Parsing and Constraints on Word Order

Lyn Frazier

PARSING AND CONSTRAINTS ON WORD ORDER

Lyn Frazier

1. Introduction

The goal of this paper is to begin to explore the constraints on the expected word order of natural languages which derive from the structure and operation of the language parsing device. The particular type of word order constraints that we might expect to find will of course depend on our view of the sentence parsing process; however, that there should exist restrictions on word order which are motivated by the exigencies of sentence parsing is strongly suggested simply by the course which psycholinguistic theories have taken in recent years.

One tendency in psycholinguistics has been to assume an ever shorter lag between the time when the parser receives some item in the incoming lexical string and the time when the parser makes a decision about the structural role and interpretation of that item. For example, in the early 1960's when the dominant psycholinguistic hypothesis was that there was a direct one-to-one correspondence between the rules of grammar and the operations hearers use to interpret a sentence, the implicit assumption must have been that the entire sentence was available for analysis at the time when the parser makes many of its decisions about the structure and interpretation of a sentence, since the ultimate domain of application of transformational rules was the entire sentence. On this view of sentence processing, there is no obvious reason to expect any one particular ordering of the elements in a sentence to be preferred by the parser to any (consistent) alternative ordering since all elements are assumed to be simultaneously available when the parser makes at least its higher order decisions about the analysis of the sentence.

178

By contrast, in recent years it has been emphasized how very quickly the parser makes even higher order decisions about the analysis of linguistic material. On this view of sentence processing, one would expect certain orderings of elements to be preferred over others by the parsing device since only certain orderings of elements would permit the parser to make rapid and accurate on-line decisions about the analysis of an item without having to wait until a large number of subsequent items had also been received. Thus, the very fact that most current theories of sentence processing assume that the parser's decision lag is very short leads us to expect that certain word orders will be preferred over others for reasons of efficient on-line sentence processing.

In this paper I will focus on Greenberg's universals concerning the word order of natural languages and show that many of these universals may be attributed to the exigencies of sentence parsing. Greenberg's universals are stated as implications of the form: given that a language exhibits some property X, it will also (tend to) exhibit some other property Y. (Rather than review all of these universals here, I will merely assume that the reader has a basic familiarity with them (cf. Greenberg (1965)).

Greenberg considered the relative order of subjects, verbs and objects, the relative order of nouns and qualifying adjectives, and the placement of non-lexical heads of phrases (prepositions and postpositions) to be central in determining the "basic order typology" of natural languages. The data pertaining to these aspects of word order are presented in Table 1 below.

		VSO	SVO	SOV
Postpositions:	A precedes N	0	1	6
Postpositions:	N precedes A	0	2	5
Prepositions:	A precedes N	0	4	0
Prepositions:	N precedes A	6	6	0

TABLE 1. (Taken from Greenberg, 1965.)

These data illustrate some of the hazards of an investigation (like the present one) which relies on implicational universals for its primary data. One problem with this type of investigation is that it is simply not clear what data an explanation of word order should account for. For example, the strength of the correlations that can be observed in the data in Table 1 varies markedly: the correlation between basic Object-Verb (OV) word order and postpositions and between basic VO word order and preposi-

tions is quite impressive and holds for 27 of the 30 languages analyzed by Greenberg; by contrast, the relative order of nouns and qualifying adjectives does not appear to be very strongly correlated with the basic (subject-verb-object) word order, except in the case of VSO languages. Hence, pretheoretically it is not clear whether the order of nouns and qualifying adjectives is related to the order of subjects, verbs and objects, or whether the partial correlation seen in Table 1 indicates that the placement of qualifying adjectives is related to some other property of languages which is found consistently only in VSO languages. Furthermore, in the case of stronger or more pervasive correlations, such as that between OV word order and postpositions, the 'direction' of the implication (i.e. whether OV word order determines the placement of non-lexical heads or vice versa) is difficult to establish. (In fact, the explanation of word order which I will propose suggests that the statement of some of Greenberg's universals may have to be revised somewhat.) Recent advances in linguistic theory (e.g. the development of $\bar{X}$ theory) also offer a much richer framework for the analysis of natural languages than was available when Greenberg first formulated the universals discussed here, and thus a third problem is that it is often unclear how the elements involved in a particular universal would be characterized in current linguistic theory.

Though one might reasonably argue that these problems constitute a very compelling reason for not pursuing this investigation, I prefer to argue just the opposite, viz. that trying to resolve these problems is itself part of the justification for undertaking the investigation, despite its many hazards. Another reason for proceeding with the study is that it raises some of the most important issues in linguistic theory since an explanation of the typology characterized by Greenberg's universals must, at least ultimately, explain why natural languages may differ from one another along certain parameters but not others; why, given the various options available to natural languages, some are preferred over others; and, finally, why certain options are linked together so that a particular alternative may be highly valued just in case the language in question also exhibits some other specific property, but is not highly valued if the language lacks that property.

Though I will address these issues only in so far as they relate to word order, the explanation I will propose is "functional" in nature and as such should be couched in terms of an explicit theory of the ways that natural languages may be influenced by the exigencies of sentence processing. Thus, the final section of the paper is devoted to general questions concerning the relation of language parsing and language structure. We will begin, however, by looking at a previous attempt to provide a "functional" or parsing explanation of word order typologies.

2. Avoidance of center-embedding as an explanation of word-order

In a well-known paper, Kuno (1974) suggests that many of Greenberg's universals can be explained by appealing to the perceptual complexity of certain syntactic patterns, notably center-embedded constructions and constructions with "conjunction juxtaposition" (i.e. constructions in which

180

two or more clause markers occur immediately adjacent to each other). Kuno argues that languages with dominant SOV word order tend to place relative clauses prenominally because postnominal positioning of relative clauses would maximize the amount of center-embedding in these languages, as illustrated in (1) and (2) (≈ Kuno's (9) and (10)).

(1) SOV with Prenominal Relative Clauses

    a. [Mary loved] boy died.
    b. [Mary loved] boy Jane Hated.
    c. Jane [Mary loved] boy hated. (center-embedding)

(2) SOV with Postnominal Relative Clauses

    a. Boy [Mary loved] died. (center-embedding)
    b. Boy [Mary loved] Jane hated. (center-embedding)
    c. Jane boy [Mary loved] hated. (center-embedding)

Similarly, prenominal positioning of relative clauses would lead to rampant center-embedding in languages with dominant VSO word order. Thus, Kuno proposes that VSO and SOV languages place relative clauses in those positions which will minimize the occurrence of perceptually complex center-embedded constructions.

In languages with SVO word order, center-embedding is equally likely to occur regardless of the placement of relative clauses. Kuno suggests that these languages adopt special syntactic devices, primarily raising and extraposition rules, to cope with center-embedding. However, this explanation is problematic since it is not clear why raising and extraposition rules should not also be available in VSO and SOV languages. And, of course, if such rules are available this would eliminate the pressure for the placement of relative clauses to be restricted in these languages. Furthermore, as has been pointed out by Rosenbaum (ms.), according to this account there is no reason to expect the placement of relative clauses in SVO languages to be systematic; but, in fact, relative clauses are quite systematically placed postnominally in SVO languages. Hence, we might expect whatever principle accounts for the consistent placement of relative clauses in SVO languages to also account for their placement in VSO and SOV languages.

Kuno provides arguments parallel to those just given for relative clauses to explain why SOV languages tend to place "conjunctions" ("forms one of whose functions is to mark clause boundaries", i.e. complementizers and adverbial conjunctions such as if, when, etc.) in clause-final position, whereas VSO languages tend to place them in clause-initial position. Again his argument is that the alternative placement of these items would lead to more center-embedding and conjunction juxtaposition than does the placement which is actually found in natural languages. He also suggests that natural languages do not mark both the beginnings and ends of clauses because this would place too great a memory load on speakers, who would have to remember how many markers (e.g. that) should occur at the end of sentences like (3).

(3) I think [that Mary said [that she believes [that the world
     is flat that] that] that].

I will suggest later that both the placement of "conjunctions" and the
fact that languages do not mark both the beginnings and ends of clauses
actually follow as an automatic consequence of the role that such items
play in sentence parsing.

The avoidance of center-embedding is also claimed to explain why
VSO languages are prepositional and SOV languages tend to be postposi-
tional. Kuno (p. 127) suggests:

> Given that attributives usually follow their head nouns in
> VSO languages and that they usually precede their head nouns
> in SOV languages, there are four possibilities for representing
> the phrase the color of the flowers in the vase on the table.

(28) VSO Language
      a. Prepositional: color[of flowers[in vase[on table]]]
      b. Postpositional: color[flowers[vase[table on]in]of]
(29) SOV Language
      a. Prepositional: [of[in[on table]vase]flowers]color
      b. Postpositional: [[[table on]vase in]flowers of]color

> From this, it should be clear that combinations of postpositions
> and postnominal positioning of attributives would produce a
> hopeless situation of center embedding and juxtaposition of
> postpositions, as seen in (28b), and that combinations of preposi-
> tions and prenominal positioning of attributives would produce an
> equally hopeless situation of center-embedding and juxtaposition
> of prepositions, as seen in (29a).

Though this explanation crucially depends on the placement of "attributives"
in VSO and SOV languages, it leaves one fundamental question unanswered,
namely, why the basic word order of a language (the order of subjects,
verbs and objects) should influence the placement of attributives in the
first place. Thus, the avoidance of center-embedding and conjunction
juxtaposition offers only a partial explanation of the data we have
considered and then only in languages with VSO and SOV word order. The
explanation of these data which I will propose does not attribute any
aspect of word order to the avoidance of center-embedding, however it
does rely heavily on the role of grammatical markers in sentence process-
ing which is therefore the topic of the next section.

3. Parsing and the placement of grammatical markers

From the perspective of language parsing, one of the basic questions
which must be addressed in an investigation of word order is why natural
languages may differ with respect to the placement of grammatical markers
such as prepositions. Given that speech is linearly ordered in time and
that parsing proceeds as the words of a sentence are received, it is
rather surprising that languages may differ in this respect. What is so

182

unexpected about this fact is that prepositions carry very important information about the type of phrase which follows, and thus we might have expected all natural languages to be prepositional, since postpositions would not be very helpful in identifying phrase types.

However, Frazier and Fodor (1978) observe that postpositions would be very useful in the two stage model of the parser which they propose. In this model, the first stage parser, or Preliminary Phrasal Packager, must incorporate incoming lexical items into a phrase marker as those items are received. Because the capacity of the first stage parser is restricted by the limitations of short-term memory, it may only hold roughly seven words at a time. Thus, unless a sentence is very short, at some point the first stage parser will have to end the phrasal package it is constructing and send it off to the second stage parser in order to have the capacity to deal with subsequent portions of the sentence. And, to structure the lexical string into phrasal packages, the first stage parser must make decisions about where to cut the lexical string on the basis of the limited amount of local information available in its restricted view of the sentence. Hence, to make intelligent decisions about where to end one phrasal package and begin a new one, it must have access to some sort of superficial marker of phrasal boundaries. But of course it doesn't matter whether it is the beginnings or the ends of phrases which are marked; when it is approaching the limits of its memory capacity, the phrasal package may either end its current phrasal package after a post-position, in a language like Japanese or Basque, or it may begin a new phrasal package with a preposition in a language like Thai or English. Thus, for purposes of intelligent phrasal packaging, either prepositions or postpositions will serve as effective cues to phrasal boundaries.
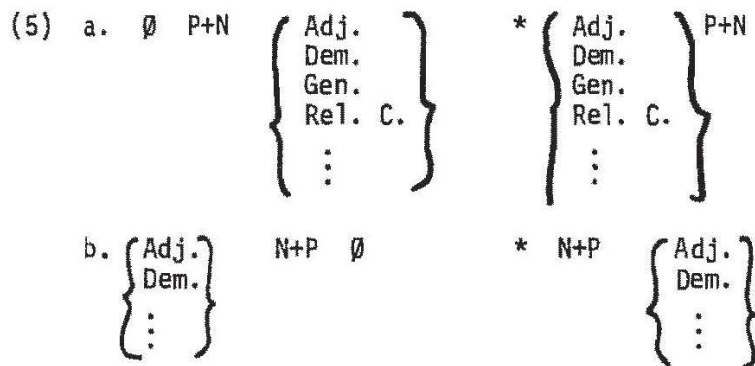
From this account of the role of prepositions and postpositions in parsing, there immediately follows a further prediction, namely, that languages should never allow "in-positions", since the occurrence of a grammatical marker in the middle of the phrase would not help the parser to construct intelligent phrasal packages. That is, an "in-positional" language would make phrasal packaging very difficult since the parser would have no readily accessible signal to rely on in deciding where the lexical string can be divided -- it could establish a phrasal package boundary neither before nor after an "in-position". Thus I propose the following Universal which I shall call the Inpositional Universal: a non-lexical head of a phrase may occur before or after the other constituents of the phrase but it may not occur embedded in that phrase. Assuming that N, V, A and P (preposition and postposition) are the only head-of-phrase categories and that of these categories only P is nonlexical (i.e. a closed class item), the Inpositional Universal may be formally stated as a constraint on $\overline{X}$ theory, as shown in (4).

(4) In the rule scheme: $X^n \longrightarrow x_i \ X \ x_j$, if X is non-lexical, either x_i or x_j must be null.[1]

I will now argue that the placement of grammatical markers is central in determining the expected word order of natural languages.

4. Grammatical markers and word order

In languages where non-lexical heads (prepositions and postpositions) occur as bound forms, the Inpositional Universal imposes a quite powerful constraint on the possible word order of the language. If non-lexical heads are prefixed to the head noun of their complements, adjectives, demonstratives, numerals, genitives, relative clauses, etc. must occur after the head noun, as shown in (5a), since otherwise these items would occur to the left of the preposition, thereby violating the Inpositional Universal. Similarly, in languages where non-lexical heads are suffixed to the head noun, these items must precede the head noun, as shown in (5b), if they are not to violate the Inpositional Universal.

(5) a. Ø P+N { Adj. / Dem. / Gen. / Rel. C. / ⋮ }   * { Adj. / Dem. / Gen. / Rel. C. / ⋮ } P+N

   b. { Adj. / Dem. / ⋮ } N+P Ø   * N+P { Adj. / Dem. / ⋮ }

Thus, in a language like Turkish, where non-lexical heads are expressed as suffixes on the head noun of their complements, the Inpositional Universal will by itself account for the fact that adjectives, relative clauses, etc. are placed prenominally.[2]

However, the Inpositional Universal will not account for the generalizations concerning word order in languages where non-lexical heads occur as free forms, because in these languages modifying items and phrases may intervene between a non-lexical head and the head of its complement without violating the Inpositional Universal, as shown in (6).

(6) a. Ø P { Adj. / Dem. / Gen. / Rel.C. } N   b. N { Adj. / Dem. / Gen. / Rel.C. } P Ø

Thus some further principle must be invoked to account for the expected word order in these languages. Specifically, the generalization in these languages seems to be that, in the unmarked case, optional nominal, verbal and sentential phrases (non-lexical nodes) must follow the head of the phrase they are in if the language is prepositional, and must precede it if the language is postpositional.[3] In the feature system of Jackendoff (1977), this generalization may be captured by a principle which specifies that phrases with the features $\begin{bmatrix} +subj. \\ +comp \end{bmatrix}$ (i.e. NP, VP and S) which are

184

optional in the phrase structure rules are placed before the head of the phrase that contains them in postpositional languages, and after the head in prepositional languages. This principle, which I shall call the Head Adjacency Constraint, will prevent optional nominal, verbal and sentential elements from intervening between a nonlexical head and the head of its complement. The Head Adjacency Constraint will apply to constructions such as relative clauses, genitives, objects and sentential complements (placing them to the right of the head of the phrase that contains them in prepositional languages, and to the left of the head in postpositional languages). The Head Adjacency Constraint will not apply to adjectival phrases or specifiers such as quantifier phrases or degree phrases (since these items are not characterized by the features $\begin{bmatrix} +subj \\ +comp \end{bmatrix}$), and, assuming that the subject noun phrase of a sentence is obligatory in most languages, the constraint typically will not apply to subjects (i.e. will not govern the relative order of subjects and verbs). Of course, unlike the Inpositional Universal, which I believe to be a true language universal, the Head Adjacency Constraint is not exceptionless and thus must be considered a principle of markedness theory.

Before examining the descriptive adequacy of this constraint, we might ask why such a constraint should exist in the first place. Earlier I argued that the role of prepositions and postpositions in parsing was to guide the first stage parser in constructing intelligent phrasal packages. In order for phrasal packages to be "intelligent", obviously they must respect the constituent structure of the sentence. In addition, since the semantic interpretation of a sentence proceeds from left to right, as Marslen-Wilson has repeatedly emphasized on the basis of data obtained from speech shadowing experiments, ideally it should be possible for the items contained in a phrasal package to be semantically interpreted as a unit, even though this interpretation may later have to be supplemented by information contained in subsequent phrasal packages. But, in this case, intelligent phrasal packaging will depend not only on the operation of the first stage parser but also on the design of the language. If, for example, a prepositional language placed the complements of a noun phrase before the head noun of that noun phrase, the first stage parser would be handicapped by the design of the language since it would have no choice but to package the complement together with the preposition. And, if the complement were fairly long, the limited capacity of the first stage parser would not permit it to incorporate the head noun of the preposition's complement in the same phrasal package. But in this case, since it is the preposition and the head of its complement which form a semantically integrated unit - not the preposition and optional modifiers such as relative clauses - the design of the language would force the parser to construct unintelligent (i.e. temporarily uninterpretable) phrasal packages. Hence, the very notion of intelligent phrasal packaging leads us to expect that there should be some constraint like the Head Adjacency Constraint which imposes a restriction on the elements which may intervene between the non-lexical head of a phrase and the head of its complement.

The Head Adjacency Constraint will account for the fact that preposi-

tional languages typically place relative clauses postnominally and
genitives after their governing noun; whereas, in postpositional languages
the ordering is just the reverse. It should be noted that this mirror-
image effect provides support for the basic notion of Core Grammar, that
setting one parameter of a language may significantly restrict the class
of possible or expected grammars of the language.

And, it appears that the correlation between basic SOV word order
and postpositions, and basic VSO and SVO word order and prepositions will
also fall out of the Head Adjacency Constraint since it predicts that
verbal objects will precede their head (i.e. the V) in postpositional
languages and will follow their head in prepositional languages. In
fact, the Head Adjacency Constraint may also account for the tendency
for prepositional languages to be right-branching and for postpositional
languages to be left-branching since consistent placement of complements
on the right (in prepositional languages) is bound to lead to a predominance
of right-branching, whereas consistent placement on the left (in post-
positional languages) will result in a predominance of left-branching
structures.

Though surely it is necessary to look at data from far more languages,
in terms of the thirty languages analyzed by Greenberg, all but four
(Finnish, Guarani, Quechua and Songhai)[4] abide by the Head Adjacency
Constraint and thus are unmarked. And, according to the account proposed
here, the fact that SVO languages tend to place relative clauses post-
nominally is not simply an accident, rather it follows automatically
from the fact that prepositional languages tend to place all optional
nominal, verbal and sentential elements after the heads of the phrases
they occur in. Moreover, it is not necessary to invoke an unexplained
correlation between the placement of attributives and basic (subject,
verb, object) word order to explain why the placement of non-lexical
heads is tied to the placement of verbal objects.

It should be emphasized that according to the explanation I have
proposed the placement of non-lexical heads plays a central role in deter-
mining the expected word order of languages. That this should be the case
is hardly surprising given that non-lexical, or closed class, items are
drawn from a small set of easily and rapidly identifiable items which
appear to be crucial to the operation of the sentence parsing device.
In fact, recent experimental findings by Bradley (1978) suggest that there
is a separate lexical retrieval system used to access closed class items
which, along with other evidence, attests to the special role that closed
class items play in the structural analysis of a sentence. And the fact
that it is a head-of-phrase category which is important in establishing the
basic ordering of elements in natural languages is also not very sur-
prising given that heads of phrases are obligatory in all natural languages.
These observations do, however, raise the possibility that it is not
simply non-lexical heads of phrases which are important in determining
word order but rather all non-lexical or closed class items.

There is in fact considerable evidence that languages tend to place
all closed class items either consistently on the left or consistently

on the right of the other constituents of the phrase they occur in.
We may account for this tendency by simply generalizing the Inpositional
Universal so that it applies to all closed class items, as shown in (7).

(7) In the rule schema: $Y \rightarrow x_i \ X \ x_j$, if X is nonlexical, x_i or x_j
must be null.

Though there are some exceptions to the Generalized Inpositional Constraint
(and thus it must be regarded as a principle of markedness, rather than
a true universal) the constraint does account for a number of quite
disparate facts. It will explain why an inflected auxiliary precedes
the main verb in prepositional languages, but follows the main verb in
postpositional languages (Greenberg's universal 16). The constraint also
explains why conjunctions (complementizer and adverbial conjunctions)
occur clause-initially in prepositional languages and clause-finally in
postpositional languages. And, assuming that (yes-no) question particles
must be generated in or moved to complementizer position, this in turn
will explain Greenberg's universal 9 (below).

Universal 9. With well more than chance frequency, when question
particles or affixes are specified in position by reference to
the sentence as a whole, if initial, such elements are found
in prepositional languages and, if final, in postpositional languages.

Kuno's observation that languages typically do not mark both the
beginnings and ends of clauses also may be explained by appealing to the
Generalized Inpositional Constraint and the role of cloased class items
in sentence processing. Assuming that there is a separate lexical retrieval
system for closed class items, when the first stage parser must decide
where to end its current phrasal package, it need only know that an item
has been accessed by means of the closed class retrieval system to decide
where the lexical string may be cut, if the language being analyzed abides
by the Generalized Inpositional Constraint. By contrast, in a language
which does not abide by the Generalized Inpositional Constraint, the parser
could not establish the phrasal package boundary until after it had
retrieved the lexical entry for that particular closed class item and
obtained distributional information about the placement of that item.
Likewise, in a language which marked both the beginnings and ends of
clauses (or phrases), simply knowing that an item had been accessed through
the closed class retrieval system would not be sufficient to determine
where the lexical string should be cut, since the phrasal packager could not
determine on the basis of that information alone whether it was dealing
with a clause-final marker, in which case the lexical string could be cut
after the closed class item, or with a clause-initial marker which would
indicate that the current phrasal package should be terminated before the
closed class item.

The time it takes to retrieve a single item from the lexicon and
extract information about its distribution may not seem very substantial,
and thus the savings to the phrasal packager may appear to be quite
insignificant. However, in view of the time pressures of sentence parsing,

this savings may actually be quite considerable, especially in light of
the fact that the reason why the first stage parser is ending its current
phrasal package is precisely because it has approached the limits of its
memory capacity and will not be able to cope with new incoming lexical
material unless it can manage to ship current material off to the second
stage parser. Hence, the consistent placement of all grammatical markers
which is stipulated by the Generalized Inpositional Constraint would
save the phrasal packager some time at exactly that point in the parsing
process when it is under the most severe time pressure.

5. Parsing and constraints on grammar

We will turn now to some general questions concerning the relation of
language parsing and the structure of natural languages. Though a variety
of different issues will be addressed, each of them bears on the question
of when, and how, the exigencies of sentence parsing will influence the
structure of natural languages.

To begin, it should be noted that the Head Adjacency Constraint is
a very peculiar constraint from the perspective of the language parsing
device. Assuming that the desire to prevent long constituents from inter-
vening between a non-lexical head and the head of its complement is in fact
the motivation for the Head Adjacency Constraint, it is quite odd that
languages have not gone a step further and fully accomodated themselves
to the needs of the parser by directly preventing all long constituents
from separating a nonlexical head from the head of its complement. In
other words, it must be explained why such a large number of natural
languages only indirectly restrict the placement of long constituents
by constraining the placement of certain types of constituents (i.e.
nominal, verbal and sentential phrases) when what the parser really wants
is for the language to impose a restriction on the length of whatever
constituents happen to intervene between P and N in the configurations
shown in (8).

(8) a.

P^n
N^n
P X N

b.

P^n
N^n
N X P
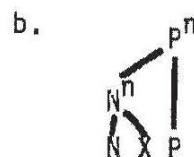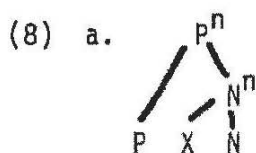
Presumably the reason why natural languages do not develop length restric-
tions of this latter variety is that they simply can not incorporate
this kind of restriction. Though long constituents may form a natural
class from the perspective of the sentence parsing device, apparently
they do not form a natural class in terms of the grammars of natural
languages. The fact that languages typically do not contain rules which
state that a constituent - regardless of its grammatical category - may
undergo some grammatical operation (i.e. movement, deletion, etc.) or may
occur in a particular type of phrase providing only that the constituent
is relatively long surely indicates that long constituents do not consti-
tute a natural class in terms of the grammar. And, therefore, it is not
particularly surprising to find that constraints on grammars can not be
formulated in terms of this (unnatural) class. Thus, what I am suggesting

188

is that a constraint like the Head Adjacency Constraint which applies to certain types of constituents is simply the best device the grammar has available for imposing a restriction on the placement of long constituents. (The importance of this observation will be discussed below.)

Apart from restricting the placement of certain types of constituents, it is not clear whether there is any other means the grammar can employ to prohibit long constituents from occurring in particular positions in a sentence. However, a principle like the Surface Recursion Restriction proposed by Emonds (1976) is at least a likely candidate.

Surface Recursion Restriction: Given a surface configuration of the form $[_{H_i} \ldots A \ldots H_i \ldots]$, if the base rules permit right sisters $H_k \ldots$ to H_i, then $A \neq XSY$, $A \neq XPPY$, where PP dominates a lexical preposition, and $A \neq WAZ$, where W and $Z \neq \emptyset$. In such cases we say that A does not exhibit free recursion.

In fact, this restriction may be motivated by essentially the same parsing considerations that motivate the Head Adjacency Constraint, since one of its effects is to limit the length of the major phrases which may intervene between a nonlexical head and the head of its complement, by restricting pre-head recursion in (prepositional) languages where major phrases may freely occur to the right of a head of phrase (and, presumably, a general formulation of the principle would restrict post-head recursion in (postpositional) languages where major phrases may freely occur to the left of the head of a phrase). Thus, restricting free recursion in phrases which occur in certain positions may be another device available to the grammar for limiting the length of constituents. Of course, like the Head Adjacency Constraint, the Surface Recursion Restriction only indirectly imposes a length restriction on constituents since it does not exclude all and only long constituents from preceding the head of a phrase. And thus it too is somewhat peculiar from the perspective of the language parsing device.

Before taking up another example where the grammar has incorporated a constraint which only indirectly manages to exclude constructions which create difficulties for the parsing mechanism, it should be noted that these "indirect" constraints provide evidence against the Performance Grammar Only (PGO) theory which claims that there is no autonomous mental representation of linguistic knowledge apart from the actual performance routines used to produce and interpret sentences (cf. Lakoff and Thompson, 1975a, 1975b). As Fodor (1978) has pointed out, the PGO theory makes the minimal assumptions concerning the mental apparatus available to speakers and hearers and thus it is incumbent on those who claim that there is a mentally represented competence grammar in addition to these performance routines to provide evidence for their position.

Fodor presents two arguments in defense of the Competence Grammar theory. First, she argues that movement and deletion rules, as opposed to copying rules, do not facilitate either speaking or hearing nor do

they increase the expressive power of the language, and thus the prevalance
of these rules in natural languages may only be explained on the assumption
that they are highly valued by the format for the mental representation
of linguistic knowledge (and, hence, there must be such a format, or
"competence grammar"). Her second argument is that it is necessary to
assume the existence of competence grammars to explain why some constraints,
though they may be motivated by parsing considerations, are "concretized"
or absolute, whereas other constraints which are motivated by the opera-
tion of the parser are "squishy" and only stylistically preferred. For
example, Clark and Clark (1968) and Bever and Weksel (reported in Bever
(1970)) have shown that sentences are stylistically preferred if the main
clause of the sentence precedes any subordinate clauses, however hearers
certainly do not judge sentences to be ill-formed if the subordinate clause
happens to precede the main clause. By contrast, a sentence which violates
the Nested Dependency Constraint (cf. Fodor, 1978) is judged to be ill-
formed, even when hearers are able to determine what the structure and
meaning of the 'sentence' would have been if it were generated by the
grammar. Fodor's argument is that the difference between an absolute
preference like the preference for nested dependencies (in cases of
ambiguity) and a stylistic preference like the preference for subordinate
clauses to follow main clauses is only explained on the assumption that
absolute preferences are those which have actually been incorporated
into the grammar.

The existence of indirect constraints like the Head Adjacency
Constraint and the Surface Recursion Restriction provides another argu-
ment for the existence of competence grammars, since if there were no
autonomous mental representation of linguistic knowledge, there would
be no explanation for the fact that these constraints do not directly
exclude all and only the perceptually complex constructions which
motivate the constraints. In other words, the only way to account for
tne particular form that constraints like Head Adjacency assume is by
appealing to the fact that long constituents do not form a natural class
in terms of the format for the representation of linguistic knowledge;
if one denies the existence of that format, then it is difficult to see
how the indirect nature of these constraints could be explained, especially
since tne length of constituents is exactly the sort of variable which
the sentence parsing routines are concerned with (cf. Frazier and Fodor,
1978; Frazier, 1979).

Perhaps one might defend the PGO theory by arguing that the language
must resort to indirect restrictions in order to maintain the expressive
power of the language or to avoid complicating the sentence production rou-
tines. However, in the case of the Head Adjacency Constraint there is
simply no reason to believe that a more direct restriction on the place-
ment of all long constituents would reduce the expressive power of the
language (since what is at issue is only the placement of long constituents)
or that it would complicate the sentence production routines (since, if
anything, we would expect that speakers, like listeners, would prefer to
place items which form a coherent semantic unit in close proximity to one
another). Thus, I see no way to explain the indirect nature of the Head
Adjacency Constraint without appealing to the properties of competence
grammars.

190

And, as would be expected according to the Competence Grammar theory, there are quite a few grammatical constraints which, though motivated by parsing considerations, do not exclude exactly the set of constructions which complicate the sentence parsing process. The prohibition against deleting complementizers in subject relative clauses in English is one familiar example of a restriction which facilitates the sentence parsing device. Bever and Langendoen (1971) have argued that the reason why complementizers can not be deleted in sentences like (9), where the subject of a clause has been relativized, is because this would lead to these sentences being misanalyzed by the Main Clause Strategy (10), i.e. hearers would incorrectly interpret the sequence <u>the girl hit the man</u> to be the main clause of the sentence (see Frazier (1979) for a discussion of the Main Clause Strategy).

(9) *The girl Ø hit the man was angry.

(10) Main Clause Strategy:

$$X_1 - \text{Nominal} - V_{finite} - (\text{Nominal}) - X_2 \rightarrow X_1[\text{Nom } V_f \text{ (Nom)}] X_2$$

What is important to note is that sentences like (11a) contain exactly the same misanalysis as sentences like (9), nevertheless these sentences have not been excluded from the grammar. (Note that hearers do accept sentences like (11a) as being grammatical once they have determined the correct structure of the sentence.)

(11) a. The horse raced past the barn fell.
     b. The horse that was raced past the barn fell.

(12) a. The horse ridden past the barn fell.
     b. The horse that was ridden...

(13) a. The woman speaking Swahili turned out to be an American.
     b. The woman that was speaking...

Thus the question again arises as to why the language did not exclude exactly the set of sentences which create problems for the parsing device; why, instead, are sentences like (9) excluded from the language, while those like (11a) are not, even though they pose precisely the same problem for the sentence parsing routines.

From the perspective of the sentence parsing device, sentences (9) and (11a) are members of a natural class of sentences, namely the class of sentences misanalyzed by the Main Clause Strategy. Thus according to the PGO theory we would expect this entire class of sentences to be excluded from the language. That is, if the only mental representation of linguistic knowledge were in terms of parsing routines like the Main Clause Strategy, then surely the prediction is that a set of constructions which run afoul of the strategy could be excluded from the language as a class. Thus, we have another example where the PGO theory can not account for the disparity between the class of sentences which create problems for the parsing device, and the class of sentences excluded from the language.

The Competence Grammar theory, on the other hand, can explain this disparity since within the grammar there is no natural class of sentences which includes sentences (9) and (11a) but does not also include sentences like (12a) and (13a); hence, for the grammar to exclude a sentence like (11a), it would also have to exclude a large set of sentences like (12a) and (13a) which do not pose any particular problem for the sentence analysis routines. In short, the restriction against deletion of complementizers provides another example of the grammar indirectly responding to the exigencies of sentence parsing in a fashion which is only explained by appealing to the properties of competence grammars.[5]

So far I have argued that a set of constructions which form a natural class from the perspective of sentence parsing may not form a natural class in terms of competence grammars and used this fact to explain the indirect nature of several constraints which are motivated by parsing considerations and to argue for the existence of competence grammars. However, I have not yet addressed the question of when the grammar will respond to the needs of the parsing device. Fodor (1978) has argued that the likelihood of a language changing in response to the exigencies of sentence parsing is a function of how severe the parsing problem is and, secondly, how easy it is for the grammar to solve the parsing problem. However, the examples discussed here suggest that a third factor is also important, namely, how "directly" the grammar can respond to the demands of the parsing device. The larger the disparity between the class of sentences which create parsing problems and the class of sentences which the grammar can exclude (given its restricted vocabulary), the less likely the grammar will be to change in response to the demands of the parser. With respect to the deletion of complementizers in English, we have already seen that the severity of the parsing problem is no different in sentences like (11a) than in sentences like (9). And presumably it would not be any more difficult for the grammar to develop a constraint which would exclude sentences like (11a) than it was to develop the constraint which excludes sentences like (9). Hence, the fact that the grammar did develop the latter constraint but not the former (i.e. a constraint which would exclude sentences like (11a) from the language) may only be attributed to the difference in the directness with which the grammar could respond in the two cases.

I have just presented evidence for the following principle, which I shall call the "Direct Exclusion" principle.

> Direct Exclusion: The likelihood of a language changing to accomodate itself to the needs of the sentence parsing device is, in part, a function of how directly the grammar may respond to the parsing problem.

Implicit in my argument for this principle was the assumption that, when the grammar of a language does incorporate some restriction in order to facilitate sentence parsing, it will incorporate a restriction which excludes from the language only as many sentences as is necessary given the parsing problem. Thus, my argument for the Direct Exclusion principle

192

(as well as the arguments presented in Fodor (1978)) implied the "Minimal Exclusion" principle:

> Minimal Enclusion: When a language responds to the exigencies of sentence parsing by developing a constraint which excludes perceptually complex constructions from the language, it will incorporate whatever constraint allows it to exclude the fewest sentences from the language and still alleviate the parsing problem.

Though both of these principles may appear to be quite obvious, I think that they are interesting for a number of reasons. First, the Minimal Exclusion principle is of interest because, as it stands, it is clearly false since it totally disregards the evaluation metric (i.e. the "minimal"-in the above sense-constraint that a language might incorporate might involve much more major changes in the grammar than some less "minimal" constraint). Hence, violations of the Minimal Exclusion principle should provide valuable information about the evaluation metric. (I will return to this point below.) Furthermore, though the predictions of these principles are not entirely explicit in the absence of a more detailed theory of universal grammar, the principles may help elucidate the reasons why different languages develop somewhat different constraints to cope with the same parsing problem. For example, it appears that there is a universal constraint which prohibits intersecting dependencies of "fillers" and "gaps" in cases of ambiguity (cf. Fodor, 1978; Engdahl, 1979 and references therein). Fodor shows that the Nested Dependency Constraint is motivated by the operation of the sentence parsing device, however she notes that it is not intersecting filler-gap dependencies per se which create parsing problems, but rather the assignment of intersecting dependencies in cases of ambiguity (i.e. intersecting dependencies are permitted, and are not particularly difficult to parse, in cases where the grammar does not license a nested assignment of dependencies).

In English and Norwegian the Nested Dependency Constraint is formulated as a straightforward No Ambiguity Constraint, i.e. an intersecting assignment of filler-gap dependencies is permitted if and only if a nested assignment is not permitted by the grammar of the language under analysis (cf. Fodor, 1978; Engdahl, 1979). For example, in English an intersecting assignment of fillers and gaps is not permitted in sentence (14), as shown in (14b), because a nested assignment of dependencies (shown in (14a)) is permitted. However, in (15), where the two fillers (wh-constituents) are of different formal types, a nested assignment of dependencies is not licensed by the grammar and thus intersecting dependencies are permitted.

(14) a. Which boxes$_i$ are these containers$_j$ easy to store $\underline{\quad}_j$ in $\underline{\quad}_i$?
     *b. Which boxes$_i$ are these containers$_j$ easy to store $\underline{\quad}_i$ in $\underline{\quad}_j$?

(15) *a. Which crimes$_i$ does the FBI not know how$_j$ to solve __$_j$ __$_i$?
    b. Which crimes$_i$ does the FBI not know how$_j$ to solve __$_j$ __$_i$?

Similarly, in cases of ambiguity, only a nested assignment of filler-gap dependencies is permitted in Norwegian, as shown in (16). However, in sentences where a nested assignment of dependencies would violate some rule of the grammar, as in (17a)(where a nested assignment of fillers and gaps would violate the selection restrictions of the verb), an intersecting assignment of dependencies is allowed, as shown in (17b).

(16) a. Dette er piken$_i$ som laereren spurte hvilken gutt$_j$ vi trodde

         __$_j$ var sint pa __$_i$?

   *b. Dette er piken$_i$ som laereren spurte hvilken gutt$_j$ vi trodde

         __$_i$ var sint pa __$_j$?

         'This is the girl that the teacher asked which boy we thought ___ was mad at __.'

(17)*a. Hvilke elever$_i$ husker du hvilke opgaver$_j$ laereren ba __$_j$

        lφse __$_i$?

     b. Hvilke elever$_i$ husker du hvilke opgaver$_j$ laereren ba __$_i$

        lφse __$_j$?

        'Which students do you remember which problems the teacher asked __ to solve __?'

The particular formulation of the Nested Dependency Constraint which is found in English and Norwegian (i.e. its formulation as a No Ambiguity Constraint) is not the simplest way for the grammar to exclude ambiguous intersecting dependencies; surely by any evaluation metric it would be simpler for a grammar to develop an absolute or across-the-board prohibition of all intersecting dependencies. Of course, though this absolute form of the Nested Dependency Constraint would be simpler, it would also amount to a form of "overkill" since it would exclude a larger range of sentences from the language than is necessary or warranted given the demands of the parsing device. Thus, we must invoke the Minimal Exclusion principle to account for the fact that the Nested Dependency Constraint was incorporated as a No Ambiguity Constraint in English and Norwegian, rather than as an absolute prohibition against all intersecting dependencies.

Though the Swedish version of the Nested Dependency Constraint is quite different from the English and Norwegian version, it too appears to have been shaped by the pressure to exclude all and only as many sentences as necessary given the parsing problem associated with ambiguous intersecting dependencies. Engdahl (1979) shows that in Swedish intersecting dependencies are permitted if and only if a bound pronoun occurs in the position of the first "gap", as shown in (18).

194

(18) Swedish: F_i F_j G_j G_i (nested) $*F_i$ F_j P_j G_i

   $*F_i$ F_j G_i G_j (intersecting) F_i F_j P_i G_j

where G=deletion site; P=bound pronoun

That is, in a sentence like (19), where there is no bound pronoun, only a nested assignment of dependencies is permitted; by contrast, in sentence (20), where there is a bound pronoun in the position of the first "gap", only an intersecting assignment of dependencies is allowed.

(19) a. Här är boken$_i$ som ingen minns vilken författere$_j$ de
   gav ___$_j$ N.P. för ___$_i$.

   *b. Här är boken$_i$ som ingen minns vilken författere$_j$ de
   gav ___$_i$ N.P. för ___$_j$.

   'Here is the book that no one remembers which author they
   gave ___ the Nobel Prize for ___.'

(20) *a. Här är författaren$_i$ som ingen minns vilken bok$_j$ de gav
   honom$_j$ N.P. för ___$_i$.

   b. Här är författaren$_i$ som ingen minns vilken bok$_j$ de gav
   honom$_i$ N.P. för ___$_j$.

   'Here is the author that no one remembers which book they
   gave him the N.P. for ___.'

   Notice that the Swedish version of the Nested Dependency Constraint goes even further than the English or Norwegian version in narrowing the class of sentences excluded from the language. That is, in addition to excluding only ambiguous cases of intersecting dependencies, Swedish has actually reduced the size of the class of ambiguous intersecting dependencies by using resumptive pronouns to disambiguate a large class of sentences with intersecting dependencies (which, because they are ambiguous, will not be excluded from the language by the Nested Dependency Constraint).

   But if universal grammar permits the use of resumptive pronouns for this purpose, why haven't languages like English and Norwegian also adopted this device? In other words, in clear violation of the Minimal Exclusion principle, English and Norwegian have not utilized a device which would have permitted fewer sentences to have been excluded from these languages by the Nested Dependency Constraint. Thus, to maintain the Minimal Exclusion principle, it must be explained why the use of resumptive pronouns was accessible in Swedish, but was not accessible in English and Norwegian.

   Engdahl (1978) notes that in Swedish resumptive pronouns may optionally occur in sentences where an element has been extracted from a site which is embedded more than two sentences down (as shown in (21) and (22)); whereas,

in Norwegian and English, the use of a resumptive pronoun is considered colloquial at best.

(21) Här är boken $[_{S_1}$ som jag läste __/*den.

    'Here is the book that I read __/*it.'

(22) Här är boken $[_{S_1}$ som läraren fragade $[_{S_2}$ om vi mindes

    $[_{S_3}$ vem some skrivit __/(den).

    'Here is the book that the teacher asked if we remembered who
    that wrote __/(it).'

Thus it may be that the grammar of a language will deploy a device which is already in use in the grammar of the language (e.g. the use of resumptive pronouns in Swedish) to narrow the class of sentences excluded from the language by some constraint, but that the grammar will not adopt some totally new device for that purpose (i.e. an operation which is not already in use in the grammar of that language, such as the use of resumptive pronouns in English). This suggests that simply because some device is permitted by universal grammar this does not guarantee that the device will be equally accessible to the grammars of all natural languages.

Lightfoot (1977) has argued that it is not necessary to develop a theory of language change in order to predict when changes will occur in a language and what those changes will be; he argues that there is no need for historical principles or a "separate" theory of language change since a suitably restrictive theory of grammar, together with some basic statements about language acquisition, will serve this purpose. The present investigation suggests that this will only be possible if we develop an evaluation metric which not only specifies which of an array of possible grammars is most highly valued (i.e. which grammar a child will choose when presented with a given set of data), but also specifies or ranks the accessibility of the various devices available to grammars with respect to a particular grammar.[6] In the absence of this latter type of evaluation or "accessibility" metric, it will be necessary to resort to an independent theory of language change to predict how a given language will respond to the exigencies of sentence parsing.

The examples discussed above illustrate that there is often a tension between the ease with which a grammar may respond to a particular parsing problem and the directness with which it may respond. To predict the precise way that the grammar of a given language will resolve this tension, it will be necessary to specify all of the factors that determine how accessible each of the devices available to a grammar is. Though at present our theories may not be sufficiently detailed to do this, given psycholinguistic theories which lead us to expect certain types of restrictions to occur in natural languages, we are in a position to ask why the grammar of some language failed to incorporate a particular restriction or why the restriction assumed a peculiar form in some language.

196

And surely this approach will place the properties of grammars and the evaluation metric in relief and lead to more detailed and explanatory theories of language structure.

Notes

*I am very grateful to Chuck Clifton, Mike Flynn, Janet Fodor and Lisa Selkirk for many helpful comments on an earlier draft of this paper.

[1]Intensifiers such as the word right in the phrase right on top constitute a counterexample to the above statement of the Inpositional Universal if it is correct to analyze these items as specifiers of prepositional phrases. Hence, it may be necessary to weaken the Inpositional Universal slightly to allow a restricted set of closed class items to occur to the left of a non-lexical head of phrase in prepositional languages and to the right of a non-lexical head in postpositional languages.

[2]If non-lexical heads of phrases are not present in the base in languages where these items occur as bound forms then, obviously, the present statement of the Inpositional Universal will have to be revised so that it will apply to surface structures rather than base structures.

[3]I must emphasize again that this generalization is based primarily on the data collected by Greenberg and thus, given a fuller array of data, it may well turn out to be possible to formulate a much broader generalization than was warranted given the limited set of data considered here.

[4]According to Greenberg (1965), Finnish, Guarani, Quechua and Songhai are each postpositional languages but, in violation of the Head Adjacency Constraint, relative clauses may precede their head noun in these languages (though apparently the alternative ordering is also permitted in Finnish) and, further, in Finnish, Guarani and Songhai S-V-O word order is dominant.
Clearly it should be determined whether it is simply an accident that these particular languages violate the Head Adjacency Constraint or whether there is some structural characteristic which distinguishes these languages from other postpositional languages (e.g. these "marked" languages may have some characteristic which mitigates the effects of non-optimal placement of constructions such as relative clauses and objects, the languages might be in the process of changing from one language type to another, etc.).

[5]Chomsky and Lasnik (1977) suggest that sentences like (i) can not be excluded from the grammar (at least by means of a surface filter) because a surface filter could not distinguish (i) from (ii), "taking

account of morphological ambiguities."

   (i) The horce raced past the barn fell.
  (ii) The horse ridden past the barn fell.

Thus their account also relies on the properties of competence grammars to explain the grammaticality of (i).

[6]On the basis of the examples discussed in this paper it would seem reasonable to suppose that it is generally simpler for the grammar to respond to a parsing problem by excluding some existing construction of the language than by "extending" a device already in use in the grammar, which in turn is simpler than adding a totally new construction to the language. If this type of principle is included in an "accessibility" metric, I see no reason to suppose that the accessibility metric will apply exclusively to children acquiring a language. That is, I see no reason to believe that every innovation (i.e. extension of an existing construction or addition or a new construction) in a language is due to the language acquisition process. Rather, adults as well as children may be responsible for certain innovations in a language and thus the accessibility metric would specify the willingness with which speakers (of all ages) will accept an innovation into the language. For example, in English one occasionally hears speakers (including adult speakers) use resumptive pronouns, although the same speakers often consider sentences with resumptive pronouns to be ungrammatical. In a situation like this, the accessibility metric may reflect how deviant speakers will judge the innovative construction to be and thus how willing speakers will be to accept the innovation into the language (e.g. to use the new construction generally and not simply as a means of coping with an unusual situation where the constraints of their language or perhaps memory constraints make it difficult for speakers to express themselves).

References

Bever, T. G. (1970) The cognitive basis for linguistic structures, J.R. Hayes (Ed), Cognition and the Development of Language. New York: John Wiley and Sons.

Bever T. G. and D. T. Langendoen (1971) A dynamic model of the evolution of language, Linguistic Inquiry, 2, 433-463.

Bradley, D. (1978) Computational distinctions of vocabulary type. MIT Doctoral dissertation.

Chomsky, N. and H. Lasnik (1977) Filters and Control. Linguistic Inquiry, 8, 425-504.

Clark, H. H. and E. V. Clark (1968) Semantic distinctions and memory for complex sentences. Q. Journ. of Exp. Psych., 20, 129-138.

198

Emonds, J. E. (1976) A Transformational Approach to English Syntax. New York: Academic Press.

Engdahl, E. (1979) The Nested Dependency Constraint as a Parsing Strategy, Papers Presented to Emmon Bach by His Students. University of Massachusetts.

Fodor, J. D. (1978) Parsing strategies and constraints on transformations, Linguistic Inquiry, 9, 427-474.

Frazier, L. (1979) On comprehending sentences: syntactic parsing strategies. Indiana University Linguistics Club.

Frazier, L. and J. D. Fodor (1978) The Sausage Machine: a new two stage parsing model, Cognition, 6, 291-325.

Greenberg, J. H. (1963) Some universals of grammar with particular reference to the order of meaningful elements, J. H. Greenberg (Ed.), Universals of Language.

Jackendoff, R. (1977) $\overline{X}$ Syntax: A Study of Phrase Structure, Linguistic Inquiry Monograph Two.

Kuno, S. (1974) The position of relative clauses and conjunctions, Linguistic Inquiry, 1, 117-136.

Lakoff, G. and H. Thompson (1975a) Introducing Cognitive Grammar, in C. Cogen, H. Thompson, G. Thurgood, K. Whistler and J. Wright, eds., Proceedings of the First Annual Meeting of the Berkeley Linguistic Society, University of California, Berkeley, California.

Lakoff, G. and H. Thomson (1975b) Dative Questions in Cognitive Grammar, in R. E. Grossman, J. L. Sam, and T. J. Vance, eds., Papers from the Parasession on Functionalism, Chicago Linguistic Society, University of Chicago, Chicago, Illinois.

Lightfoot, D. (1977) Explaining syntactic change, McGill University manuscript.

Rosenbaum, H. (ms.) On perceptual difficulty as typological explanation.