR-Configurational). If we accept tree diagrams as a valid notation for representing syntactic relations, then at least a minimal degree of S-Configurationality is necessary since it is forced by the notation.

Preliminary to developing and exploring an approach to GR-Configurationality, let us consider the basic concept of the θ-grid. A θ-grid is a list of θ-roles which is associated with a predicate in the lexicon. θ-roles are sets that include such thematic relations as *agent, patient, instrument,* etc. These relations abbreviate the actual roles of participants in the predicate event (*ie.* the event described by the predicate). For example, the verb **put** in English has a θ-grid as in (3).

(3)

$$\textbf{put: } [\ \{agent\}, \{theme\}, \{locative\} \ ]$$

This θ-grid carries the information that, for the verb **put**, there are three principal and obligatory participants. The terms *agent, theme* and *locative* serve as indices into a conceptual structure. At the same time the thematic relations

comprise a useful level of abstraction. Generalizations can be captured by this level of abstraction that would be obscured by a theory that included all the details of the specific and idiosyncratic roles of the participants in a predicate event. Hence, thematic relations are abstractions of roles and relationships and they function as indices associating syntactic arguments with the semantics of the predicate event.

The three principal participants, reflected in the three-position θ-grid for **put**, are mapped onto three arguments in the syntactic representation. If we view the thematic relations as abbreviated indices into the semantics of the predicate, then, since we must be able to recover the semantics of an utterance, we must be able to recover these indices. This suggests that we need rules of the sort presented in (4).

(4) For X, a syntactic argument and α, a thematic relation: X bears the θ-role containing α if and only if X is identified by $\mathbf{R}^{\alpha}$, where $\mathbf{R}^{\alpha}$ is some relation identifiable in the syntax.[1]

This says that if X is linked to a particular θ-role, there must be some way of identifying that θ-role in the syntax. Since $\mathbf{R}^{\alpha}$ must be identifiable in the syntax, it must be stated in terms of relations evidenced in the syntax. Syntactic structures convey information about lexical items (leaf nodes), hierarchical constituency, and linear order. $\mathbf{R}^{\alpha}$ must therefore be conveyed by one or more of these syntactic information types. GR-Configurational languages express the relation $\mathbf{R}^{\alpha}$ in terms of hierarchical constituency, and I will refer to such relations as *local GR-Configurational relations*. The term *local* is used to emphasize that this is not a global property of a language, but a property local to the thematic identification of a particular argument by a predicate.

Consider the sentence in (5).

(5) The man put the box on the table.

The θ-role {*agent*} is identified with the subject grammatical relation. This is identified in a local GR-Configurational manner as the specifier of **IP**, a position that is structurally identifiable. The θ-role {*theme*}, the object grammatical relation, is identified in a local GR-configurational manner as the complement of the verb phrase[2]. The θ-role {*locative*} is identified in a local GR-NonConfigurational manner by means of the overt preposition.

In (4) it was implicitly claimed that there are rules associating a particular thematic relation with a particular syntactically defined relation. The work of

---

[1] As examples of this, the structural relationship *subject of the verb* is a type of R relation, as is a *nominative* suffix. These can both occur simultaneously on a single argument, as occurs in languages that have both structural subjects and nominative markers.

[2] Linear order as well as hierarchical constituency may figure in the GR-Configurational phenomena illustrated here. The relative linear sequencing of the arguments in the syntax is generally assumed to follow from headedness properties of the syntax. It remains unclear how ordering facts about direct and indirect objects can be accounted for with only reference to headedness properties and no reference to linear sequencing.

Jackendoff (1972) and Gruber (1965) has indicated that there is no such one-to-one correspondence between a particular thematic relation and a particular instance of **R**.

Two possibilities present themselves to avoid this objection. One possibility is to associate a set of thematic relations with each **R** relation. This set would contain all the thematic relations that can be represented by that instance of **R**. The members of each set could have associated with them a markedness value. This value would be used to predict the least marked thematic relation from a lexical θ-grid of a predicate to be represented by the **R** relation. In this way, the least marked mapping of **R** relations for a particular θ-grid can be selected for the syntactic arguments (Brunson, 1986).

Another possibility would be to structure the θ-grid itself in a manner consistent with the **R** relations. For example, the θ-grid for **put** might be represented as in (6).

(6)

$$\text{put:} \, [\, \{agent\} \, [\, \text{put} \, \{theme\} \,] \, P\text{-}locative \,]^3$$

In this representation, the θ-roles $\{agent\}$ and $\{theme\}$ are identified by a structuring of the θ-grid and the $\{locative\}$ is identified by 'P-locative' (*ie.* a locative preposition), so we have lexical identification as well. In this analysis, rather than **R** constraining the mapping of the θ-grid to the syntax, **R** is actually realized on the θ-grid itself. The mapping to the syntax would then involve preservation of dominance relations as well as constraints imposed on the selection of lexical items. I know of no empirical evidence that might help to decide between these options.

The three possible forms of expression of the **R** relations yield the three general strategies for linking arguments to θ-grids shown in (7).

(7)   i) structural/hierarchical configuration (local GR-Configuration)
      ii) linear sequencing (template matching)[4]
      iii) overt markings (case frames)

These strategies may be combined in a language. In English, we have hierarchical configuration yielding subject/object positional asymmetry, linear sequencing distinguishing direct and indirect objects[5], and overt markings determining the roles of arguments within prepositional phrases.

---

[3] This structured representation is similar in some ways to the LS structures proposed in Hale (1983); however, I do not claim that this structure is universally configurational.

[4] It is an open question whether linear sequencing is used in this manner at all in natural language. Since only the structural configuration and overt marking types of **R** relations will figure in the analysis that will be presented, I will only consider these two types of relations in the remainder of this discussion.

[5] This may be the result of linear ordering facts, or it could be construed as evidence for structural configuration where the two levels of structure within V'' are used to distinguish direct from indirect objects. This would require that the indirect object be the sister of V while the direct object is the sister of V' in a sentence with two objects.

English makes full use of the **R** relations available to it. Notice, however, that (7i) is only used to distinguish subjects from objects[6]. If all the roles in a predicate event were distinguished in a local GR-Configurational manner we would need to have many different structural argument positions defined in the hierarchical structure of the sentence[7]. This proves unnecessary since languages tend to supplement (7i) with other **R** relations taken from types (7ii) and/or (7iii). If a language were to express all of the roles in a predicate event by means of overt markings (7iii), we would only need to have as many such markings as the maximal number of participants in the event.

If a language expresses an **R** relation in a local GR-Configurational manner (7i), we find that it must either complicate the phrase structure by defining many argument positions that are somehow distinguishable (which would most likely involve lexical marking), or supplement (7i) with other possible expressions of **R** relations. The least costly option for a grammar that has chosen to make use of local GR-Configurationality is to make use of local GR-NonConfigurational **R** relations (7ii) and/or (7iii) as well.

Generally, what we find is the relationship expressed in (8).

(8)

   local GR-Configurationality implies local GR-NonConfigurationality

As defined in (2), if a language makes use of local GR-Configurationality as an **R** relation, we will call such languages GR-Configurational. If a language does not make use of local GR-Configurational **R** relations, we will use the term GR-NonConfigurational. Notice that a GR-Configurational language may make use of local GR-NonConfigurational **R** relations, but, by definition, a GR-NonConfigurational language does not make use of local GR-Configurational **R** relations.

Accepting the standard assumption that sister positions of the verb are universally argument positions, GR-Configurational languages normally define at least one other structurally distinct argument position. This is commonly taken to be the specifier of **IP**. GR-NonConfigurational languages cannot have argument positions defined outside of the sisters of **V**, since otherwise there would be contrasting argument positions defined·in a local GR-Configurational manner. If there were such contrasting argument positions, there would be no way to prevent the base generation of arguments in such positions, which would contradict the GR-NonConfigurational nature of the language. Accepting, with Chomsky (1986), that the specifier of **IP** is an argument position, then a GR-NonConfigurational language cannot have specifiers of **IP**. Alternatively, a language without specifiers of **IP** cannot have local GR-Configurational relations

---

[6] And possibly direct from indirect objects.

[7] This is further complicated by the fact that hierarchical configuration is, in some ways, an artifact of the notational device used to represent syntactic structure, namely tree diagrams. To the extent that this hierarchy is string-vacuous (*ie.* not visible in the string), we would still need some observable indication of the structurally distinct positions required for the identification of the various arguments, most likely in the form of lexical markers.

defined with respect to this position. The projected structure for **I** then is as in (9) and never as in (10) for a GR-NonConfigurational language.

(9)

$$I^{MAX}$$
$$\diagup\diagdown$$
$$I \qquad V^{MAX}$$

(10)

$$I^{MAX}$$
$$X^{MAX} \qquad I'$$
$$\text{(argument)}$$
$$I \qquad V^{MAX}$$
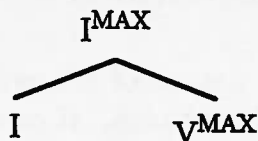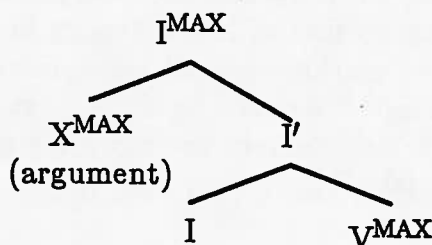
The question then arises as to how the shallow structure in (9) will be represented in the grammar.

In Chomsky (1986), the general assumption is that $X^{MAX} = X''$ for all **X**. If an intermediate category X' is unnecessary in a particular projection, then it is omitted for that projection. We have the situation as in (11).

(11)

$$X'' \qquad\qquad\qquad X''$$
$$X' \qquad\quad\equiv\qquad X \qquad \ldots$$
$$X \qquad \ldots$$

This opens the possibility that I^{MAX} in (9) is I'' and that (9) is as in (12).

(12)

$$I''$$
$$I \qquad V^{MAX}$$
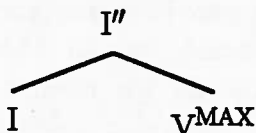
However, the convention illustrated in (11) is just that, a convention for simplifying the bracketed representation of a particular projection. We would still need to stipulate that there is never a specifier for I''.

One way to get around this is to allow I^{MAX} in (9) to be I' for GR-NonConfigurational languages. This precludes the existence of a specifier of I'', but raises a further question about whether or not I' really counts as a maximal projection or if it is some sort of degenerate projection. I will return to this issue in chapter 3.

A further issue concerns the nature of the arguments — that is, the sisters of V, in a GR-NonConfigurational language. If overt lexical markers are used to identify the roles of the various arguments, we find that the sister constituents of V contain not only the argument, but also the identifying marker. These markers occur in the forms of prepositions, postpositions, and case marking affixes. Accepting the strong version of X-bar theory in which all categories project, I also assume that these markers are all instances of the same category, P. Prepositions occur in left-headed structures while post-positions occur in right-headed structures. Case marking suffixes are morphologically bound instances of P. Note that P projects to P^{MAX} as in (13) even if P is morphologically bound[8].

(13)
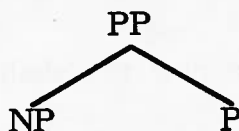
In this way, the NP is the argument of the P and an indirect argument of the verb, mediated by the lexical marker, P. In this sense, PP is an argument of the verb and occurs as a sister of the verb — that is, in an argument position. I set aside the issue of non-subcategorized PPs and whether they occur in argument positions (see Brunson, 1986 for discussion).

2. Case Theory and Configurationality

Each of the three strategies (presented in (7)) for representing arguments in the syntax involves consequences for Case Theory.

Local GR-Configurationality is defined according to common asymmetries of structural case. For local GR-Configurational relations, case is assigned by a governing adjacent case assigner (usually V or **Infl**). These are the situations that Case Theory was designed to handle. The Case Filter requires that all overt arguments have case. This provides yet another explanation of why it is that languages tend to only contrast two or three arguments in a local GR-Configurational manner. Not only do we need to have identifiable structural positions in order to recover the thematic roles of the arguments, there must also be a structural relationship between the argument and an adjacent case assigner.

---

[8] In the case of morphologically bound case suffixes, this structure may exist in the morphology before insertion into the syntactic tree structure.

If lexical marking is the relevant strategy for linking arguments and thematic roles, the lexical marker itself (*ie.* the **P**) would assign case, under adjacency and government, to its complement. In this way, the verbal arguments (*ie.* the **PPs**) have case internal to the structure that occurs in the sister positions of the verb. This avoids the need to compromise the adjacency requirement on case assignment. These arguments are not receiving case from the verb, but rather have case internally by virtue of the very marker that is necessary for them to occur in a GR-NonConfigurational manner. Since case is internal to the argument (taken together with its lexical marker), then the free-order facts about these arguments are predicted since they need not occur in any particular location in order to receive case.

One problem occurs when the lexical markers for GR-NonConfigurational relations are morphological (*ie.* case marking suffixes). Two possibilities present themselves. We could treat these suffixes as post-positional clitics that are really constituents in their own right but, incidentally, are morphologically bound. Another alternative is to allow case assignment to happen in the morphology, where constituent structures exist anyway. The **P'** constituent would then be inserted from the lexicon with inherent case.

By allowing the lexical marker that identifies the thematic role of the argument to perform the additional function of assigning case to the argument, we can predict the free-order facts that are evidenced by certain languages, including Warlpiri. In the case of Warlpiri, there is a phonetically null case marker (the absolutive). Since absolutives in Warlpiri do not need to be in a position to receive structural case, we are left with an empty category of type **P** for absolutive arguments. However, since Warlpiri allows discontinuous expressions, we can have more than just one such argument. If we can propose a null **P** for any nominal expression requiring case, it becomes impossible to violate the Case Filter in Warlpiri s-structure. All arguments will either receive case from an overt **P**, or else a null **P** will be introduced. Of course, it is possible for the semantics of an argument to render it ungrammatical in a particular role, but this is totally independent of the Case Filter. In this analysis there is no way for the Case Filter to block a sentence in Warlpiri.

The role of intonation figures crucially in such situations, since it appears that, in Warlpiri, there are intonational conditions controlling when a nonovert **P** can be introduced into the s-structure. The importance of the intonational facts go far beyond the analysis of Warlpiri nominal phrases and the Case Filter. It also proves to be crucial in avoiding rampant ambiguity and non-determinism in parsing Warlpiri s-structure. This will be discussed in more detail in a later chapter.

3. Theta Theory and Configurationality

Discontinuity of arguments tends to occur only with arguments that are identified in a GR-NonConfigurational manner. In this section I will explore

some possible reasons for why this might be the case, as well as propose a preliminary analysis of discontinuous expressions.

With local GR-Configurational **R** relations, the structural location of the argument, to the extent that it is well defined, must be unique. Since local GR-NonConfigurational **R** relations rely on lexical marking to link their thematic role to the predicate, provided that two or more arguments bear the same type of lexical marking, they will be linked to the same thematic role.

This is fairly straightforward, essentially capturing the fact that agreement of markings is necessary to interpret discontinuous expressions. This was evident in the examples of discontinuous expressions in Warlpiri that were presented in chapter 1.

Where this analysis runs into difficulty is with the θ-Criterion which is stated in (14).

(14) θ-Criterion: Each argument bears one and only one θ-role and each θ-role is assigned to one and only one argument. (Chomsky, 1981, p.36)

The θ-Criterion requires a bijection between the lexically subcategorized arguments and the syntactic arguments for a particular predicate. However, at some level of representation, discontinuous arguments must be discontinuous. If the θ-Criterion applies at all levels of representation, then it is violated by discontinuous arguments, since then there may be a many-to-one mapping between syntactic arguments and subcategorized θ-roles. A reasonable level of representation to have the actual discontinuity of discontinuous arguments is at s-structure, since s-structure represents the surface syntactic relations of the sentence. But then it seems that we have the θ-Criterion not applying at s-structure in those languages that allow for discontinuous expressions.

This seems to be a rather strong exception to a fundamental principle of grammar. But there is a way around making the θ-Criterion optional at s-structure. Suppose instead that the θ-Criterion applied universally at s-structure, but to a percolated argument structure at the level of **C″**. Constraints on percolation that are derivative of the different types of **R** relations can provide us with an instantiated θ-grid at the uppermost node of the sentence, which is then subject to the θ-Criterion's requirement that there be a bijection between syntactic arguments and subcategorized arguments.

Since there are only single subject and object structural locations in GR-Configurational languages, we can only have one constituent occupying these positions. This disregards such afterthought constructions as shown in sentences (15) through (16).

(15) The child, the orphan, wanted a new toy.

(16) I gave the committee, the executive, the report.

In these cases the second of the two noun phrases in the subject or indirect object position can be seen as correcting and replacing the initial noun phrase. This contrasts with such multiply borne thematic relations as the locatives in

sentences (17) and (18).

(17) I sat in the park on the bench.

(18) I stayed in the city in a hotel.

In each of these sentences, the two locative **PP**s express further restrictions on the locative role, rather than the second replacing the meaning of the first.

Certain thematic relations that can be expressed in both a GR-Configurational as well as a GR-NonConfigurational manner, seem to have the same restriction when they are in a GR-NonConfigurational relation as they do when in a unique structural location. If there is more than one such argument, the second is generally taken as correcting the first in an afterthought structure rather than as further restricting the first. Consider examples (19) through (20).

(19) I presented report Z25G to the committee to the executive.

(20) The report was filed by the committee by the executive.

This discussion has been focused on English, so we cannot make claims about universals of linguistic theory. However the emerging analysis of the situation in English is that for each GR-Configurational **R** relation, as well as for each GR-NonConfigurational lexical identifier, there is an associated list of thematic relations that can be expressed by that relation or lexeme. If a particular thematic relation has a GR-Configurational **R** relation as a possibility, then this thematic relation cannot be multiply borne in a discontinuous, further restricting manner. GR-Configurationality, then, defines a partition of thematic relations, isolating those that can freely be multiply borne from those that cannot.

The situation with Warlpiri is that there are no GR-Configurational **R** relations, hence there are no constraints on discontinuity, so all argument types may be discontinuous.

The constraint on whether a thematic relation can be multiply borne is no longer an issue for the θ-Criterion, but rather is a constraint on the indexing of overt arguments into a thematic structure during the course of percolation. The partition of thematic relations according to GR-Configurationality defines two types of indexing behavior. Those thematic relations that can be expressed in a GR-Configurational manner (that is, those thematic relations that cannot be multiply borne) must index biuniquely to subcategorized positions on the percolating grid. This predicts the sort of interpretation assigned to the afterthought constructions discussed above, since, if the second argument is to index into the structure, then the condition on biuniqueness will force it to override any previous indices.[9] In the case of the thematic relations that cannot be expressed in a GR-Configurational manner (*eg.* locatives, temporals, *etc.*, in English, and all thematic relations in Warlpiri) the biuniqueness restriction does not hold and multiple arguments are free to link into the same position on the thematic structure.

---

[9] This is not a direct consequence of the theory proposed here, but it is a performance strategy that is consistent

In this way, at s-structure the discontinuous arguments are, in fact, discontinuous. A thematic structure is percolated, starting at the level of the verb's θ-grid, to a complete thematic structure at the level of **C''**. This percolation is sensitive to restrictions on biuniqueness that are derivative of the GR-Configurational relations used to identify certain thematic relations. The θ-Criterion then applies to the percolated structure, checking that the subcategorized arguments are satisfied.

What has been done is to take some of the power out of the θ-Criterion at s-structure. The θ-Criterion verifies that subcategorization frames are satisfied, while the uniqueness of arguments is controlled by a parametrized percolation convention that applies equally to subcategorized and non-subcategorized arguments.

Given this percolation convention, we are now in a position to handle the optional pronouns found in Warlpiri. Consider the Warlpiri sentences (21) through (23).

(21)

parnka-mi   ka-rna
run-NONPAST   (AUX)Pres-1sg(subj)

'*I am running.*'
(Hale, 1982, p.230)

(22)

ngaju   ka-rna   parnka-mi
I   (AUX)Pres-1sg(subj)   run-NONPAST

'*I am running.*'
(Hale, 1982, p.217)

(23)

parda-rni   ka-rna-jana
wait-NONPAST   (AUX)Pres-1sg(subj)-3pl(obj)

'*I am waiting for them.*'
(Hale, 1982, p. 231)

In Warlpiri, pronouns are usually optional with the agreement on the auxiliary providing the information that is supplied redundantly when there are overt pronouns.

If we use an extended percolation convention to allow the auxiliary itself to instantiate argument positions in the θ-grid, then we can avoid the traditional problems with the θ-Criterion and non-overt arguments. The θ-Criterion at s-structure need check nothing more than that all subcategorized arguments in the projected grid are legally indexed into the syntactic arguments of the

sentence without introducing empty categories in the syntactic structure. Whether that indexing is constrained by the biuniqueness constraint or not is irrelevant to the θ-Criterion, as is whether the index arrives from an argument itself or from agreement on the auxiliary. I discuss the details of the θ-grid percolation conventions for Warlpiri in chapter 3. For now it suffices to have shown that the θ-Criterion need not prove problematic for the diverse argument structure phenomena that are evidenced by Warlpiri s-structures.

By defining the term *GR-Configurational*, I have been able to isolate certain types of syntactic behaviour involving Case Theory and Theta Theory. The division between GR-Configurational and GR-NonConfigurational languages is not proposed as a typological division of languages. It is merely a convention for describing local relationships used for argument identification. Languages vary considerably in how they use GR-Configurational and GR-NonConfigurational relations, but there is enough common ground that we can retain the general Case Theory and Theta Theory. What is striking about Warlpiri is the total absence of GR-Configurationality. The analysis of Warlpiri presented in the next chapter will assume that Warlpiri is GR-NonConfigurational, and proceed to analyze it within the GB framework.

---

with the biuniqueness requirement on certain arguments.

CHAPTER 3

Configuration in Warlpiri

The study of languages with properties such as free order and discontinuity — properties not generally not found in the widely studied European languages — has led to proposals that there are typological classes of languages. Common to such proposals is the assumption that the different properties of these languages must be captured by a parametrization of the most fundamental principles of linguistic theory. I take issue with this general line of approach. I maintain that the best way to evaluate the status of a language within a linguistic theory is to assume that the theory applies equally well to all languages. Overt differences that prove problematic for the theory of grammar as it stands can be used to indicate necessary changes in the theory itself. In this way, data from languages such as Warlpiri can be used to single out the principles of grammar that need to be revised or parametrized so that the theory might be extended to account for a larger range of phenomena evidenced in natural languages. Unless there are fundamental problems with the theory, these revisions should be changes of detail and not fundamental changes to the theory.

Accepting this line of argument, I find that the properties of putative *nonconfigurational* languages are principled differences and they do not indicate that such languages should be classified as typologically distinct. What emerges is a set of parametric choices in the principles of grammar that can yield a gradation of language types with respect to properties such as free word order and discontinuity.

1. Why Warlpiri?

In recent years, more and more languages that had been thought to be quite different from European languages have been studied in the context of linguistic theory. As they are studied, it has become evident that these languages aren't so different after all. The approach to Warlpiri has been somewhat different in this respect. Most analyses of Warlpiri have assumed that it is exceptional, so Warlpiri remains as a language that has been extensively studied and documented, yet which has not yet been analyzed as conforming within the existing theory.

A significant factor in the choice of Warlpiri has been the extent to which data is available on the language. In this analysis, I rely on data from previously published works on Warlpiri, and am greatly indebted, in particular, to Ken Hale and David Nash for their extensive fieldwork and writings on this

language. Another contributing factor is the extreme differences that Warlpiri exhibits compared to, say, English. This makes it a challenge not only from a linguistic theory point of view, but also from a processing perspective. Warlpiri illustrates properties that have not generally been dealt with in language processing models. For these reasons, a detailed study of Warlpiri syntax has implications for linguistic theory as well as the theory of language processing and parsing.

2. X-bar Theory and Warlpiri Tensed Clauses

Simple tensed sentences in Warlpiri demonstrate a remarkable freedom of word order. With the exception of the auxiliary any order of content words in such sentences is allowed.

(1)

ngaju	ka-rna	parnka-mi
I-(ABS)	AUX:pres-1sg(subj)	run-NONPAST

I am running.

(2)

parnka-mi	ka-rna	ngaju
run-NONPAST	AUX:pres-1sg(subj)	I-(ABS)

I am running.
(Simpson, 1983, p.90)

The most common X-bar type schema used to account for this free order is stated in (3).

(3)

$$V' \rightarrow AUX \; X'^* \; V \; X'^*$$
(Hale, 1983, p. 7)

The X'^*s are essentially restricted to the arguments of the verb. This schema produces a fan structure for tensed clauses with a fixed position for the auxiliary and a free position for the head (*ie.* the V). Presumably the V' is the uppermost node in the tree structure of tensed clauses. From this phrase structure rule, the status of the auxiliary is somewhat questionable. It seems to be part of a discontinuous verb since it is non-maximal and a sister of the verb. In Hale's analysis, whenever the auxiliary occurs in second position, it is assumed that a rule of the PF component has taken place to move it from first to second position. In the syntax, the auxiliary is always initial, and it is only in the PF level of representation that it occurs in second position.

The auxiliary, being an element in the tensed sentence with a relatively fixed position, is an obvious starting place in attempting an X-bar analysis of Warlpiri. The factors that determine the position of the auxiliary are rather complex, but it is generally restricted to being either the first or second

constituent in a tensed sentence. The auxiliary is the locus of tense and aspect morphemes as well as agreement morphemes that agree in person and number with arguments of the verb. The category **Infl** normally includes tense and agreement elements. Consistent with the current approach to non-lexical categories, we can analyze the Warlpiri auxiliary as an instance of **Infl**.

If we accept the approach to the structure of **S** taken in Chomsky (1986), then **Infl** projects in the syntax. If in Warlpiri, **Infl** projects to **I″** (as in (4)), then there would be a structurally defined argument position before the auxiliary in the specifier of **I″**.

(4)

$$
\begin{array}{c}
I^{MAX} \\
\diagup \quad \diagdown \\
X^{MAX} \qquad I' \\
(\text{argument}) \quad \diagup \diagdown \\
I \qquad V^{MAX}
\end{array}
$$
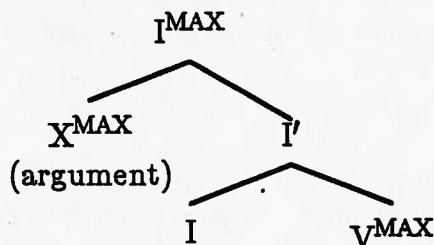
It is tempting to use this structure to account for second position auxiliaries in Warlpiri, allowing the specifier (which is an optional node) to hold the first position constituent when the auxiliary comes in second position. But generally the specifier of **I″** is an argument position that identifies a GR-Configurational relation. As discussed in chapter 2, grammatical relations are GR-NonConfigurational in Warlpiri and are identified by morphological marking rather than by structural configuration. That is, it is not sufficient that an argument occur in the pre-auxiliary position to identify it in a particular thematic function. This first position element can range over any independently identified thematic role, and extra evidence (in the form of a lexical case marker) is required to determine which thematic role is, in fact, borne by this argument. To say that Warlpiri syntax is GR-NonConfigurational is not to set a high-level parameter for the exceptionality of Warlpiri. What this means instead is that, in Warlpiri, there are no structural positions that consistently link to a predictable element in a θ-grid, and hence structural position will not serve to identify the thematic function of an element. This is not a global parameter, but a local fact that must be acquired for each thematic type. The expression GR-NonConfigurational is not, then, a typological classification via a parameter of fiat, it is instead an observational classification by means of universal generalization on a local property that consistently obtains. The pre-auxiliary constituent in Warlpiri is freely selected from any of the verbal arguments, or even the verb itself. If the specifier of **IP** is universally a thematically relevant position, it is doubtful that this is the appropriate location for the pre-auxiliary constituent in Warlpiri tensed clauses. This suggests that **Infl** in Warlpiri only projects to **I′**. The instantiated schema for **Infl** phrases in Warlpiri is as in (5).

(5)

$$I' \rightarrow I \; VP$$

Although Hale (1981) has claimed that Warlpiri is right-headed, notice that this is a left-headed rule.

The structure of a sentence is assumed to have **C″** as the uppermost node, with **IP** as the complement of **C″**. Is there any evidence to support this claim in Warlpiri? There are not normally any overt complementizers in simple tensed clauses; however, dependent tensed clauses do have overt complementizers. This is the case even if the clause is dependent in the discourse and not in the particular sentence.

(6)

yi-lpa-na	ya-ntala	wilinji
COMP-AUX:past-1sg(subj)	go-irrealis	hunting

I should go hunting; I would like to go hunting.

(7)

yali	kutja-ka	kari-mi	tjapananka
that(removed)	COMP-AUX:pres	recognize-Nonpast	Japanangka

There stands Japanangka.

(Hale, 1976, p.96, with morpheme glosses from Hale, 1974)

The complementizers, *yi* and *kutja* above, are prefixed to the auxiliary base. These are the same complementizers that function as relative clause complementizers in subordinate constructions. I propose that, although these are phonologically bound complementizers, they are dominated by a **C** node in the syntax. This category projects to the structure presented in (8).

(8)

$$
\begin{array}{c}
C^{MAX} \\
\diagdown \\
X^{MAX} \quad C' \\
\diagdown \\
C \quad I^{MAX}
\end{array}
$$
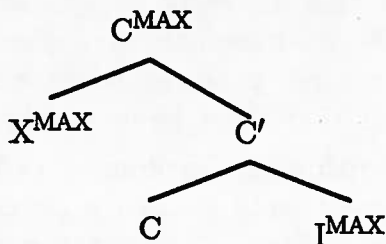
In this case the specifier is not an argument position, so the objection raised with respect to **I″** does not apply here. This specifier of **C″** is the correct position for the pre-auxiliary constituent, as can be seen from example (7), where the pre-auxiliary constituent comes before both the auxiliary and the complementizer. In general, if there is a complementizer, then it occurs as a prefix on the auxiliary and no constituent can intervene between it and the auxiliary.

I will now present some additional evidence that the pre-auxiliary position in Warlpiri is actually the specifier of **CP**. The specifier position of **CP** is traditionally the location of fronted wh-phrases. In Warlpiri, question words must occur in pre-auxiliary[1] position if they are to receive sentential scope (Nash, 1980). When they do, they satisfy the requirement on some auxiliaries that they must be in second position. The presence of a question word in initial position precludes any other constituent occurring before the auxiliary. For this reason, it is clear that there is a single pre-auxiliary constituent that can be either a fronted question word or the standard pre-auxiliary constituent.

To complete the discussion of the structure of Warlpiri tensed clauses, we need to account for the free order of constituents after the auxiliary. I accept the standard assumption that **VP** is the complement of **IP**. However, in the current theory the status of specifiers of **VP** is unclear. If these are argument positions[2], then, for the same reason that we did not want specifiers of **IP**, we do not want to allow specifiers in **VP**. In Warlpiri, there does not appear to be any need for internal constituency in the verb phrase. Rather than taking a stand on the nature of specifiers of **VP**, I will assume that $V^{MAX} = V'$ for Warlpiri since there is no need for the constituency that a two-level projection would give us in Warlpiri verb phrases. This category will be the locus of free-order and discontinuity in Warlpiri, and is exceptional in a number of ways. It is an n-ary branching fan structure with no apparent requirement that the head be peripheral. The sister categories of the verb are arguments with inherent case (*ie.* **PPs**) or embedded clauses (*ie.* **CPs**).

2.1. Pre-auxiliary Placement as Syntactic Movement

Having resolved that the pre-auxiliary constituent is in the specifier of **C″**, we are committed to a movement analysis to account for any first-position constituents.

Hale (1983) maintained that there is no syntactic movement in Warlpiri, and treated the pre-auxiliary phenomenon as a phonetic adjustment rule that moved the auxiliary into second position when necessary. This analysis is motivated by the requirement that there be no empty categories in Warlpiri.

Clearly, in a language with such freedom of order as Warlpiri exhibits, to propose empty syntactic constituents creates a potential problem for recoverability of syntactic structures. After all, np-trace can occur wherever syntactic arguments occur, so if there are few constraints on where arguments may be realized, then there are few constraints on traces. For this reason, empty categories could be generated virtually anywhere in the phrase structure.

Nevertheless, as we shall see, scoping facts for questions dictate that the pre-auxiliary constituent be positioned through syntactic movement. If we

---

[1] I will continue to use the term *pre-auxiliary* for this position even though it is before both Infl and Comp.

[2] As is assumed in GR-Configurational analyses of dative constructions in English.

accept the general structure of the grammar presented in (9), we are committed to having the syntax handle any phenomena that requires access to both PF and LF information.

(9)

```
                    d-structure
                        |
                    s-structure
                       / \
                      /   \
                     /     \
                   PF       LF
            (phonetic form)  (logical form)
```
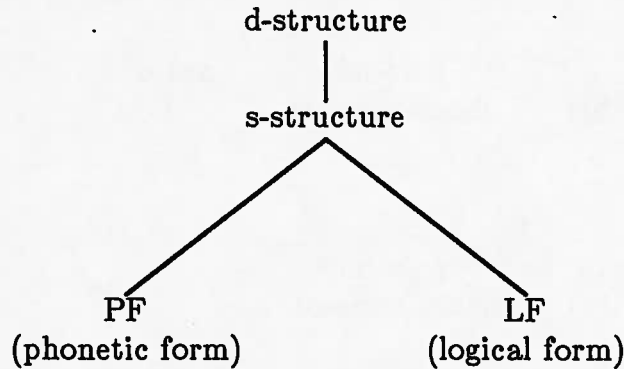
The placement of the auxiliary involves access to information normally represented in PF. According to Hale (1973)

"... if an auxiliary base .. is disyllabic or longer, the auxiliary may remain in initial position." (p.312)

Hale assumes that the auxiliary is initial in Warlpiri finite clauses and that a rule of AUX-Insertion moves the auxiliary into second position if its base is monosyllabic or null. This rule is optional if the auxiliary is disyllabic or longer. If syllable counting is an appropriate criterion[3] for the placement of the auxiliary, then this is information from the PF component. Since the environment for the rule is stated in terms of a PF condition, then the rule must apply at a level of representation that can access the PF component. This restricts the possibilities to PF or s-structure (which has direct access to the PF level of representation).

Scope information is represented at LF. Clearly the placement of question words with sentential scope in the pre-auxiliary position involves information from the LF component. It is not immediately as clear that the placement of non-question pre-auxiliary constituents involves scope information. However, there are circumstances where scope is definitely involved in the placement of the auxiliary, even when question words are not involved. There is a restriction with the negative complementizer, *kula*, (which is prefixed to the auxiliary) that the verb of the sentence cannot occur in the pre-auxiliary position if *kula* is to have sentential scope. This situation is illustrated in sentences (10) through (14).

---

[3] This is clearly not the type of rule that provides any explanation about the position of the auxiliary, and suggests a sensitivity in the grammar to prosodic constituency.

(10)

natju   kula-ka-na   pula-mi
  I   neg-AUX:pres-1sg(subj)   shout-nonpast

I am not shouting.

(11)

      kula-ka-na   pula-mi   natju
neg-AUX:pres-1sg(subj)   shout-nonpast   I

I am not shouting.

(12)

      kula-ka-na   pula-mi
neg-AUX:pres-1sg(subj)   shout-nonpast

I am not shouting.

(13)

  *pula-mi   kula-ka-na   natju
shout-nonpast   neg-AUX:pres-1sg(subj)   I

(14)

  *pula-mi   kula-ka-na
shout-nonpast   neg-AUX:pres-1sg(subj)

(Hale 1973, p. 313)

The relevant observation here is that we cannot have the verb located before the negative complementizer. Verbs are generally free to occur in initial position, so the situation with *kula* is somewhat exceptional and has more to do with the scope of the negation than with any phonological criteria. Hale accounts for this situation by adding an exception to his AUX-Insertion rule. However, if this is a PF rule, it should have no access to scope information, not even in an exception clause. If we accept that this restriction on the placement of the verb in negative clauses is an LF condition, then LF constrains the movement to the specifier of C″. Clearly, if we attempt a uniform analysis of question movement and pre-auxiliary movement, then we require access to LF information. What this means is that our movement rule must happen in a level of representation with access to LF information; the only two candidates being s-structure and LF itself.

So, the movement of a constituent into the specifier of C″ must access PF and LF, simultaneously. The only option, then, is for this to be an s-structure rule. This is intuitively correct as well, since s-structure should represent the surface relationships of the sentence, and clearly the order of constituents is a surface phenomenon.

The analysis proposed here is that the constituent that may occur before the auxiliary is moved freely in s-structure by move-α. The resulting structure is then subject to PF and LF well-formedness conditions that can restrict the acceptable instances of move-α. The movement is to the specifier of C'' and hence is an instance of $\bar{A}$-movement[4]. This movement rule, as should be expected, leaves a trace at the original location of the constituent. Given the free-order of arguments in Warlpiri, recovering the exact location of this trace is a problem. Without loss of generality, I will adopt the convention that the trace is located in the nearest possible position (in a linear sense) to the final position of the moved constituent. This is not a theoretical claim, but merely a convention to allow the traces to be unambiguously put into the tree structures[5]. This will amount to the trace being initial in V'.

I do not think that the recoverability issue for the exact location of the trace poses any real problem to this analysis. In analyses of English, recoverability of traces is not considered a problem, yet with preposed prepositional phrases, and oblique wh-questions the same recoverability questions can be raised. Prepositional phrases can occur in fairly free order in English, as can all arguments in Warlpiri, so to the extent that we do not worry about the exact location of the trace in sentences like (15) we need not concern ourselves about the exact location of traces in Warlpiri.[6]

(15) Where$_i$ did John go (t_i) quickly (t_j)?

If there is a trace in the Warlpiri sentence, then we must guarantee that it is properly governed to satisfy the ECP. In the next section, I will demonstrate how this trace is not problematic for proper government.

2.1.1. The ECP and Movement in Warlpiri

The analysis of pre-auxiliary constituents in Warlpiri involves the movement of a constituent from within V' to the specifier of CP. This situation is illustrated in (16). Recall that the trace is located in initial position in V' merely as a convention.

---

[4] The movement of the head V adds an extra complication, but this will be dealt with in section 2.1.2 of this chapter.

[5] As far as the syntax is concerned, the traces can be located anywhere within V', but since a parser needs to have more specific information about the location of the trace, this convention is necessary.

[6] I am indebted to Richard Sproat for discussion of these issues.

(16)

$$
\begin{array}{c}
C'' \\
\diagup\;\diagdown \\
X_i^{MAX}\qquad C' \\
\diagup\;\diagdown \\
C\qquad I' \\
\diagup\;\diagdown \\
I\qquad V' \\
\diagup\diagdown \\
t_i \ldots V \ldots
\end{array}
$$
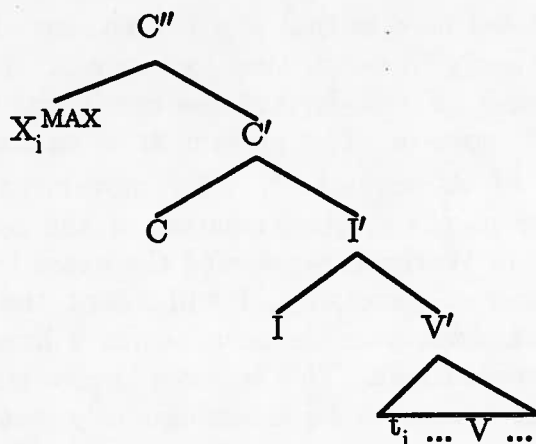
The ECP requires that this trace be properly governed. I will work with the most recent version of the ECP and proper government as found in Chomsky (1986) and briefly presented in chapter 1.

In this section I will only be concerned with the movement of a maximal category from within V' to the specifier of CP, and return in a later section to the movement of the zero-level category, V, to pre-auxiliary position.

The ECP requires that all traces be properly governed. Since we accept that Warlpiri is GR-NonConfigurational, then the arguments in V' are directly θ-marked by the lexical marker that also assigns case to the argument. The verb only indirectly θ-marks the arguments within V'. Since direct θ-marking is a requirement of θ-government (see definition 13, chapter 1) the verb does not θ-govern arguments within V'. The NP within the PP is θ-governed by the P but it is the entire PP that moves, and the constituent PP is not θ-governed.

The only way to ensure proper government of t_i is through antecedent government by its antecedent, X_i^{MAX}. So, to satisfy the ECP, all we need to do is ensure that X_i^{MAX} governs t_i.

Clearly we have m-command between X_i^{MAX} and t_i since CP is the only maximal projection dominating X_i^{MAX} and this also dominates t_i. Are X_i^{MAX} and t_i sufficiently local for government? Since locality for government is defined in terms of barriers (see chapter 1, section 1.2.6) we have at most two potential barriers intervening between X_i^{MAX} and t_i, namely the two maximal categories I' and V'.

One way to guarantee antecedent government is to assume that universally $X^{max} = X''$. This suggests that the single-level projections I' and V' are not really maximal as far as the definition of blocking category is concerned. In this way I' and V' are not potential barriers, so X_i^{MAX} properly governs the trace.

This analysis predicts that the only barrier from the categories that we have seen in Warlpiri is **CP**, which in turn predicts the fact that there is no mixing of arguments across clauses in Warlpiri. We don't find situations such as that shown in (17).

(17)
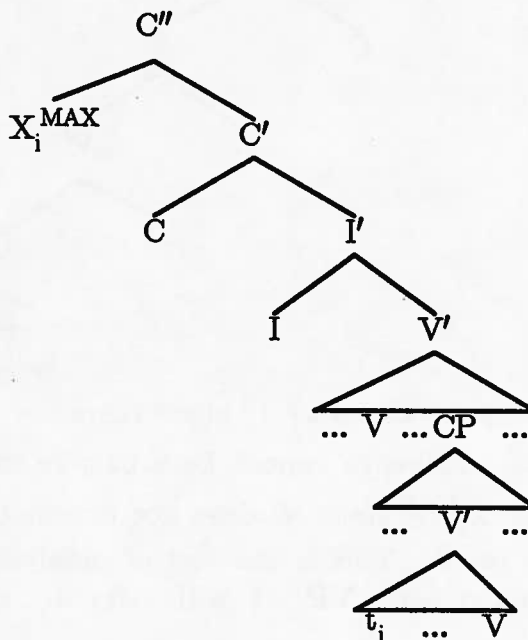
Of course, the entire embedded **CP** can occur in the specifier of the matrix **CP**, but we cannot extract an argument from within the embedded **CP** to the specifier of the matrix **CP**. This is exactly what we would expect with **CP** being a blocking category and hence a barrier for the trace.

By requiring that maximal projections have two bar-levels, we can predict the freedom of movement of a constituent to the specifier of **CP**. As well, we also predict the single apparent constraint on that freedom — namely, that clauses do not mix arguments. I will refer to this possibility as the *degenerate projection hypothesis.*

There is another possibility for ensuring antecedent government between X_i^{MAX} and t_i while allowing single-bar projections to be treated as maximal in the absence of a two bar-level projection. What we need to do is prevent **V'** from being a blocking category for the trace. If **V'** is not a blocking category, then **I'** will not be a barrier, regardless of whether it is a blocking category or not, since I^{MAX} can only become a barrier by inheritance. Since I^{MAX} immediately dominates **V'**, if **V'** is not a blocking category for the trace, then I^{MAX} cannot be a barrier by inheritance (clause i of definition 16, chapter 1).

What may happen, then, is for the argument to move first from within **V'** to adjoin to **V'** before moving to the specifier of **CP**. This results in two traces as shown in (18).

(18)

$$C''$$
$$X_i^{MAX} \qquad C'$$
$$C \qquad I'$$
$$I \qquad V'$$
$$t_i^1 \qquad V'$$
$$t_i^2 \ldots V \ldots$$
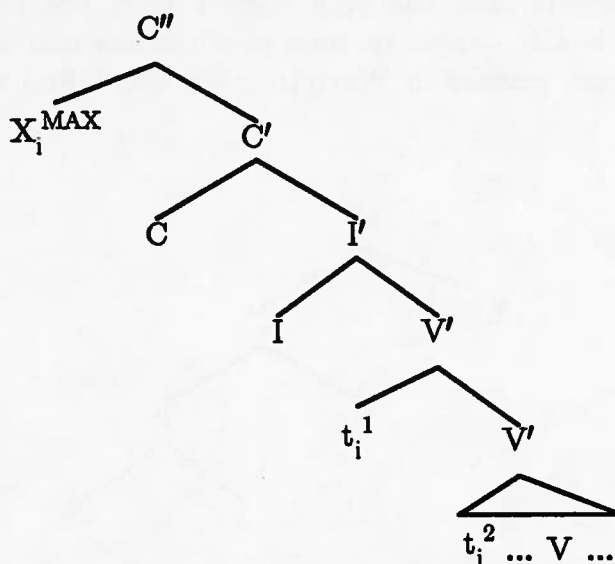
Notice that t_i^2 is properly governed by t_i^1 since there are no intervening barriers (V' does not exclude t_i^1 and hence cannot be a barrier to government). Also t_i^1 is properly governed by X_i^{MAX} since V' does not dominate t_i^1 and hence cannot be a blocking category for it. This is the sort of analysis that Chomsky (1986) proposes for movement out of **VP**. I will refer to this as the *adjunction hypothesis*.

The *degenerate projection hypothesis* captures both the fact that these projections do not have specifiers as well as the fact that they do not interfere with government. These two facts both follow from the degenerate nature of the projections which only project to a single bar-level. The *degenerate projection hypothesis* clearly involves fewer mechanisms, but it is not clear what far reaching consequences this hypothesis may have. It has implications for the general X-bar schema since these degenerate projections are treated as maximal with respect to properties of the X-bar component. The main problem for this hypothesis is that these constituents are not treated uniformly as maximal projections in all the components of grammar. However, there is a precedent in the theory for exceptional projections — consider, for example, the stipulated clause about **IP** in the definition of barriers.

The *adjunction hypothesis*, on the other hand, is entirely analogous to other analyses proposed within the theory. It is not at all clear what sort of empirical evidence would help to decide between the *degenerate projection hypothesis* and the *adjunction hypothesis*.

We must now consider the special case of the movement of the zero-level verb into pre-auxiliary position.

2.1.2. Verbs in Pre-auxiliary Position

When verbs occur in pre-auxiliary position, they cannot be in the specifier of **CP**. Specifiers can only contain maximal projections, so a verb, being an X^0 category, cannot occur in a specifier position. This is a result of structure preservation (Emonds 1976, Chomsky 1986), which requires that X^0 categories can only move to adjunction structures. A constraint on adjunction requires that X^0 categories can only adjoin to other X^0 categories, hence the movement of the verb to pre-auxiliary position must be an instance of head adjunction to the complementizer. The resulting structure is shown in (19)[7].

(19)
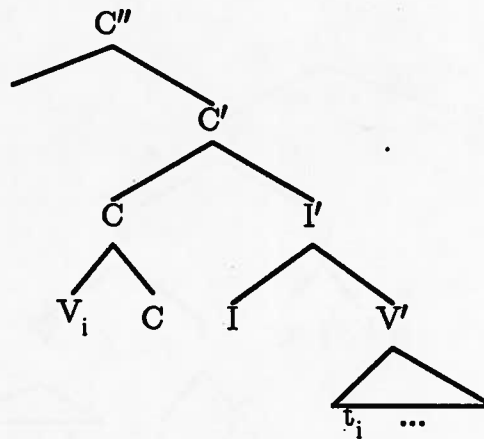
If we accept the *degenerate projection hypothesis*, then the trace of the verb is properly governed by the moved verb since there are no intervening maximal projections that could be barriers. The ECP is satisfied and there are no apparent problems.

If, however, we maintain that X' can be maximal, the analysis is once again more complex. The *adjunction hypothesis* will not work here since the X^0 category cannot adjoin to V', V' being maximal in this analysis. What we need, then, is to arrange for V' to be L-marked. The only candidate for an L-marker for V' is I, since I is its only sister. Let us assume that I directly θ-marks V'. This might be a realization of the argument structure expectations that are set by the auxiliary's person and number agreement markers. If I directly θ-marks V', then I θ-governs V'. We only need I to be lexical to satisfy the requirements for L-marking. This is a major problem since I is generally assumed to be non-lexical.

---

[7] Since there is only a single pre-auxiliary constituent, we cannot have movement to the specifier of CP as well as movement of a verb into adjunction with C. This constraint may be derivative of a requirement for spec-head agreement in CP (Chomsky, 1986). If the specifier and the head of CP must agree, then the [+V] feature that the moved verb gives the C would not agree with the [−V] feature of any arguments in the specifier. Since there is generally only one verb, and X^{MAX} cannot head-adjoin, we can predict that there will be either a constituent in the specifier of CP or a verb adjoined to C, but not both.

Following the analysis of V-raising presented in Chomsky (1986 pp. 69-71), **V** could first move into head adjunction with **I**. When **I** has a lexical category (the verb) adjoined to it, it is then capable of L-marking **V'**. This removes **V'** as a potential blocking category or barrier for the trace of the verb within it. The verb would then move from head adjunction with **I** into head adjunction with **C**. The intervening **IP** cannot be a barrier for proper government since there are no other maximal projections for it to use to satisfy the definition of inherited barrierhood.

In this way, there are two traces of the moved verb. The intermediate trace properly governs the lowest trace, and the verb, itself, properly governs the intermediate trace. This situation is illustrated in (20).

(20)
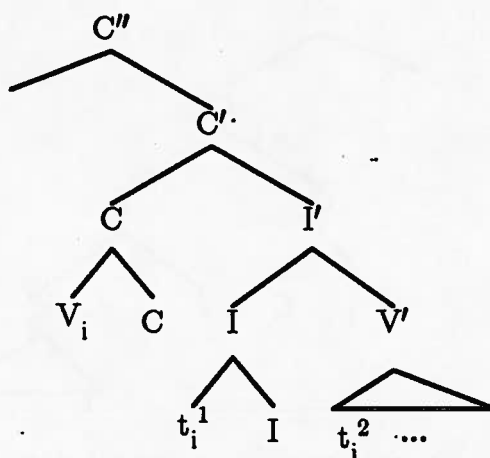
Again it is hard to find empirical evidence supporting one analysis over another. Clearly, the *degenerate projection hypothesis* handles both movement of arguments and movement of the verb with relative ease, but again the successive movement analyses have precedents in the theory. The choice here remains an open question, but either way, it should be clear that GB theory does not preclude the movement analysis that I have proposed for Warlpiri.

2.2. The Internal Structure of Arguments

So far, I have not discussed the internal structure of Warlpiri **PPs** and embedded **CPs**. These are the sister constituents of the verb, the constituents that can occur in free order or in discontinuities. In the syntax of tensed clauses discussed above, the X-bar schema was instantiated as left headed — both the complementizer and the auxiliary occurred before their complements. **V'** has no apparent directionality of headedness. Now we will find that with the internal structure of arguments in **V'**, the X-bar schema is instantiated as right-headed.

2.2.1. The Structure of PP

As discussed in chapter 2, Warlpiri is a GR-NonConfigurational language with the thematic roles of arguments identified by means of overt lexical markings. These markings are in the form of case suffixes. Since these suffixes are morphologically bound to the nominal, it is possible that the entire **PP** comes from the lexicon as a single constituent. This does not preclude there being internal structure to **PP**. In fact, given that **PP** is right-headed, and morphology is claimed to be universally right-headed (Di Sciullo and Williams 1985), we can account for the apparent change in headedness from left to right as resulting from different headedness requirements of the syntax and the morphology, respectively.

The structure of Warlpiri **PP**s will be assumed to be as shown in (21), with $\mathbf{P^{MAX}} = \mathbf{P'}$, since there is no apparent need for the extra structure that a two-level projection would provide.

(21) .

```
        P′
       / \
      /   \
    NP  .  P
```
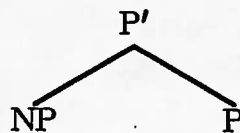
Case assignment is internal to **P′**, and perhaps not in the syntax at all since the internal structure of **P′** appears to follow the rules of the morphology.

The internal structure of **NP**s also has properties more typical of morphology than syntax. Since there is no apparent category distinction in Warlpiri between traditional adjectives and nouns, we find situations where **NP**s do not have a single unique head. I am not aware of any need for internal constituency among the various **N**s that can occur within an **NP**. The literature is unclear as to whether any order of **N N** sequences within a single **NP** is preferred. When the two nominals are in discontinuous expressions, the order is apparently free, but order preferences, if they exist, are not documented for contiguous nominals in a single constituent. If we accept that there is no internal constituency between nominals in complex nominal phrases, then the structure shown in (22) can be used.

(22)

```
        N′
      / | \
     /  |  \
    N   N   N
```
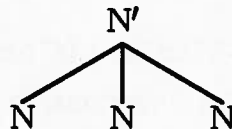
The notion of a *relative head* (defined in (23)) has figured prominently in recent discussions of morphological structure.

(23) *Relative head:* The *head*/F (*ie.* the head with respect to the feature F) of a word is the rightmost element of the word marked for the feature F. (Di Sciullo and Williams, 1985)

Using this definition, if N's are constructed in the morphology, then the X-bar schema poses no problem for the structure of N'. The right-most N instantiates the category feature of N', while other properties (especially semantic properties) would percolate to the phrasal category from the rightmost constituent possessing a value of that property. This is particularly useful in providing the information about plurality to the phrasal category of a complex N'. Consider the examples given in (24) and (25).

(24)

kurdu	wita-jarra-rlu
child	small-dual-ERG

'*(two) small children*'

(25)

kurdu-jarra-rlu	...	wita-jarra-rlu
child-dual-rlu	...	small-dual-ERG

'*(two) small children*'

If we accept that nominals have a default feature, say [*sg num*], then, in the absence of an overt number marker, the phrasal category receives this feature from the right-most nominal that it contains. The presence of an overt number marker to the right of the last nominal will result in its feature for number being percolated to the phrasal category. If the number marker is unmarked for category, then the right-most nominal will still be the head of the phrase with respect to the category.

2.2.2. The Structure of Infinitival CP

Infinitival clauses occur in the same structural positions as **PP** arguments in a matrix clause. They can be moved into pre-auxiliary position, and, as with pre-auxiliary PPs, a pre-auxiliary **CP** can have a fairly complex constituent structure[8]. Embedded clauses usually have an obligatorily controlled subject (*ie.* the subject is missing but construed as coreferential with an argument in the matrix clause). The general sequential order of elements in complex infinitival clauses is as follows:

$$\text{ARG}^* \text{ V (Infinitival-Marker) (Complementizer)}$$

The complementizer contains the information needed to resolve the con-

---

[8] There appears to be a statistical preference in published data on Warlpiri for post-auxiliary constituents to be as short as possible. This preference is realized in the occurrence of short discontinuous constituents in post-auxiliary position, while pre-auxiliary position has longer and more complex PPs and CPs.

troller of the missing subject of the embedded clause[9]. It also contains information about the nature of the dependency relationship between the matrix clause and the infinitival clause. The complementizer in Warlpiri embedded clauses is entirely consistent with the usual sense of complementizer, so I will consider it to be the head of the embedded clause (**CP**).

The infinitival marker contains the information that the clause is an infinitival (*ie.* not tensed). The category **Infl** is generally assumed to be the locus of tense information, so the infinitival marker will be treated as an instance of **Infl**.

The structure that emerges is shown in (26).

(26)

```
                    C″
                   /  \
                       C′
                      /  \
                    I′     C
                   /  \
                 V′     I
```
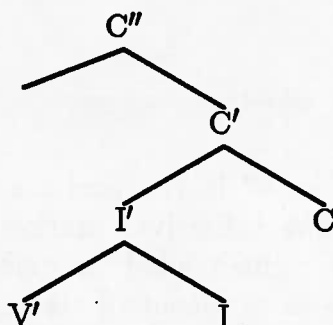
This structure is the right-headed analogue to the structure of tensed clauses.

The specifier of **CP** is reserved for the controlled empty category **PRO**. Using the abstract notion of spec-head agreement, the role of the complementizer in the resolution of the reference of **PRO** becomes more clear. The agreement requirement between the controller of **PRO** and the case selected by the complementizer could be viewed procedurally as forcing the index of **PRO** to be functionally determined by the head of **CP**. In the lexicon, the complementizer will be marked for the grammatical relation that it selects from the matrix clause. In resolving the reference for **PRO**, this information is used to construct the appropriate index on **PRO**.

The internal structure of **V′** is much the same in embedded infinitival clauses as it is in matrix tensed clauses. Arguments are presumably able to occur in free order, but most sentences have only one argument in the embedded clause. Examples (27) and (28) illustrate the general situation with arguments in embedded infinitival clauses in Warlpiri.

---

[9] Each complementizer selects a case (*eg.* ergative, absolutive, dative). The particular case that is selected will need to be marked in the lexical entry of the complementizer. The argument from the matrix clause that is marked with the case selected by the complementizer is construed as the subject of the embedded CP.

(27)

[karli   jarnti-rninja-rlajinta]
[boomerang   trim-inf-comp(reflex)]

            Ø-rna-ju   paju-rnu
AUX:perf-1sg(subj)-1sg(obj)   cut-past

'*I cut myself while trimming the boomerang.*'
  (Hale, 1981, p.5)

(28)

karnta   ka-ju   wangka-mi   [yarla
woman   AUX:pres-1sg(obj)   speak-nonpast   [yam-(abs)

  karla-nja-karra]
dig-inf-comp(subj)]

'*The woman is speaking to me while digging yams.*'
  (Hale, 1983, p.21)

One exception to the free order in **V'** is the position of the verb — the verb must occur immediately before the infinitival marker in embedded infinitival clauses. This means that **V'** is right-headed in embedded clauses. Another exception to the free order occurs in non-control clauses where there is an overt subject phrase. In such cases, the subject phrase must occur in the first position of the clause.

(29)

ngarrka-ngku   ka   karli   jarnti-rni --
  man-erg   AUX:pres   boomerang-(abs)   trim-nonpast --

kurdu-ku   maliki   wajilipi-nja-rlarni
child-dat   dog-(abs)   chase-inf-comp(oblique dat)

'*The man is trimming the boomerang, while the child is chasing the dog.*'
  (Hale, 1981, p.9)

(30)

ngarrka-ngku   ka   karli   jarnti-rni --
  man-erg   AUX:pres   boomerang-(abs)   trim-nonpast --

karnta-ku   kurdu-ku   miyi   yi-nja-rlarni
woman-dat   child-dat   food-(abs)   give-inf-comp(oblique dat)

'*The man is trimming the boomerang, while the woman is giving food to the child.*'
  (Hale, 1981, p.9)

In sentence (30), the first dative is unambiguously taken to be the subject of the embedded clause, even when the embedded verb subcategorizes for a dative argument (as *yi* does).

The specifier of **CP** gives just the right position for initial subjects in non-control clauses. In the case of these non-control situations, there would be no **PRO** in the specifier of **CP**. Notice also that the case suffix on the subject is not ergative as would be expected, but dative. The complementizer, *rlarni*, when in a control structure, selects an (unsubcategorized) dative from the matrix clause as the controller. The same sort of abstract requirement of spec-head agreement is consistent with the actual markings on the subject argument in non-control clauses. In this case, the agreement is realized through an overt marker rather than as a selected controller.

The general structure for embedded infinitival clauses is shown in (31).

(31)
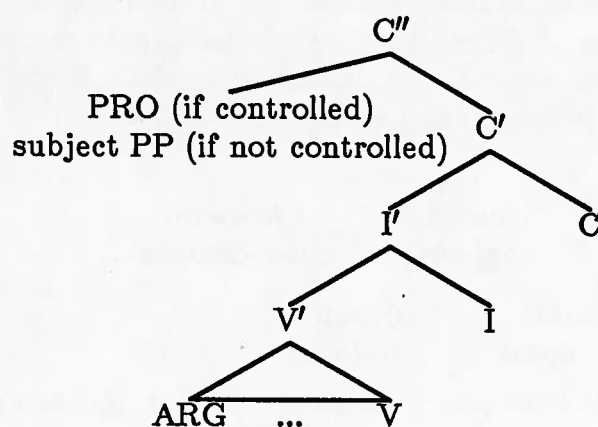
3. Complete Functional Complexes

The term *complete functional complex* refers to those constituents that have all of their arguments contained within them. This is generally assumed to single out **CP** and **NP** as complete functional complexes (Chomsky 1986), while, for example, **VP**, which may have an argument in the specifier of **IP** and hence external to itself, is not a complete functional complex.

Using this definition, Chomsky (1986) is able to indirectly specify the exceptional behaviour of **NP** and **CP** with respect to the theory of bounding. Where these were the bounding nodes for subjacency, they are now complete functional complexes which are assumed to never be adjunction nodes. In this way, we can move by adjunction over **VP** but this strategy does not work with movement outside of **NP** or **CP**.

The question then arises whether **NP** and **CP** are universally complete functional complexes, or if this is only relevant for languages like English. Considering the case of Warlpiri and the general definition of complete functional complexes, it would seem that since arguments are identified in local GR-NonConfigurational relations in sister positions of the verb, that the Warlpiri verb phrase must contain all of its arguments. Since there are no argument positions defined external to **VP**, then **VP** must be a complete functional complex. If this is the case, and if complete functional complexes cannnot be

adjunction nodes, then the adjunction analysis proposed for movement to the specifier of **CP** proves unworkable for Warlpiri (since it required adjunction to **VP**). This would lend indirect support for the degenerate projection hypothesis.

The situation with **NP** and **CP** in Warlpiri is difficult to assess. Since discontinuity is allowed in Warlpiri, it is possible to have a **CP** that does not contain all of its arguments. Presumably, Warlpiri **NP**s that are discontinuous do not contain all of their arguments either. At the level of representation where discontinuity is represented, **CP** and **NP** need not be complete functional complexes[10].

If **CP** is not a complete functional complex then we would expect to find adjunction structures with **CP** as the adjunction node. This adds some support to Hale's analysis of tensed relative clauses in Warlpiri (Hale, 1976). The analysis that Hale proposes is for adjunction of the dependent clauses in sentences with multiple tensed clauses. So, in the examples of (32) and (33), the structures of (34) and (35) (respectively) would be used.

(32)

natjulu-lu	kapi-na	maliki	luwa-ni,
I-erg	AUX	dog(-abs)	shoot-nonpast,

katji-∅-nki	yalki-ni	njuntu
COMP-AUX	bite-nonpast	you-(abs)

'*I will shoot the dog that bites you/I will shoot the dog, if/when it bites you.*' (Hale, 1976, p.80)

(33)

maliki-li	katji-∅-nki	yalki-ni	njuntu,
dog-erg	COMP-AUX	bite-nonpast	you(-abs)

natjulu-lu	kapi-na	luwa-ni,
I-erg	AUX	shoot-nonpast

'*I will shoot the dog that bites you/I will shoot the dog, if/when it bites you.*' (Hale, 1976, p.80)

---

[10] An interesting issue is whether the property of being a complete functional complex is one of node types or a local property of specific nodes. Are all NPs and CPs not complete functional complexes in Warlpiri, or are only discontinuous NPs and CPs not complete functional complexes while the continuous ones are? If there is such a local distinction in the property of being a complete functional complex, one would expect to find systematic differences between continuous and discontinuous expressions in Warlpiri. I know of no evidence that shows any such differences.

(34)

```
              CP
             /  \
          CP      CP
    dependent  main clause
     clause
     adjunct
```

(35)

```
              CP
             /  \
          CP      CP
    main clause  dependent clause
                 adjunct
```
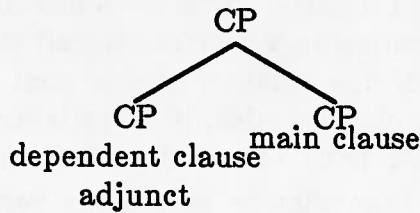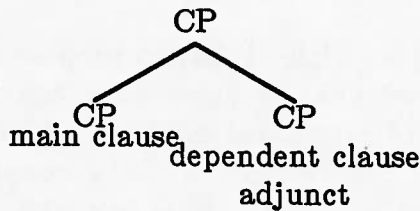
4. Principles of Grammar — Syntax or Morphology

In much of the above disussion of Warlpiri syntax, it appears that Warlpiri is immune from the full force of some of the commonly accepted universals of linguistic theory. The θ-Criterion is compromised to account for Warlpiri, and the Case Filter is rendered virtually vacuous. In this section I wish to discuss the principled nature of this exceptionality; in particular addressing the insight that Warlpiri grammar can give us regarding the complementary nature of the syntactic and the morpho-syntactic components of the grammar.

The Case Filter simply requires that overt noun phrases be in a particular structural configuration to a case assigning element. I have suggested that Warlpiri noun phrases are sub-constituents of P′ constituents that are headed by the morphological case marking affix. The word formation rule that affixes the case marker can be readily viewed as the morphological realization of the Case Filter. In this way, the Case Filter functions within the morphology, defining well-formedness prior to the insertion of a word into the syntactic phrase-marker.

Any discussion of the status of Warlpiri with respect to the θ-Criterion would be incomplete without addressing the *Extended Projection Principle*. The *Extended Projection Principle* is a requirement that all VPs have subjects. With *subject* defined in terms of structural configuration, it appears that something has to go — either the *Extended Projection Principle* (at least as far as the language Warlpiri is concerned) or the analysis of Warlpiri syntax presented above. It is clear that the basic tenet that I maintain about Warlpiri syntax is that there are no structural configurations that can be used in thematic identification — hence, no subjects.

There is, however, some evidence in Warlpiri that would suggest an asymmetric treatment of subjects and objects. This evidence has been used to argue for a type of structural configuration in Warlpiri. Recall that in embedded relative clauses it is the *subject* of the relative clause that is controlled by the selected element from the main clause. Also, if there is any order restriction in embedded relative clauses, apart from the subject first requirement, it is that the object of the clause occur immediately before the verb[11]. Also recall that the Warlpiri auxiliary exhibits a nominative/accusative distinction in the person and number markers that occur on it — in fact, making a distinction of subject and object.

These asymmetries have led Hale (1983) to propose a level of Lexical Structure (LS) — an abstract level that is universally configurational and exhibits subject/object asymmetries of structural position. I maintain that we do not need an abstract level of representation to account for the facts about subject/object asymmetries in Warlpiri. What is common to all the instances of apparent subject/object asymmetry is that they occur in constituents that are constructed according to the rules of the morpho-syntax rather than the syntax proper. This is particularly striking in the case of relative clauses, as compared to syntactic matrix clauses. We have seen that relative clauses are strictly right-headed (**Infl**, **Comp**, and **V** occur on the right periphery), whereas main clauses are left-headed (**Infl** and **Comp** occur at the leading edge, modulo movement to initial position). Right-headed structures are consistent with the general rule of word formation — this switch from right to left-headed structures can be viewed as resulting from a different component being responsible for the structural well-formedness of these two types of constituents.

Similarly, within the auxiliary we find evidence for a subject/object distinction in the person and number markers. These markers occur as part of the auxiliary word, so again would come under general well-formedness of the morpho-syntax rather than the syntax proper.

Within the main clause where the constituents are subject to well-formedness of the syntax (as opposed to the within word morpho-syntax), we find no evidence for any asymmetry of subjects and objects. When an embedded relative clause subject is controlled by a matrix clause argument, the argument is selected from a range of argument cases, with no special distinction made for control by matrix clause subjects.

What emerges is an analysis wherein the syntax and the morpho-syntax together can ensure that the principles of universal grammar are satisfied. They work together in a complementary nature, conspiring, as it were, to satisfy all principles of grammar. Across languages we can find great variability as to the relative burden placed on the morpho-syntactic as opposed to the syntactic

---

[11] This is difficult to verify from the published data since most of the embedded relative clauses only contain what would be the direct object argument, and hence it's occurrence consistently adjacent to the verb may be an artifact of data set limitations.

component. In Warlpiri we find a large burden placed on the morpho-syntax, while the syntax itself exhibits greater freedom and flexibility of well-formedness. What this amounts to, in essence, is parametrizing certain principles of grammar as to the domain of application — morpho-syntax, syntax or both. Certain universal properties guide, in general, the nature of representations in each of these domains. So, for example, the right-hand head rule will be a general guiding principle of morpho-syntactic well-formedness.

It remains to address just how the θ-Criterion fits within this schema. Suppose that the θ-Criterion were a principle that, in Warlpiri, applies within the morpho-syntax. Note that person and number markers in the auxiliary word are obligatory, even if the syntactic arguments themselves are missing. So, a Warlpiri sentence can consist of only an auxiliary together with a verb, where the person and number markers on the auxiliary provide a pronominal interpretation. Each person and number marker is unique — that is, we do not find multiple copies of a particular person and number marker in the same auxiliary, even if the syntactic argument that it agrees with is discontinuous. The θ-Criterion would accurately account for the obligatory and unique nature of these markers. However, the auxiliary is not the verb of the sentence, and if the θ-Criterion is applying to account for these markers within the morpho-syntactic constituent, then the verb simply isn't a part of this constituent. It is questionable how the θ-Criterion can apply in these circumstances, in particular because, in the absence of a verb, there is no applicable θ-grid. I would like to sketch a speculative solution to this problem.

Suppose that the Warlpiri auxiliary instantiated a very general θ-grid — selected from a small set of possible θ-grids covering major predicate types (eg. intransitive, transitive, bitransitive). This essentially amounts to the auxiliary being a predicate in its own right, only with very general semantics[12]. The θ-Criterion could then apply in the morpho-syntax within the auxiliary word, ensuring that all necessary person and number markers were present. The auxiliary is an instance of a syntactic atom of type *Infl*, and would, in this analysis, have associated with it a general θ-grid instantiated with person and number markers. Within the syntax, the verb in VP would have a specific θ-grid that would be instantiated at the level of VP with specific arguments, if present, and possibly indexed to multiple pieces of discontinuous expressions. So, at VP we have a specific θ-grid, possibly with gaps in it where a subcategorized argument is not realized as a constituent in VP. As a percolation convention, we would need to compose the specific θ-grid on VP with the general θ-grid on the auxiliary word, ensuring that the person and number features agreed and instantiating each slot in the resulting composed θ-grid with the extensional intersection of the corresponding slots in the θ-grids of the auxiliary and V. In most cases,

---

[12] Supporting evidence for this is found in nominal clauses in Warlpiri, where a main clause has no actual verb. The auxiliary is present in such cases, and the meaning of the nominal clauses is consistent with the auxiliary being a very general predicate (eg. do something).

extensional intersection will amount to just feature agreement. However, when a subcategorized argument is not realized as a constituent in **VP**, extensional intersection of the unspecified slot with the features of person and number will result in just those features of person and number, and hence give us the pronominal interpretation of these cases[13]. In this way, the percolated θ-grid is guaranteed to satisfy the θ-Criterion if the θ-Criterion defines well-formedness of the auxiliary in the morpho-syntax.

5. Summary and Overview

The main difference between the analysis presented above and most syntactic analyses is the extent to which the morphology is used to construct constituent structure and define well-formedness. The general theory that was proposed involved syntactic structure to account for the location of the complementizer and auxiliary in tensed clauses and morpho-syntactic structure to account for the internal argument structure. If we look at a general tree structure of a Warlpiri sentence, the proposed analysis can be viewed as in (36).

---

[13] In the parser implementation this will need to be further refined. Since third person singular marking is null in Warlpiri, this leads to ambiguity about the type of general θ-grid on the auxiliary, since we cannot tell immediately whether there is a null object and a null subject marker on the auxiliary, or if there is just a null subject marker in an intransitive verb type. To avoid such ambiguity, we need to require that the actual predicate agree in type with the auxiliary θ-grid — that is, we can only compose like types of θ-grids. In a left-to-right deterministic parse, if we encounter the auxiliary prior to the predicate (as occurs in the vast majority of Warlpiri clauses), we will not be able to determine the appropriate type for the general θ-grid at the necessary time as far as the deterministic parser is concerned. To handle this situation, I instantiate a minimal general θ-grid for the auxiliary consistent with the actual markings that occur on it. The composition of this θ-grid with that of the VP can potentially involve composition on unlike θ-grids. In such cases the θ-grid on VP is dominant and defines the actual slots that percolate. Any subcategorized slots in the composed θ-grid that are uninstantiated are simply instantiated by a default rule that adds the third person singular features to such slots.

(36)

SYNTAX

MORPHOLOGY

```
                          C″
                         /  \
                          C′
                         /  \
                        C    I′
                            /  \
                           I    V′
                               /|\
                            P′ ... V ... C″
                           /  \        /  \
                          N′   P          C′
                         /_\             /  \
                        N ... N         I′   C
                                       /  \
                                      V′   I
                                     /|\
                                   P′ ... V
                                  /  \
                                 N′   P
                                /_\
                               N ... N
```
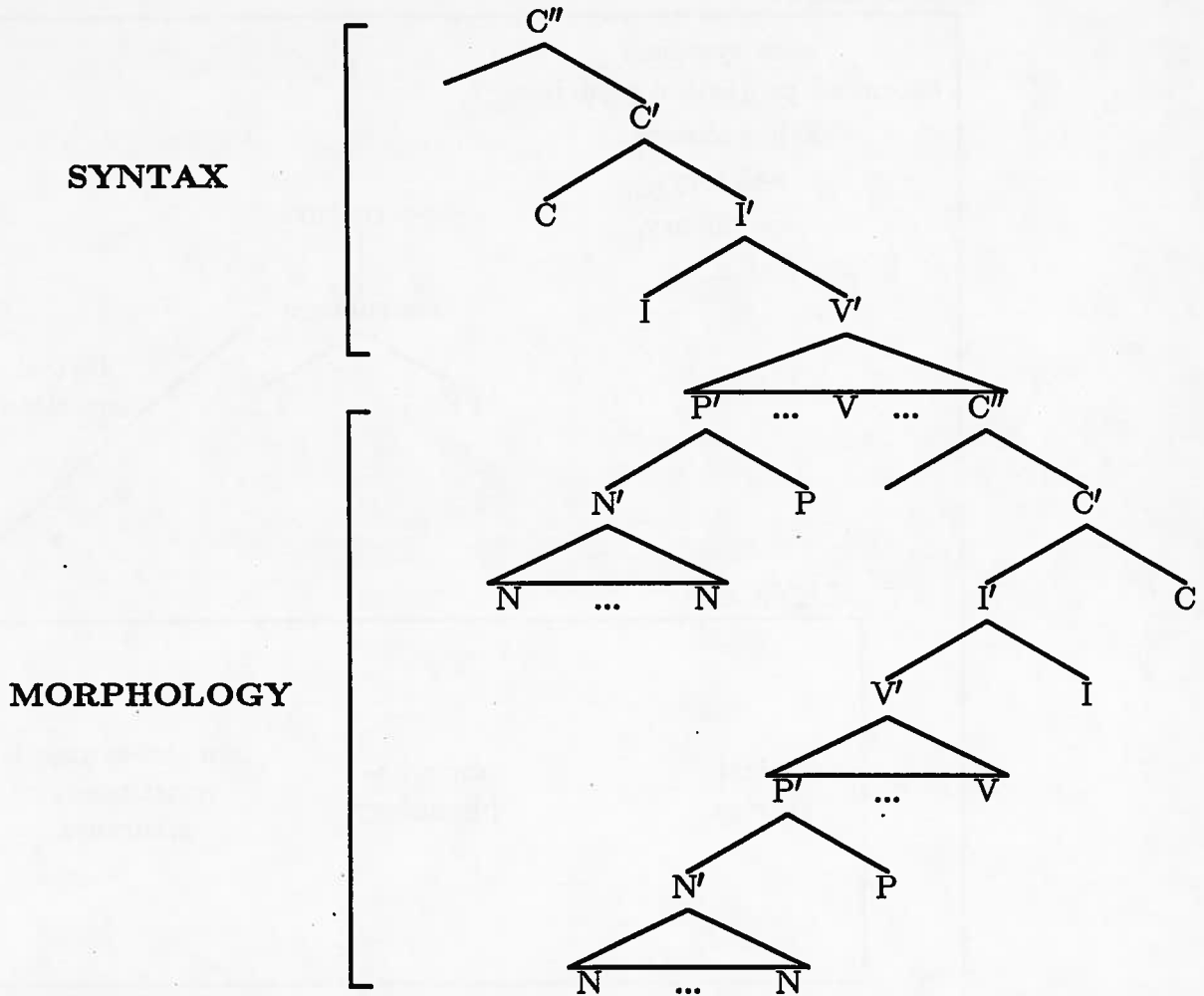
Notice that the level of structure where the two systems meet is **V′**, the only constituent that does not have a clear directionality of headedness. It is also interesting that this constituent is the locus of free word-order and discontinuity.

Since the θ-Criterion has been changed to a condition on percolated argument structures, it remains a general theorem of grammar. The percolation conventions are parametrized: languages may restrict certain arguments to be in a biunique relationship with syntactic positions, or they may allow non-biunique relationships (allowing for optional arguments and discontinuities).

Case theory, we have seen, cannot be violated by Warlpiri syntax. Case assignment, and presumably the case filter itself, is a condition that applies more aptly to Warpliri morphological structures. If we view the theory of grammar as in (37) then, by allowing core systems of grammar to be indexed to the particular component of grammar (*eg.* the lexicon, the syntax, the entire grammar), we do not need to burden the entire grammar with mechanisms that only apply in a particular component. The indexes in (37) are set for Warlpiri.

(37)

GRAMMAR

core systems:
extended projection principle$_L$
X-bar theory$_G$
θ-theory$_G$
case theory$_L$
...

d-structure

|

s-structure

PF   LF

lexical
insertion

LEXICON

lexical
entries

morpho-
phonology

morpho-syntactic
constituency
structure
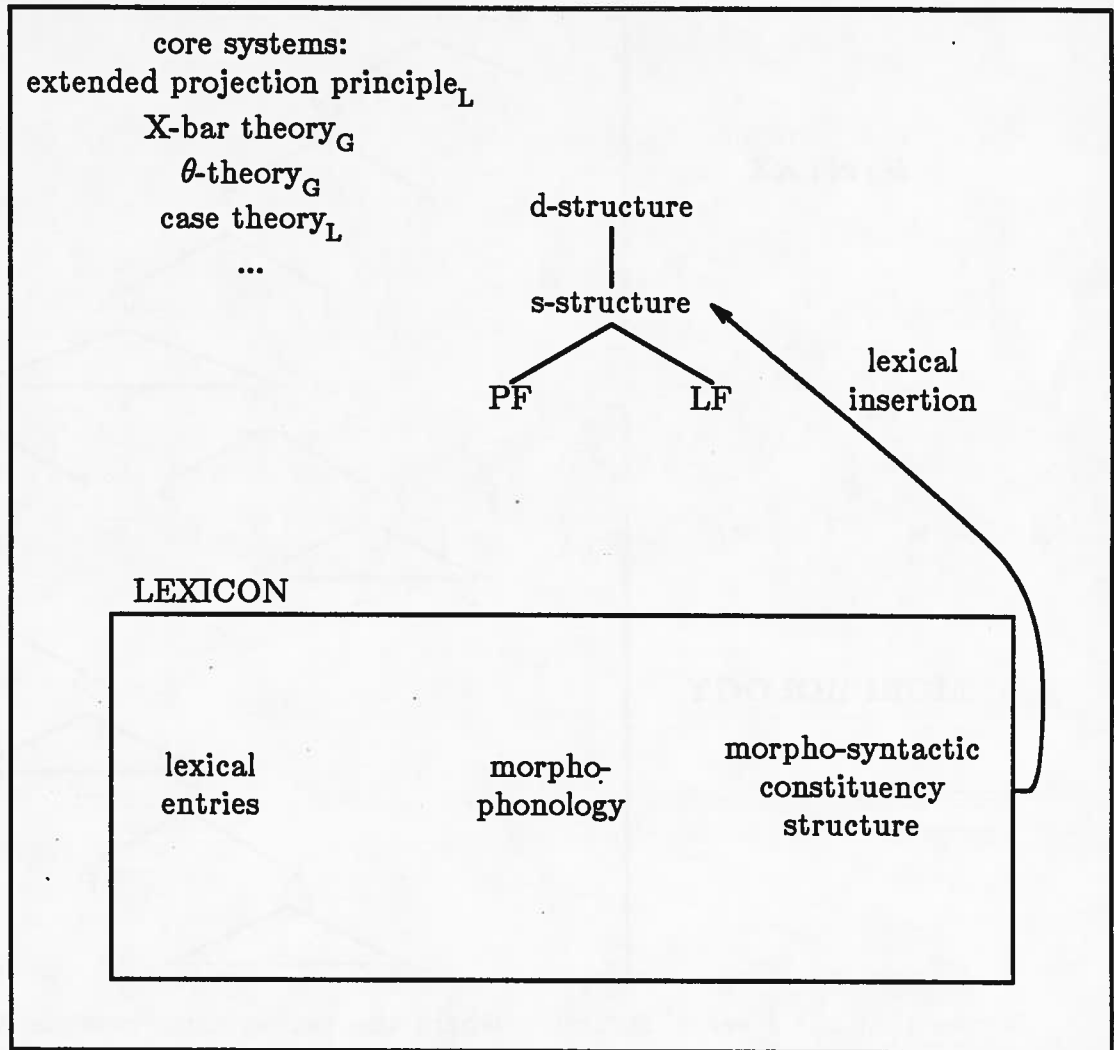
The following summarizes the four major changes to the theory of grammar that have been proposed to account for Warlpiri syntax:

1) X^{MAX} may be X' (and this may be *degenerate* with respect to the theory of bounding).

2) A greater burden of constituent structure and well-formedness has been put in the morpho-syntax.

3) Subcomponents of grammar may apply within the morpho-syntax.

4) The θ-Criterion has been replaced by a constraint on percolated argument structures, with parametric percolation conventions.

CHAPTER 4

A Parsing Model for Warlpiri

This chapter presents the parser that has been developed for Warlpiri. The theory outlined in the two preceding chapters was developed with parsing considerations in mind, particularly keeping in mind the degree of specification that a parsing implementation requires. In approaching a language from a theoretical point of view alone, it is much easier to isolate and describe a particular phenomenon without ever specifying the theory to handle a wider range of phenomena. Often the limitations of the theory go unnoticed since, it is common practice to only hand-test the theory for linguistic data coverage. A parsing model, however, forces a considerably greater degree of specification and explicitness in the theory, and provides a ready test of the coverage of the theory.

It could be argued that the required degree of specification for a parser is more detrimental to the progress of the theory, forcing *ad hoc* specifications of theory prematurely. This need not be the case if the implementation keeps separate those routines that are theoretically founded from those routines that are merely necessary to allow the parser to deal with a phenomenon in the absence of any linguistically well-founded account of that phenomenon. In fact, such a modular approach to the development of a parsing model can provide indications as to just where the theory itself breaks down in the coverage of linguistic data. By keeping separate the understood phenomena from the waste basket of *ad hoc* stipulations, the development of a parser can serve as a tool in the extension of theory to actually account in a principled way for a wider variety of linguistic phenomena.

The parser that has been implemented is, as far as possible, a transparent mapping of the theory that has been developed. Of course, the static statements of well-formedness found in the linguistic theory have been translated into dynamic procedures that actually construct representations that satisfy the well-formedness conditions; however, the modularity of the theory and division of labour among the various principles of grammar has largely been preserved. The strengths of the current GB theory were particularly noticeable when a routine designed to handle one particular phenomena could be reused, without any major modifications, to handle a variety of phenomena that the theory, itself, predicted to be related. Equally important, however, some of the limitations of current linguistic theory are highlighted by the difficulties and stop-gap measures that were required in the implementation.

1. Overview of the Parser

The parser is written in Zetalisp and implemented on the Symbolics Lisp Machine.

The input to the parser consists of a series of morphemes, separated by '-', as well as punctuation markers. The punctuation markers mark the endpoints of intonational phrases, as well as the endpoint pause of the falling/rising melody intonation that indicates adjunction (for topics and adjoined clauses).

The parser assumes that the input is a complete sentence, and returns a phrase marker for the input sentence — complete with syntactic and morpho-syntactic structure. It also returns the percolated θ-grids for each clause of the input sentence.

The parser operates in a left to right deterministic fashion, constructing syntactic and morpho-syntactic representations in the same pass of the sentence. The morphology is not preprocessed except to the extent that morpheme boundaries are provided in the input. Although this is somewhat of an artificial abstraction from real speech and text, it can be justified in the same way that spaces can be justified in multiple word sequences, or the use of the apostrophe can be justified in parsing English possessive constructions. Presumably, phonological cues can aid in the location of morpheme boundaries. Principles of Lexical Phonology (Kiparsky, 1982, Mohanan, 1982 *inter alia*) provide a differential treatment of morpheme edges in the application of phonological rules, and this could be used by a morphological processor to narrow down the choice of morpheme boundary positions. Similarly, morpheme structure constraints would provide valuable information about possible locations of morpheme boundaries. For example, with a very limited class of exceptions, all Warlpiri morphemes are independently syllabifiable (Nash, 1980). To make full use of this sort of information would involve encoding phonological information in the input. It is unclear whether a morphological processor that located morpheme boundaries as well as constructing morpho-syntactic representations could operate deterministically[1] even if provided with phonological information. This is certainly beyond the scope of this thesis where the input to the parser — an annotated orthographic abstraction — allows these issues to be side-stepped and earmarked for future consideration.

Unlike standard deterministic parsers, there is no buffer (as such) in the parser. There is, however, a mechanism that allows a single completed morpho-syntactic unit (*ie.* a single syntactic atom) to be held in the construction of the syntactic representation.

The parser is implemented to be data driven as well as expectation driven. A particular parser state can create expectations (consistent with the principles

---

[1] A problem for extending conventional deterministic parsing algorithms to such domains as phonology is deciding just what counts as buffer contents and what size the buffer should be. Three positions works for syntactic parsing, but other linguistic domains may require different sized buffers. When the input is a continuum of speech sounds, there is an

of linguistic theory) that dictate subsequent parser states. Similarly, as each morpheme is processed, expectations are set that also dictate the course of subsequent processing.

Data structures are defined to enforce structural well-formedness, with a separate data structure for each unique syntactic and morpho-syntactic configuration type. So, for example, if a morpho-syntactic complementizer suffix is encountered, it is incorporated into the parse first in a schematic data structure consistent with the rules of morpho-syntactic projection. Lexical selectional properties of complementizers are retrieved from the lexicon and these set expectations that are used to either unify the schematic structure with a previously parsed constituent (normally the case for right-headed configurations) or to provide direction in subsequent stages of the parse.

Thematic projection occurs while constituents are constructed in the bottom up part of the parse. The default θ-grid of the auxiliary is instantiated when the auxiliary is processed, and it is composed with the projected θ-grid of the VP when VP is incorporated into the IP constituent.

2. Prosody and Determinism

As mentioned in the previous section, the parser makes use of a limited amount of prosodic information. This prosodic information is encoded into the input as edge-marking symbols inserted into the input sequence. The prosodic information that is available to the parser consists of edges of phonological words (marked by spaces in the input), intonational phrase edges (marked by * in the input), and 'adjunction' phrase edges[2] (marked by ! in the input).

In Warlpiri, there are constructions that appear as if they should be ambiguous, but prosodic information resolves the apparent ambiguity.

Consider, for example, a sentence initial sequence as in (1).

(1) **PP PP ...**

In such a sequence, the parser would normally infer a null auxiliary base with default person and number markings, since maximally one constituent is permitted in pre-auxiliary position. Thus, under normal circumstances, the sequence in (2) is ill-formed.

(2) **PP PP AUX**

There is, however, one construction that does permit the sequence in (2); namely, if the first **PP** is set off by a falling/rising melody intonation ending in a pause — that is, if it is marked prosodically as a topic. So, the sequence in (2) is well-formed if and only if it is intonationally marked. Without access to the prosodic information, the parser would not have any way of resolving whether (2) was well-formed or ill-formed. To distinguish these two cases, the end point pause of the falling/rising melody intonation is marked. Thus, (3) is ill-formed,

---

additional problem of segmentation of this continuum into buffer positions.

[2] *ie.* the end point pause of the falling/rising melody that indicates an adjoined element (either a clause or a topic).

while (4) is well-formed.

(3) **PP PP AUX**
(4) **PP ! PP AUX**

Similarly, in cases of an auxiliary that may occur either sentence-initially or in second position, sequences as in (5) present another ambiguity.

(5) **PP AUX**

In such cases, it is impossible for the parser to know whether this is an instance of a second position AUX with the initial PP in the spec of **CP** (6), or a first position AUX with an adjoined topic in sentence-initial position (7).

(6)

```
              C''
           /        \
         PP           C'
                    /     \
                  C         I'
                  Ø       /    \
                        I       ...
                        |
                       AUX
```

(7)

```
              C''
           /        \
         PP           C''
                       |
                       C'
                    /     \
                  C         I'
                  Ø       /    \
                        I       ...
                        |
                       AUX
```
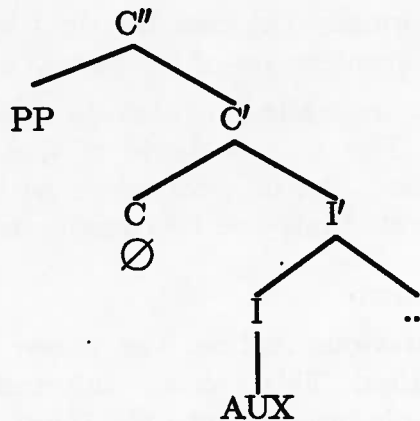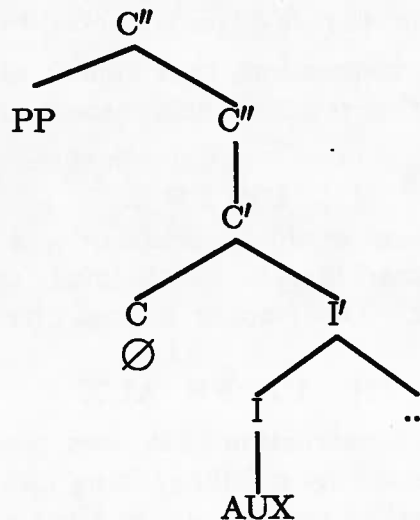
Once again, these cases are distinguished intonationally by the adjunction intonation. In order to distinguish these cases in the parser, the input corresponding to (6) is (8), while that corresponding to (7) is (9).

(8) **PP AUX**
(9) **PP ! AUX**

Adjunction intonation can also set off an entire clause. Such cases are treated structurally the same as **PP** topic adjuncts, and they are also marked intonationally in the same way[3].

Another situation arises that requires the use of prosodic information in order to preserve the deterministic parse. This ambiguity is introduced by the phonologically non-overt absolutive case marker. In general, a noun may modify another noun[4], in which case the two nouns are placed in sequence with the rightmost noun bearing the case marker, as in (10).

(10)

```
            P'
          /    \
        NP       P
       /  \
      N    N
```

However, if just presented with the sequence in (11),

(11) **N N P**

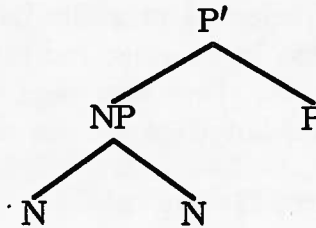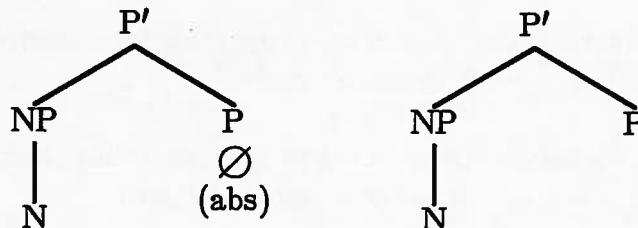there are two possible syntactic structures — the one in (10), or a structure in which the first noun of the sequence is parsed as an absolutive (null case marked) **PP**, separate from the second overtly case marked **PP**. The structure corresponding to this is shown in (12).

(12)

```
          P'                          P'
        /    \                      /    \
      NP      P                   NP      P
      |       Ø                    |
      N     (abs)                  N
```

Preliminary evidence suggests that the intonation of Warlpiri resolves this apparent ambiguity as well. It seems that if the structure in (10) is correct, the sequence in (11) is produced as a single intonational phrase; if the structure in (12) is correct, it is produced as two intonational phrases. To resolve this, the input sequence is annotated with an intonational phrase boundary marker (*).

---

[3] With **PP** topics, it is normal for the topic to be repeated within the clause. I have no account of topics, and merely treat them as adjoined to CP, as if base-generated there. The parser does not coindex a topic and a trace or constituent in the clause. As it turns out, thematic percolation provides the appropriate percolated θ-grids without necessitating any fully worked out theory of topicalization.

[4] Recall that there is no class of adjectives in Warlpiri, so nominals can serve the functional role of adjectives in

73

Thus, the input sequence in (13) corresponds to the structure in (10) , while the input sequence in (14) corresponds to the structure in (12).

(13) N N P

(14) N * N P

Certainly, further testing of the intonational facts is required to determine if these intonational cues are reliable, but the preliminary findings presented here do argue for the crucial use of prosodic information in a deterministic parsing model of Warlpiri. Similar results obtain regarding the need for intonational information to deterministically parse Japanese[5].

The boundary approach to what is really a prosodic continuum amounts essentially to a diacritic use of selected prosodic information. This is clearly a stop-gap measure, but its success in practice indicates that, although the form of this approach may be incorrect, there is a need for access to prosodic information that any complete natural language system cannot ignore.

3. Lexical Representation and Retrieval

The basic lexicon consists of a list of words and morphemes, each with their own properties. Each lexical item is classified as to its lexical category or part of speech, followed by any idiosyncratic properties of the word. Some of these properties are required for items of a particular category — for example, each item with the category, **verb**, must have an associated θ-grid.

Each lexical entry is further classified as either a syntactic atom, a morphological unit, or both (as is the case with clitics that are subject to syntactic as well as morpho-syntactic well-formedness). These are represented as in (15).

(15)

(<item> <category> general-properties (:syn syntax-specific-properties))
             — *a syntactic atom*

(<item> <category> general-properties (:morph morpho-syntax-specific-properties))
             — *a morpho-syntactic unit*

(<item> <category> general-properties (:syn syntax-specific-properties)
                      (:morph morpho-syntax-specific-properties))
             — *eg. a clitic*

---

modifying other nominals.

[5] M. Marcus and D. Hindle, personal communication.

The atom *:syn* indicates that the lexical element is to be incorporated into the parse consistent with the general rules of syntactic well-formedness, while *:morph* indicates that the lexical element is to be treated according to the general rules of morpho-syntactic well-formedness. Dual status items that are categorized as both *:syn* and *:morph* will be subject to well-formedness conditions of both the syntax and the morpho-syntax.

Any idiosyncratic properties associated specifically with the syntactic or morpho-syntactic well-formedness of constituents involving this lexical item are listed in the item's *:syn* and *:morph* lists, respectively. If there are no idiosyncratic properties appropriate to this level of lexical representation, then these lists are represented as the singleton lists *(:syn)* and *(:morph)*, and the general principles of syntactic and morpho-syntactic well-formedness apply.

In addition to the lexicon of words, there is a second lexicon that lists the lexical categories, together with their general properties. Again, the use of *:syn* and *:morph* lists distinguishes purely syntactic categories from morpho-syntactic categories, and these can each have specific idiosyncratic properties. Items in the category lexicon are represented as in (16).
(16)

> (<cat> general-properties (:syn syntax-specific-properties)
>   (:morph morpho-syntax-specific-properties))

The lexicon is implemented in LISP as property lists. When a lexical item is to be looked up in the lexicon, all that is required is to retrieve the property list using the item, itself, as the key. If this list is null, then the item is not in the lexicon, otherwise the property list is returned. The category label of the lexical item functions as a pointer to the category lexicon, where the entry is retrieved that contains the general properties of that category. These general category properties are united with the specific lexical properties of the item, yielding a more complete property list for the item. This, in turn, is united with the general properties of syntax or morpho-syntax, as applicable. As a general principle, any specific properties of the lexical item override any counter-indicative properties of the category — in essence, the Elsewhere Condition of Kiparsky, 1973 (*inter alia*). Similarly, any specific property in the *:syn* or *:morph* list of a category will override any counter-indicative general property of syntax or morpho-syntax. What emerges is a hierarchy of specification of properties as in (17), with conflicting properties resolved in favour of the more specific value.

(17)

```
┌─────────────────────┐
│    properties of    │
│   lexical items     │
└─────────────────────┘


┌─────────────────────┐
│    properties of    │
│  lexical categories │
└─────────────────────┘


┌──────────────────────────┐
│      properties of       │
│  syntax/morpho-syntax    │
└──────────────────────────┘
```

3.1. Sample Lexical Entries

The following are some sample lexical entries:

('karli N (gloss 'boomerang) (:morph))

('ka   I (gloss 'aux:present) (:syn)
                          (:morph (second-position)))
        — a second position auxiliary clitic

('jarnti V (gloss 'trim) (V2) (case (erg abs)) (θ (agent theme)) (:morph))
        — a verb of the second conjugation class (Nash, 1980)

('kurra C (gloss 'comp) (:morph (control object)))
        — a complementizer suffix indicating object control of the embedded
            subject

('ninjtja I (gloss 'infinitive) (:morph (1-bar) (binary)
                              (c-select oblig VP)))

('ngku   P (gloss 'erg) (:morph))

('rni tense (gloss 'nonpast) (V2 V4) (:morph))
        — *tense* is a non-projecting category that converts a
            morphological verb into a syntactic verb.

('*   inton (:syn))
        — the intonational phrase edge marker

The following are some sample entries in the category lexicon:

```
('P  (c-select oblig NP CP)          — complement selection of an
                                        obligatory NP or CP
     (non-overt (gloss 'abs))        — there is a non-overt surface variant,
                                     glossed by 'abs'
     (binary)                        — binary selection
     (1-bar))                        — projects to 1 bar-level


('I
     (:morph (n-ary)
             (0-bar)
             (c-select opt aux-morph))
      (:syn (binary)
            (1-bar)
            (c-select oblig VP)
            (non-overt (gloss 'perf))))


('V
     (:morph  (0-bar)
              (binary)
              (c-select opt tense))
      (:syn  (1-bar)
             (n-ary)
             (c-select opt PP CP)))


('N
     (:morph (n-ary)
             (1-bar)
             (c-select opt N)))
```

4. Implementation of X-bar Theory

X-bar theory is implemented by a general mechanism that interprets the properties of a category in the category lexicon consistent with either a morpho-syntactic or a syntactic projection. The X-bar mechanism contains procedures that handle projection of the category, and selection of complements and specifiers.

Each category in the category lexicon is specified as to whether it projects to a single bar-level or two bar-levels[6]. In addition, categories are specified if there is a phonologically non-overt item of this category. Category lexical entries indicate the type of branching (binary or n-ary) that the category allows

---

[6] Actually, only those that project to a single bar-level are marked in the lexicon, while the default unmarked assumption is that projection is to two bar-levels.

in its projection. The normal situation is binary branching (*ie.* a unique sister constituent for each bar-level). N-ary branching indicates a category that allows iterative complements or modifiers.

Categories are also specified for selectional properties of complement and specifier postions and optionality of these selectional properties. The selectional properties contain a possible disjunction of expected elements for the complement position, and a similar disjunction for the specifier position, if applicable.

When a lexical item is encountered, it is projected to the specified maximal level in accord with the general X-bar schema, the general properties of syntactic or morpho-syntactic projection (as the case may be), and any category specific properties. Selection properties of the lexical item or the category set expectations for complements (and specifiers, if applicable). These expectations can be optional or obligatory — obligatory expectations dictate successive parser actions.

Expectations for a maximal category result in a sort of inverse projection to the lexical category.

In general, the difference between the morpho-syntactic and syntactic X-bar interpretation is a matter of directionality of headedness. Morpho-syntactic representations are constructed by the X-bar projection component as right-headed structures; syntactic representations are constructed as left-headed structures. The switch from left to right headed structure corresponds, in the parser, to a division between expectation driven processes (top down) and data driven processes (bottom up). The uppermost clause structure (*ie.* the complementizer and auxiliary in a tensed clause) is parsed in an expectation driven fashion from an initial state where the parser expects a C''. At the locus of interface of these two components — the level of V' — there is no clear headedness of constituency. In order to parse the V' deterministically, a set of necessary and sufficient leading edges for V' were determined. Encountering a leading edge for V' is, then, a sufficient condition to posit a V' constituent. The leading edges for V' are any sequence of two consecutive constituents that could be in V'; namely, **V**, **PP**, or **CP**.

The X-bar mechanism described here, together with the category lexicon, comprises the heart of the phrase structure grammar of the parser.

5. The Structures

One of the main differences between the design of this deterministic parser and PARSIFAL (Marcus, 1980) is the parser *structures*. Where PARSIFAL has a single data type and a fairly large inventory of rules, the Warlpiri parser presented here has a set of data types — each defined by linguistically relevant and distinct structural conditions — and a smaller set of principles that are interpreted by these *structures*

The *structures* keep separate the following configurations:

    i)   Syntactic adjunction
    ii) Syntactic modification
    iii) Syntactic complementation
    iv) Morpho-syntactic modification
    v) Morpho-syntactic complementation
    vi) Morpho-syntactic closed constituents

5.1. Syntactic Adjunction

The syntactic adjunction *structure* is the most limited in application of all the *structures*. The syntactic adjunction *structure* is defined to construct and contain only those configurations compatible with the general definition of structural adjunction. It is activated under two separate conditions: when the adjunction intonational marker (!) is processed, and when the trace-handling routines require trace adjunction.

When an adjunction intonational marker (!) is encountered in the input, the preceding constituent is closed (*ie.* all obligatory expectations are satisfied and all optional expectations are discarded). This constituent, together with the top-down expected projection, is passed to the adjunction structure. The projection is first tested to determine if it is a possible adjunction node[7] and if it passes this test, the constituent and the maximal projection node instantiate the adjunction structure as in (18).

(18)

$$
\begin{array}{ccc}
 & \mathrm{XP} & \\
 & \diagup\;\diagdown & \\
\mathrm{Y} & & \mathrm{XP} \\
\end{array}
$$

where XP stands for the maximal projection node to which the constituent, Y, is adjoined.

Topics and adjoined relative clauses in Warlpiri can also be right dislocated, that is, post-posed at the end of the sentence. Such cases are presumably also cases of adjunction, but these cases are not handled by the Warlpiri parser.

The other condition that activates the adjunction *structure* is the trace-handling routine. This is used minimally in Warlpiri; however, in general, trace-handling adjoins the trace of a maximal projection to an embedded maximal constituent (if possible) whenever that level of embedding would produce a barrier between the antecedent and a trace in the embedded constituent. The trace and maximal constituent are passed off to the adjunction structure — the adjoined construction is created and returned to the same stage in the parse, from which processing resumes as before. If the trace is a trace of a zero-level

category, then the trace and the head of the constituent are passed off to the adjunction *structure*, and the resulting adjoined head is returned as the head of the constituent.

5.2. Syntactic Modification

The syntactic modification *structure* is used to construct syntactic fan-like constituents where more than just a single sister constituent are allowed. This is used, for example, to handle constituents within V', as well as iterative modifiers and non-subcategorized adverbials. The expectations for this structure tend to be rather permissive and are usually optional (unless the lexical entry of the head of the construction dictates otherwise). Essentially, this *structure* is where syntactic constituents are put after their complements are parsed, but prior to any assurance that the constituent can be closed. So, for example, we find the syntactic modification *structure* in the state shown in (19) in the parse of the sentence *karnta ka-ju wangka-mi yarla karla-nja-karra* 'The woman is speaking to me while digging yams.' (example sentence (28) from chapter 3). The parser has just parsed *wangka-mi*, the verb of the sentence, and is proceeding left to right.

(19)
Syntactic Modification Structure:
optional expectations: PP CP
obligatory expectations: nil
content: ...$(V'$ (PP t_i) (V wangka-mi) +)

The '+' in the above indicates the location in the content where the optional expectations would be located. The optional expectations list indicates a disjunction — either a **PP** or a **CP** is allowed as a continuation of the constituent. Once a optional expectation is satisfied, it remains as an optional expectation. This allows for iterative modification until the structure is closed, either by sentence-final routines or by an incompatible continuation.

5.3. Syntactic Complementation

The syntactic complementation *structure* constructs binary branching, head-complement configurations, as well as head-specifier configurations where applicable. Again subject to lexical specification to the contrary, this *structure* normally has an obligatory expectation list that presents a disjunction of possible obligatory continuations. When a syntactic element is encountered in the input sequence, it is projected by the X-bar mechanism into this *structure*. The expectation list is instantiated by the lexical selectional properties of the element, together with any grammatical principles involving the role that this particular element plays in the grammar. So, for example, the projection of the auxiliary, an instance of the functional category **I**, results in the lexically specified selection of a complement V'. This situation is shown in (20).

(20)

Syntactic Complementation Structure
optional expectations: nil
obligatory expectations: V′
content: ...(I′ (I ka) +)

Again, the '+' in the *structure* content indicates the location of the obligatorily expected continuation. The headedness of this constituent is derivative of a global left-headedness parameter that is set for all syntactic projections in Warlpiri — it is not specific to this particular *structure*.

5.4. Morpho-Syntactic Modification

The morpho-syntactic modification *structure* is analogous to the syntactic modification *structure*, only it is used for constructing morpho-syntactic representations. Again, the resulting construction is a fan-like configuration, where the sister constituents are normally not subcategorized, but are instead optional continuations. This handles any modification that occurs within a morpho-syntactic constituent. So, for example, noun-noun modification is constructed within the morpho-sytactic *structure*, since NP must be affixed by a case marking suffix to satisfy the Case Filter (which is subsumed in the morpho-syntax in Warlpiri).

5.5. Morpho-Syntactic Complementation

The morpho-syntactic complementation *structure* is the analog of the syntactic complementation *structure*. This is where selection properties regarding subcategorized complements for morpho-syntactic constituents are resolved. When a morpho-syntactic element is encountered in the parse, it normally projects in this *structure*, setting any subcategorization and obligatory selection from the lexical entry as expectations on this *structure*.

The morpho-syntax is universally right-headed. Hence, when elements project in this *structure*, they do so consistent with this general constraint on the morpho-syntax. When a right-headed constituent projects, the subcategorized complement will occur to its left; that is, it will already have been processed in the left to right pass of the input. In such right-headed constructions, the obligatory expectations that the lexical subcategorization set are not continuations of the phrase, but are instead indications of how the current contents of the morpho-syntactic complementation *structure* will unify with the rest of the parse. So, for example, when a postposition is encountered in the input, it will immediately project to **P′**, and set a backwards obligatory expectation for an NP or a CP constituent. This is subject to immediate verification by checking the adjacent constituent — the structures are unified, and the **P′**, with its expectations satisfied, becomes a closed morpho-syntactic constituent.

---

[7] It must not be a Complete Functional Complex.

Even with the strictly right-headed nature of the morpho-syntax, certain forward expectations are handled by the morpho-syntactic complementation *structure*. These are forward expectations required by the grammar principles to ensure well-formedness. So, for example, whenever an NP is processed, the grammar immediately sets an expectation for a case-assigner. In general, a global expectation is set requiring that the NP receive case from any of the possible ways that case can be assigned to NPs in the language. In Warlpiri, case is assigned in the morpho-syntax only, and it is assigned by postpositional affixes. Whenever an NP is encountered, then, case theory sets an obligatory expectation for a case-assigning postposition. This expectation is set in the morpho-syntactic complementation *structure*, essentially projecting the postposition in the *structure* together with any backwards obligatory expectations, but adding an additional expectation for the postposition itself. Processing continues, left to right, until the NP is closed by some inadmissible continuation. Under normal circumstances the closing element will be the postposition itself, satisfying the expectation of the morpho-syntactic complementation *structure* by providing it with, not only its head, but also closing the NP and making it accessible to unify with the sister position in P'. If an inadmissible continuation of NP is encountered, but it isn't a postposition, then the NP is closed and unifies with the expectation on the position within P'. There is a phonologically non-overt postposition, namely the absolutive — this is marked in the lexical entry for the category P. This essentially means that, although the expectation of the head P in the morpho-syntactic complementation *structure* is obligatory, it can be satisfied string-vacuously (as a last resort). Because of the existence of the unmarked absolutive case, the NP can satisfy the case filter through the inference of this null postposition.

5.6. Morpho-syntactic Closed Constituent Structure

The morpho-syntactic closed constituent *structure* is defined to be a single constituent with no outstanding expectations. This *structure* allows one closed morpho-syntactic constituent to be held, pending the closure of another constituent. This constituent is accessible to satisfy any backwards expectations set in subsequent parsing, but must be unified with the rest of the parse at the time that another constituent becomes closed, since only a single constituent can be held in this *structure*. The main function of this constituent is to allow for the leading edge projection of V'. Unless the verb occurs initial in the V', then it is necessary to hold off projecting a V' until a second constituent is parsed. Once a second constituent is closed, however, then the leading edge requirement for projecting a V' can be verified.

6. Priorities and Scheduling

One of the main design features of this parser is the *structures* — each linguistically defined — and generalized rules that can take on a slightly different implementation depending on what *structure* interprets them.

The *structures* divide into two basic classes — those that are primarily relevant to the syntax, and those that are primarily relevant to the morpho-syntax. The most basic part of the scheduler is that expectations set by the morpho-syntax take priority over expectations that are set by the syntax — that is, there is a preference to preserving the boundaries of the phonological words in the syntactic constituency, wherever possible. This scheduling preference can be overridden in the event of processing a lexical item that is marked as both *:syn* and *:morph*. In such cases, as long as any lexically specific syntactic and morpho-syntactic requirements are satisfied, such an element can override the priority of morpho-syntax before syntax. This creates a mismatch between the phonological word edges and the constituency of the syntactic representation — the desired result for processing clitics.
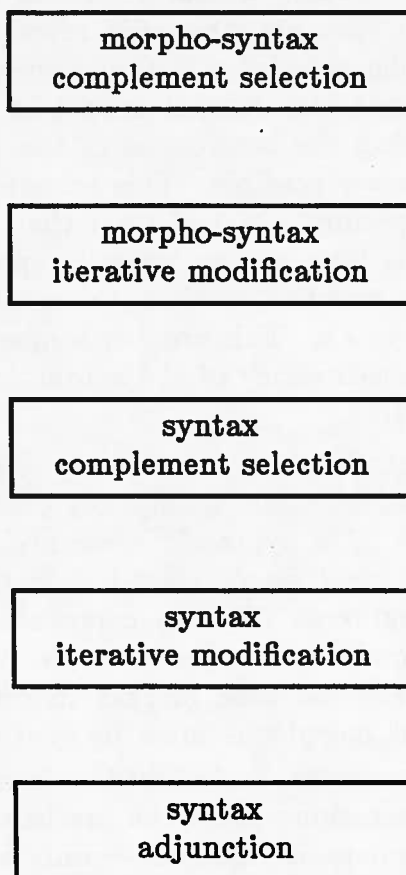
If there are unsatisfied expectations within a morpho-syntactic *structure* and an element is encountered that does not satisfy any of these expectations, but it does satisfy an expectation of a syntactic *structure* — any obligatory expectations in the morpho-syntax must be satisfied before the element can be incorporated into the syntactic *structures*. This is accomplished by the inference of empty morphemes, wherever possible. So, for example, if there is an outstanding expectation for a postpositional case marker in one of the morpho-syntactic *structures* and the current morpheme must be syntactic, then the null postpostional case marker (the absolutive) is inferred[8]. In other words, if an obligatory morpho-syntactic expectation must be satisfied, whenever the category allows it, the required element is inferred — only as a phonologically null variant. This phonologically non-overt element projects into the morpho-syntax in the same way that overt element of the same category do, setting expectations and projecting structure. The parse then continues inferring non-overt elements until all obligatory expectations are satisfied, including any that have been introduced by the phonologically null elements themselves.

The use of the intonational phrase boundary symbol (*) functions in a similar manner. Outstanding morpho-syntactic expectations cannot be maintained across an intonational phrase break. When an intonation phrase boundary symbol is encountered in the input sequence, all morpho-syntactic expectations must be resolved. This is accomplished by listing the intonational phrase boundary marker in the lexicon as having the sole property *:syn*.

Within the morpho-syntactic *structures* and syntactic *structures*, there are further priorities that facilitate scheduling the parser. In each component, *structures* that contain specific lexical selection for a particular content are prioritized before the iterative expectations typical of recursive modification *structures*, and these iterative expectations are prioritized before adjunction *structures* (syntax only). Thus we find the priorities in (21).

---

[8] Those categories that have a phonologically null surface variant are marked as such in the lexicon of grammatical categories.

(21)

```
┌─────────────────────────────┐
│      morpho-syntax          │
│   complement selection      │
└─────────────────────────────┘

┌─────────────────────────────┐
│      morpho-syntax          │
│   iterative modification    │
└─────────────────────────────┘

┌─────────────────────────────┐
│         syntax              │
│    complement selection     │
└─────────────────────────────┘

┌─────────────────────────────┐
│         syntax              │
│    iterative modification   │
└─────────────────────────────┘

┌─────────────────────────────┐
│         syntax              │
│        adjunction           │
└─────────────────────────────┘
```

Activation of the adjunction structure in the syntax is highly restricted and must be accompanied by an adjunction intonational marker (!), unless the adjunction is string vacuous (*ie* trace adjunction).

7. Trace Handling

A number of structurally defined positions can initialize a trace-handling procedure. In Warlpiri, the only such position is the specifier of **CP**. Trace initialization involves parsing the constituent *in situ* and returning a coindexed trace of the same phrasal type or lexical category. This trace sets up an overriding priority in the scheduler with respect to priorities of the syntactic *structures*. If a syntactic expectation for an element with the same category as the trace arises — the trace immediately satisfies this expectation and is incorporated into the *structure* as if it were the full phrasal or lexical antecedent of the trace. This amounts to an early gap location strategy.

As subsequent syntactic constituents are constructed, each new maximal projection of embedding is pushed onto a stack. The stack definition includes a test that determines whether a trace within that constituent would violate the 'no barriers' condition between an antecedent and its trace. If a violation would

occur, then the projection cannot be pushed onto the stack. If the projection is not a Complete Functional Complex[9] then it is passed through the adjunction structure, where an adjoined trace is inserted — XX type traces adjoining to the maximal projection and X type traces adjoining to the head of the projection. This then instantiates a new stack, again defined according to the 'no barriers' criterion. As projections are closed, they are popped off the stack. Using this approach, we conceivably produce spurious adjunction traces in constituents that do not contain the ultimate trace of the D-structure position.

The full power of this mechanism cannot be tested on Warlpiri data. It certainly works for Warlpiri, but it also has sufficient power to work on cases that do not obtain in Warlpiri. Hand-testing of this method on simple cases in English has proven promising, but rigorous testing on a wide range of cases is required, as is thorough assessment of the consequences that the spurious traces have in linguistic theory. No doubt, as these considerations are made, considerable refinement to this mechanism will prove necessary.

8. Thematic Projection

As the syntactic and morpho-syntactic representation is constructed, the parser also constructs an instantiated θ-grid for each C'' in the parse. The θ-grid is instantiated by means of an intermediate case grid. The case grid is merely a record of exactly which morphological case markings are encountered in the left to right parse. The case grid is linked to the θ-grid by means of default linkings of morphological cases to thematic relations (as in Nash, 1980). Since the linking of the morphological cases to the thematic relations[10] is largely dependent on the lexical entry of the verb, the case grid collects the necessary information until the verb is encountered — functioning as a buffer in the construction of the thematic representation. The case grid is ordered according to the *case hierarchy* (Nash, 1980):

$$erg - abs - dat .$$

The auxiliary instantiates a minimal θ-grid consistent with the person and number markings that it contains. This θ-grid is underspecified as to exact thematic content. The entries are merely associated with numbers that indicate relative positions in the thematic hierarchy. This grid picks up full specification when composed with the θ-grid of the verb when the verb is incorporated into I'.

After the auxiliary is parsed, the parser continues, left to right, indicating in the case grid any case-marked nominal phrases (*ie* P's) as they are encountered. Once the verb is encountered, its θ-grid is retrieved from its lexical entry, and the case grid is linked to it[11]. Subsequent case marked nominal phrases can then link directly to the θ-grid. Certain case markers carry the

---

[9] A language specific list of Complete Functional Complexes is provided to the parser.

[10] I assume a hierarchy of thematic relations as presented in Ostler, 1979 and applied to Warlpiri in Nash, 1980.

[11] I do not deal with 'doubly classified' verbs (Nash, 1980, p.195), or similar cases of lexical ambiguity.

information about their thematic role independent of the predicate's θ-role[12] (*eg.* locatives). **P'**s marked with such case markers can link directly to the θ-grid.

Once **V'** is closed, it is normally unified with an expectation set for a complement of **I'**. When this unification occurs, the θ-grid associated with the **V'** is composed with the θ-grid associated with the auxiliary. If all subcategorized arguments in the **V'** θ-grid are realized, then this amounts to feature checking. If any subcategorized arguments are missing in the **V'** θ-grid, then these are instantiated with the corresponding person and number values from the auxiliary grid. Default values (3rd person, singular) are then filled in on all remaining unrealized subcategorized arguments in the θ-grid, now projected to **I'**.

Discontinuity is handled by instantiating the θ-grid slot with the intersection of the discontinuous pieces. Intersection would need to be defined rigorously by an interpretive semantic component that is not implemented in this parser.

In the case of a **P'** topic or pre-auxiliary constituent, the postpositional case marker instantiates this constituent into the case grid in the left to right parse of the sentence. If the topic is repeated in the clause, as it quite frequently is, then the topic provides no additional information to the percolated θ-grid (since $X \cap X = X$). If the topic is not repeated, then it serves to fill the θ-role associated with its morphological case.

The process that constructs θ-grids in embedded clauses is similar to the one just discussed for matrix clauses. In the case of infinitival clauses, there is no auxiliary, hence the θ-grid of **V'** is the θ-grid of the clause. Discontinuous pieces of infinitival clauses link into a single θ-grid by means of agreement of their complementizers. Adjoined relative clauses return their own instantiated θ-grid in the same manner as main clauses. Both of these subordinate clause constructions are dependent on the matrix clause. They are not combined with the θ-grid of the main clause except insofar as the dependent clause complementizer dictates a switch-reference type coindexation between θ-grid positions in the subordinate and main clauses. The complementizers act as functions, operating on θ-grids (mediated by case grids) dictating coindexations that reflect subordination relations between the two clauses.

Although the θ-grid is not instantiated until the predicate is encountered, there are frequently earlier hints as to the type of θ-grid that will be required. The auxiliary is the main locus of such information. The schematic representation of a θ-grid on the auxiliary is used in the parse to set expectations for subsequent thematic processing. Similarly, complementizers on embedded clauses set expectations for the type of controller that they select in the main clause. This suggests that ultimately the approach taken to the syntactic parse could extend to thematic parsing — syntactic information could be used to set thematic expectations in a left to right parse. Not all syntactic information is

---

[12] These are the 'semantic cases' (Nash, 1980) and are most like English prepositions.

relevant to thematic parsing — similarly we have found that only a small amount of prosodic information was actually relevant to the syntactic parse. With the syntactic parse, it was found that general principles can supplant the need for *ad hoc* stipulation. However, there are few such general principles governing thematic well-formedness. In the current implementation, this is a highly stipulative component with very few general principles governing it.

9. Extending the Parser

One of the main areas that the parser is deficient in is the range of phenomena from Warlpiri that it deals with. There are many subtleties to this language that I have not addressed at all — many of these I simply do not understand, and I am certain that there are a great many phenomena in Warlpiri that I am not aware of. I have attempted to deal with the basic types of clause structures that come up over and over again in the literature on Warlpiri, as well as the properties that are cited as fundamental to Warlpiri. Clearly a major outstanding task is to extend the parser to deal with a larger fragment of Warlpiri syntax and morpho-syntax.

Although some of the routines in the parser are specific to Warlpiri, there are a number of general strategies that could be applied to a parser for other languages, including English. The balancing of lexical and grammatical selection with parsing expectations extends in a general fashion to parsing other natural languages. In addition, the two-fold nature of projection — by general principles of X-bar theory together with leading edge projection — can be extended to supplant specific phrase structure rules in a general parser. The division of the parser into distinct *structures* guarantees the well-formedness of a construction in a particular configuration. Trace-handling routines, although somewhat impoverished as far as Warlpiri is concerned, suggest a generally extendible approach to gap location. The 'Elsewhere Condition'-inspired priorities and scheduling suggest that even such implementation details as the scheduling of actions in the parse can take linguistic principles into consideration.

In particular, the balance of burden placed on syntactic and morpho-syntactic structure building is used extensively in Warlpiri where there is rich morpho-syntactic structure. Many languages use bound morphemes for syntactic functions, and, hence, the construction of the S-structure representation on the morphemes rather than on phonological words could be exploited to deal with a wide range of such phenomena interlinguistically. Such an approach also provides a way of dealing with cliticization and bracketing paradoxes (Sproat, 1985).

The crucial use of prosodic information in the parse is generally extendible and is particularly important if parsers are ever to extend beyond being merely reading machines. In addition, the simultaneous construction of the S-structure representation with the thematic representation (instantiated θ-grids) suggests an interleaved deterministic model to simultaneously construct different levels of

linguistic representation.

These are speculations of the generality of this parser, and its extendibility to other languages. The remaining task is to actually test these by converting the implementation to deal with other languages. To facilitate such an endeavour, I have kept what I felt were Warlpiri specific routines separate from what seem to be more generally applicable routines. The parser was designed to test conventional linguistic theory on an apparently unconventional language. To the extent that it is possible to specify the principles of linguistic theory as parsing routines independent of any particular language, I have endeavoured to do so. Although a far cry from a model that merely lists a few basic parameter values that would readily allow toggling from one language to another, the approach taken here is superior to many 'kitchen sink' parsers that are designed exclusively for one particular language, losing any general principles in among language specific routines.

The choice of Warlpiri as a test case for a parser was originally motivated by the properties of adverbial modification in English — properties such as free order and discontinuity. These properties extend to all argument types in Warlpiri. Linguistic theory, in general, under-addresses adverbials, focusing on subcategorized arguments instead. To address Warlpiri at all in the current linguistic framework involves addressing adverbial-type properties, since they apply to subcategorized arguments in Warlpiri as well. Some of the routines developed to handle Warlpiri are immediately extendible to deal with nonsubcategorized prepositional phrases in English.

10. A Sample Parse

The parser is initialized to expect a C'' — that is, it assumes that the input is a complete clause. This is accomplished by setting the syntactic complementation *structure* to C'', which, in turn, being a maximal projection, sets an obligatory expectation for a head (*ie.* a C'). This will be provided by a procedure of *inverse projection*. The XP in the optional expectations indicates a possible landing site of movement of a maximal projection. The head need not be linked to the second of these two expectation sites, since general principles of X-bar well-formedness in the syntax will derive this restriction.

> *Syntactic Complementation Structure:*
> optional expectations: XP
> obligatory expectations: head
> content: $(C'' \ + \ +)$

Consider a simple clause in Warlpiri (example 23, chapter 1):

kurdu-ngku-ka   wajilipi-nyi   maliki
        child-erg-AUX:present chase-nonpast   dog(abs)
        'The child is chasing the dog.'

With the above initialization, the parser commences on a left to right parse of
the sentence:

   Current Morpheme: kurdu

This morpheme projects in the morpho-syntactic modification *structure.*

        Morpho-syntax Modification Structure:
            optional expectations: N
            obligatory expectations: nil
            content: (N' (N kurdu) +)

As soon as an N' is projected, the procedure that implements the Case Filter is
activated, setting an expectation for a case-assigner for this N'. In Warlpiri, this
amounts to setting an expectation for a postpositional case-marker. Since the
category **P** is marked in the category lexicon as projecting in a binary
configuration, the expectation for a postposition is passed to the morpho-
syntactic complement *structure,* where it projects, setting not only the expecta-
tion for its complement, but also an expectation for a head for this constituent
(a **P**, whether overt or non-overt).

        Morpho-syntax Complementation Structure:
            optional expectations: nil
            obligatory expectations: N' C''
            obligatory expectations: head null-variant
            content: (P' + (P +))

The parser proceeds to the next morpheme:

   Current Morpheme: ngku

This is incorporated into the morpho-syntactic complementation *structure,* and
all preceding expectations are resolved. The morpho-syntactic modification
structure is closed, and the resulting **N'** is used to satisfy the obligatory expecta-
tion in the morpho-syntactic complementation *structure.* The resulting closed
constituent is transferred to the morpho-syntactic closed constituent *structure,*
and processing continues to the next morpheme. The case grid for the sentence
is set to *[erg].*

   Current Morpheme: ka

ka is an auxiliary base, and combines in a n-ary fashion with various morphemes marking person and number agreement. It initially projects into the morpho-syntactic modification *structure* with expectations set for subsequent auxiliary morphemes. All auxiliary morphemes are incorporated into an n-ary branching structure, dominated by the *I* node. The auxiliary morphemes are used to set a default θ-grid for the auxiliary, in this case the minimal grid $\{<1>3rd,sg\}$. The occurrence of an impossible continuation of the auxiliary forces its closure. In this example this occurs when the next morpheme is encountered.

Current Morpheme: wajilipi

Since the auxiliary is now a closed constituent, it should be passed to the morpho-syntactic closed constituent *structure*, which at the moment contains the **P′**. In order for processing to continue, the **P′** must be removed from this *structure*. The only possibility is for it to be incorporated into the syntactic complementation *structure*, satisfying the optional expectation for a maximal projection. This is done, initializing the trace-handling routine with *(P′ t)*. Having done this also satisfies the second position requirement on the auxiliary.

The auxiliary can now move to the morpho-syntactic closed constituent *structure*. It is now a syntactic atom of type **I**, and should project in the syntactic structures. It projects in a binary fashion, and is a candidate for projection in the syntactic complementation *structure*. This triggers the inverse projection of the **C″**, expanding the content of the syntactic complementation *structure* until the auxiliary can be incorporated into this *structure*.

Syntactic Complementation Structure:
    obligatory expectations: head
    obligatory expectations: V′
    content: (C″ (P′ (N′ (N kurdu))(P ngku))
              (C′ (C ∅) (I′ (I +) +))

ka satisfies the expectation for a head for I′, and is unified with the head position in the content of this *structure*.

The verb is a morphological atom that projects to zero bar-levels[13] in a binary fashion. It is incorporated into the morpho-syntactic complementation *structure*, setting an optional expectation for a tense affix. The θ-grid is retrieved from the lexicon for this verb — {*agent theme*} — as is its case grid — *[erg abs]*.

---

[13] This is the way the parser deals with certain types of morphological structure that seems to violate the strict X-bar type morpho-syntactic projection. Any category that projects to a single bar-level, is one that may have suffixes but the category of the resulting constituent is the same type as the earlier category. In a sense, this implements the notion of *relative head* — the subsequent affixes are unmarked for category.

Morpho-syntax Complementation Structure:
    optional expectations: tense
    content: (V (V wajilipi) +)

Current Morpheme: nyi

The current morpheme is the nonpast tense morpheme — it is unified into the morpho-syntactic complementation *structure*, closing this constituent, which is then transferred to the morpho-syntactic closed constituent *structure*.

Current Morpheme: maliki

This noun projects into the morpho-syntactic modification *structure* triggering the expectation of a case-assigner. This results in the projection of a **P′** in the morpho-syntactic complementation *structure* as before:

Morpho-syntax Complementation Structure:
    optional expectations: nil
    obligatory expectations: N′ C″
    obligatory expectations: head null-variant
    content: (P′ + (P +))

In this case, the end of the input is encountered — an impossible continuation for N′, which closes and unifies with the expectation of the morpho-syntactic complementation *structure*. The non-overt variant of **P** satisfies the obligatory expectation set in the morpho-syntactic complementation *structure*, and this constituent closes, adding *abs* to the accumulating case grid. At this point, the closed morpho-syntactic constituent *structure* must be cleared to make way for this new closed constituent. The priority on trace-handling results in the trace of the **P′**, together with the verb[14] projecting (by means of the leading edge) a **V′** in the syntactic modification *structure*. The θ-grid is instantiated with the **P′**/trace chain, which links to the *agent* role by default linking rules.

Syntax Modification Structure:
    optional expectations: PP CP
    obligatory expectations: nil
    content: (V′ (P′ t) (V (V wajilipi) (TENSE nyi)) +)

---

[14] The verb would have sufficed to trigger a syntactic V′ on its own.

The contents of the morpho-syntactic complementation *structure* can now move to the closed morpho-syntactic constituent structure, and then be incorporated into the syntactic modification *structure*. This links to the *theme* role by default linking of the absolutive case. The end of input is an inadmissible continuation for the syntactic modification *structure*, which closes and unifies with the expectation for **V′** set in the syntactic complementation *structure*. At this point the default θ-grid on the auxiliary composes with the grid on the **V′**, verifying feature agreement in this example. This closes the sentence in a successful parse:

> *Syntactic Complementation Structure:*
>   obligatory expectations: head
>   obligatory expectations: V′
>   content: (C″ (P′ (N′ (N kurdu))(P ngku))
>   (C′ (C ∅)
>   (I′ (I ka)
>   (V′ (P′ t)
>   (V (V wajilipi)
>   (TENSE nyi))
>   (P′ (N′ (N maliki))
>   (P ∅)))))

CHAPTER 5

Conclusions

This thesis has investigated the status of Warlpiri, a putative non-configurational language, with respect to linguistic theory, and with respect to an interpretation of linguistic theory into a parsing model.

The proposal here is that Warlpiri is not exceptional as far as linguistic theory is concerned — its apparent exceptionality derives primarily from a greater than average burden being placed on morpho-syntactic structure for the satisfaction of linguistic well-formedness conditions. In the parsing model, this necessitates the construction of the s-structure representation on the morphemes of the sentence, rather than on the words of the sentence. This finer grained structure highlights the similarities between Warlpiri and prototypical configurational languages — indicating that a more unified treatment of this non-configurational language within the current framework is possible.

I depart from Hale's claim that Warlpiri cannot have syntactic movement by arguing that pre-auxiliary position is best analyzed as syntactic movement to the specifier of **CP**. Certain principles of grammar (the Case Filter, the Extended Projection Principle) can be preserved virtually intact for Warlpiri if they are restricted to apply only within the morpho-syntactic representation. This is proposed as a parametrization on these principles, allowing them to define well-formedness within the limited domain of the morpho-syntax. In this way, the syntax, together with the morpho-syntax, conspire to ensure that the principles of grammar are satisfied, without placing the entire burden on the word level phrasal syntax.

Warlpiri grammatical relations are morphologically defined, and not positionally or structurally defined with respect to **I** or **V** in the syntax. Grammatical relations serve to identify the thematic role of arguments with respect to their predicate. I argue that, rather than a global parameter of 'nonconfigurationality', the choice of lexical identification of thematic relations is a local property of each particular thematic relation. The exceptional nature of Warlpiri with respect to this stems merely from the use of lexical identification to the exclusion of all other means of thematic identification of arguments. Similarities between the properties of arguments in Warlpiri and properties of prepositional phrase modifiers in English stem from the fact that both of these constructions carry their own case assigner and thematic identifier (the case marking postpositional affix in Warlpiri, and the lexical preposition in English).

In this way, we can account for the similarities without giving up the argument status of the subcategorized nominal phrases in Warlpiri. That these nominal phrases act like optional modifiers is derived from other principles, and in no way indicates that they are optional modifiers[1].

The Warlpiri parser was designed to follow the theory as transparently as possible. Effort was made to keep those aspects of the implementation that are processing translations of theory separate from the aspects of the implementation that were necessary merely to facilitate the parse itself. In a similar fashion, effort was made to keep separate those aspects of the implementation that are general implementations of linguistic principles from those that are specific to Warlpiri. The intent here is to make the parser as modifiable as possible — both to handle changes and additions in the theory, as well as to extend to other languages.

The parser constructs s-structure representations on the morphemes of the input. The structure immediately dominating the morphemes is constructed by the implementation of morpho-syntactic well-formedness conditions; while the structure above this is constructed by the implementation of syntactic well-formedness conditions. A non-isomorphism between the s-structure bracketing and phonological bracketing (as evidenced by the mismatch between constituents in the s-structure and phonological word boundaries) allows for the parsing of clitics. However, phonological and prosodic constituency do provide limited information that guides the s-structure parse. In particular, I argue for the crucial use of certain prosodic boundaries to preserve determinism in the parse.

In addition to the s-structure parse, an instantiated θ-grid is returned for each clause in the s-structure. This is constructed using linking rules (Ostler, 1976; Nash, 1980) and conventions of thematic percolation.

The parser does not, by any means, account for the full range of phenomena found in the Warlpiri language. The focus here has been on basic properties of Warlpiri clauses, as cited in the literature, and what these properties dictate about the nature of universal grammar and a parser that relies on it.

---

[1] This was the position argued for by Jelinek (1984), who claimed all of the nominal phrase constituents in Warlpiri are merely optional 'adjuncts'.

References

Abney, S. and Cole, J. 'A Government-Binding Parser'. In *Proceedings of the Fifteenth Annual Meeting of the North Eastern Linguistic Society*, 1985.

Anderson, S. and P. Kiparsky (eds.). *A Festschrift for Morris Halle*, Holt, Rinehart and Winston, 1973.

Bach, E. and R. Harms. (eds.) *Universals of Linguistic Theory*, Holt, Rinehart and Winston, 1968.

Berwick, R. and Weinberg, A. *The Grammatical Basis of Linguistic Performance: Language Use and Acquisition*, MIT Press, 1984.

Bouchard, D. *On the Content of Empty Categories*, Foris, 1984.

Bresnan, J. (ed.) *The Mental Representation of Grammatical Relations*, MIT Press, 1982.

Brunson, B. *Thematic Argument Structure*, ms. University of Toronto, 1986.

Chomsky, N. 'Remarks on Nominalization'. In D. Napoli & E. Rando (eds.), 1979.

------- *Lectures on Government and Binding: The Pisa Lectures*, Foris, 1981.

------- *Some Concepts and Consequences of the Theory of Government and Binding*, MIT Press, 1982.

------- *Barriers*, MIT Press, 1986.

Cowper, E. *An Introduction to Syntactic Theory: The Government-Binding Approach*, ms. University of Toronto, 1984.

Di Sciullo, A. and E. Williams. *Three Types of Words: The Relation of Morphology to Syntax*, ms. 1985.

Dixon, R (ed.). *Grammatical Categories in Australian Languages,* Humanities Press, 1976.

Dowty, D., Karttunen, L., and Zwicky, A. (eds.) *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives,* Cambridge University Press, 1985.

Emonds, J. *A Transformational Approach to English Syntax,* Academic Press, 1976.

Farmer, A. *Modularity in Syntax: A Study of Japanese and English,* MIT Press, 1984.

Fillmore, C. 'The Case for Case'. In Bach and Harms, 1968.

Ford, M., Bresnan, J., & Kaplan, R.M. 'A Competence-Based Theory of Syntactic Closure'. In Bresnan, J. (ed.), 1982.

Frazier, L. *On Comprehending Sentences: Syntactic Parsing Strategies,* Indiana Linguistics Club, 1979.

Frazier, L., & Fodor, J.D. 'The Sausage Machine: A New Two-stage Parsing Model'. In Cognition 6, pp. 291-325, 1978.

Gruber, J. *Studies in Lexical Relations.* PhD dissertation, MIT, 1965.

——. *Lexical Structures in Syntax and Semantics.* North-Holland Linguistic Series, 1976.

Hale, K. *Preliminary Remarks on Walbiri Grammar,* ms. MIT, 1967.

——. 'Person Marking in Walbiri'. In Anderson and Kiparsky (eds.), 1973.

——. *Warlpiri-English Vocabulary,* ms. MIT, 1974.

——. 'The Adjoined Relative Clause in Australia'. In Dixon (ed.), 1976.

——. *On the Position of Walbiri in a Typology of the Base.* ms. MIT, 1981.

————.'Some Essential Features of Warlpiri Verbal Clauses'. In Swartz (ed.), 1982.

————. 'Warlpiri and the Grammar of Non-configurational Languages', *NLLT 1*, 1983.

————. *Preliminary Remarks on Configurationality*. ms.

Hirst, G. *Semantic Interpretation Against Ambiguity*. Technical Report CS-83-25, Brown University, 1983.

Jackendoff, R. *Semantic Interpretation in Generative Grammar*, MIT Press, 1972.

————. *X-Bar Syntax: A Study of Phrase Structure*, MIT Press, 1977.

Jelinek, E. 'Empty Categories, Case and Configurationality.' *NLLT 2*,1984.

Kaplan, R.M., & Bresnan, J. 'Lexical Functional Grammar: A Formal System for Grammatical Representation'. In Bresnan, J. (ed.), 1982.

Kashket, M. 'Parsing a Free-Word Order Language: Warlpiri'. In Proceedings of the *24th Annual Meeting of the Association for Computational Linguistics*, 1986.

King, M. (ed.) *Parsing Natural Language*. Academic Press, 1983.

Kiparsky, P. 'Elsewhere in Phonology'. In Anderson and Kiparsky, (eds.), 1973.

————. 'Lexical Phonology and Morphology'. In *Linguistics in the Morning Calm*, Linguistic Society of Korea. Seoul, Hansin, 1982.

Lasnik, H., & Saito, M. 'On the Nature of Proper Government', *Linguistic Inquiry*, 8.3, 1977.

Marantz, A. *On the Nature of Grammatical Relations*, MIT Press, 1984.

Marcus, M. *A Theory of Syntactic Recognition for Natural Language*. MIT Press, 1980.

Mohanan, K.P. *Lexical Phonology*, Ph.D. dissertation, MIT, published by IULC, 1982.

Napoli, D., & Rando, E. *Syntactic Argumentation*, Georgetown University Press, 1979.

Nash, D.G. *Topics in Warlpiri Grammar*, Ph.D. dissertation, MIT, 1980.

Ostler, N. *Case Linking: A Theory of Case and Verb Diathesis, Applied to Classical Sanskrit*, Ph.D. dissertation, MIT, 1979.

Pranka, P. *Syntax and Word Formation*, PhD. dissertation, MIT, 1983.

Radford, A. *Transformational Syntax*, Cambridge University Press, 1981.

Reinhart, T. 'Definite NP Anaphora and C-Command Domains', *LI* 12:4, 1981.

van Riemsdijk, H. and E. Williams. *Introduction to the Theory of Grammar*, MIT Press, 1986.

Safir, K. *Syntactic Chains*. Cambridge University Press, 1985.

Sells, P. *Lectures on Contemporary Syntactic Theories*, CSLI, 1985

Simpson, J. *Aspects of Warlpiri Morphology and Syntax*, Ph.D. dissertation, MIT, 1983.

Simpson, J. and J. Bresnan. 'Control and Obviation in Warlpiri', *NLLT 1*, 1983.

Sparck Jones, K. and Wilks, Y. (eds.) *Automatic Natural Language Parsing*. Ellis Horwood Ltd. 1983.

Sproat, R. *On Deriving the Lexicon*. Ph.D. dissertation, MIT, 1985.

Stowell, T. *Origins of Phrase Structure*, unpublished Ph.D. dissertation, MIT, 1981.

Swartz, S (ed.). *Papers in Warlpiri Grammar*, Working Papers of SIL-AAB, Series A, Vol. 6, 1982.

Wehrli, E. *A Government-Binding Parser for French*, Working Paper 48, Institut pour les Etudes Semantiques et Cognitive, Universite de Geneve, 1984.

Postscript to Technical Report Version

This postscript summarizes some of the directions that this work has taken since 1986.

Immediately after the completion of the original version of this report, I spent a term at AT&T Bell Labs in Murray Hill, New Jersey (Sept, 1986 — Dec, 1986). During this time I did preliminary intonational studies of three tapes of native Warlpiri speakers[1]. From the pitch tracks of this data, it is apparent that the claims about intonational phrases in Warlpiri made in this thesis are essentially correct. It is true that the sequence N N P has a different intonational contour when intended as a complex nominal as opposed to when it is intended as two separate nominals. If the sequence is to have the complex nominal interpretation, there is a pitch peak on the first syllable of the first noun, with falling pitch across the remaining phrase. If the sequence is to be interpreted as two separate nominal phrases, there are two pitch peaks — one on the first syllable of the first noun, and one on the first syllable of the second noun. Of course, only a few such cases were obtained from the tapes so no claim can be made about the absolute reliability of these intonational cues.

On single word utterances, the pitch peak corresponds to primary word stress. In multiple word intonational phrases with a single peak on the stressed initial syllable of the first word, there is no evidence in the pitch track of any other primary stress on subsequent words in the same phrase. Similarly, there is no evidence for secondary stress. One possibility is that what I am calling the intonational phrase is really just the domain of the prosodic word in Warlpiri. If this is the case, then N N P sequences will be said as a single prosodic word if a single constituent, and as two prosodic words if two constituents. This would serve equally well in the parser and eliminate the need for the intonational phrase marker. The input to the parser would consist of N-N-P if a single constituent, and N N-P if two constituents.

In addition, while at AT&T Bell Labs, the Warlpiri parser was interfaced with a morpho-phonological parser for Warlpiri. The input to this parser is the sequence of phonemes, together with stress information. No word boundaries or morpheme boundaries are provided to the morpho-phonological parser. The morpho-phonological representation of the sentence is constructed, and from this representation is extracted a *flattened* representation that serves as input to the syntactic and morpho-syntactic parser (for details, see Sproat and Brunson, in the 1987 ACL proceedings).

---

[1] I wish to thank Mary Laughren for providing me with these tapes, as well as Chilin Shih for teaching me how to use the pitch tracking programs at the Labs.