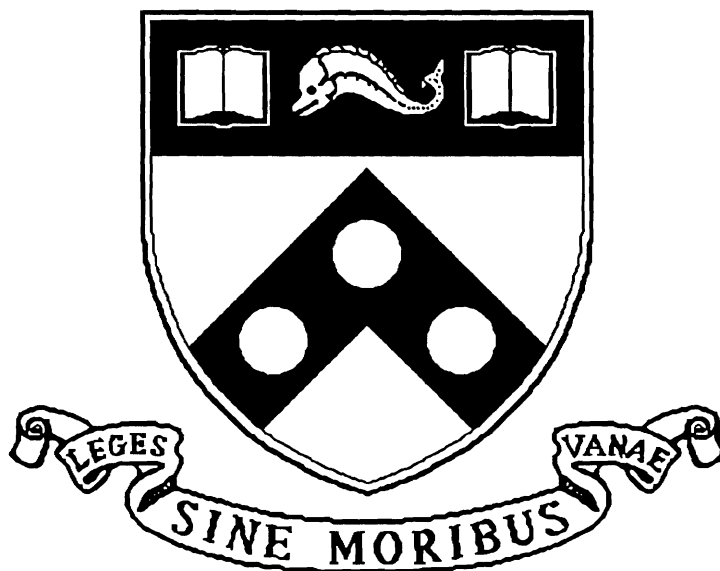


Grammars and Processors

MS-CIS-92-53
LINC LAB 231

Mark Steedman



University of Pennsylvania
School of Engineering and Applied Science
Computer and Information Science Department
Philadelphia, PA 19104-6389

September 1992

GRAMMARS AND PROCESSORS *

MARK STEEDMAN
University of Pennsylvania

September 2, 1992

§1 INTRODUCTION

The function of rules of syntax is simply to identify the corresponding rules of semantics. It is therefore to be expected that the rules of natural language syntax and semantics should be related in a straightforward way. Less obviously, but for similar reasons, it is also to be expected that the rules and structures involved in processing should be equally directly related, a version of what is usually called the “Competence Hypothesis”. The latter position has repeatedly been endorsed within the generative tradition (cf. Chomsky [20, p.10], [22, p.7]). These expectations arise not only from considerations of theoretical parsimony, but also from the observation that the development of language in children (and as far as anyone can tell the evolution of the language faculty in the species) is extremely rapid. Any increase in the complexity of the relations between these modules increases the explanatory burden on the theory of acquisition, evolution, or both.

It does not of course follow that observations concerning the different members of this trinity can be expected to be equally productive of insights into the nature of the system as a whole. Modern linguistics is founded on

*To appear in: H. Kamp and C.Rohrer, (eds.), *Aspects of Computational Linguistics*, Springer Verlag. Thanks to the following for reading the draft and for many helpful comments: Aravind Joshi, Mitch Marcus, Michael Niv, Phil Resnik, Bonnie Lynn Webber. research was supported in part by NSF grant nos IRI90-18513 and CISE IIP, CDA 88-22719 , DARPA grant no. N00014-90-J-1863, and ARO grant no. DAAL03-89-C0031.

the assumption that there is, in Chomsky's words, "little reason to question the traditional view that investigation of performance will proceed only so far as understanding of underlying competence permits" ([20, p.10]) and that equally (albeit in a weaker sense) there is "little evidence that 'intuition about meaning' is at all useful in the actual investigation of linguistic form" ([19, p.94]).

The insight that syntactic phenomena alone offer an effective and reliable source of data concerning the workings of the language faculty has been so productive of generalisations and results concerning its nature that it is quite inconceivable that we should abandon it now. It is worth calling to mind the reasons why it continues to bind us.

The methodological priority of competence over performance seems to follow from the exceedingly fortunate fact that the process of mapping natural language strings onto their interpretations happens to be one for which the abstract nature or "theory" of the computation can be specified at a more general level than the particular algorithm that is actually used to carry it out. It is quite unlikely that we could ever analyse a system as ramified as natural language if it were *not* decomposable, as Marr [58] points out.¹ We know that even for quite simple classes of language, there are very many processing algorithms. Exactly how many there are, and the number of degrees of freedom that they offer us to account for performance data, depends upon what further assumptions we are willing to make. However, in advance of knowing exactly what it is that the corresponding algorithms for natural language actually compute, we have no clear grounds for limiting these assumptions, so the search space is bound to be very much larger if we start from the performance end.

The methodological priority of syntax over semantics is not a necessary truth in the same sense, and is by no means total. In fact, however, we do not seem to have very reliable access to semantics at the level of sentence grammar. That is not to deny that we have strong intuitions about meaning

¹It is ironic that Marr, who offered a particularly clear interpretation of competence vs. performance as "theory of computation" vs. "algorithm", expressed some doubt whether natural language was in fact one of the systems that are decomposable in this way. While many more neurally-embedded and less symbolically-mediated aspects of cognition may have this character, it seems unlikely that any symbolic system like language could fail to be decomposable in this way.

at the level of propositions, referring expressions, and relations among them, such as coreference and entailment. However, this level of meaning seems to arise very far down the chain of language understanding. It is evidently too coarse to explain the very diverse ways in which the same propositional meaning can be grammatically realised as an (active, passive, topicalised, cleft, or whatever) sentence, subject to the demands of context. For that reason, this notion of meaning gives us very little help in understanding how sentences give rise to meanings. That process just seems to be very opaque to introspection.

These principles thus very strongly constrain the proper conduct of linguistic investigations. Nevertheless, commonsense continues to suggest that both semantics and the operations of the human sentence processor will in fact turn out to stand in the closest possible relation to natural language syntax, when we come to understand them properly. In particular, in the absence of other factors, we would certainly expect syntax and semantics to stand in what Bach [5] has called a “rule-to-rule” relation, according to which each syntactic rule corresponds to a rule of semantic interpretation. In the first part of the paper, I shall argue that this commonplace observation, if taken seriously, has some far-reaching implications for the theory of grammar. More contentiously, the second part of the paper will argue that we would expect the structures manipulated by the processor to be isomorphic to the constituents licensed by the grammar, an extreme version of Bresnan and Kaplan’s [17] “Strong Competence Hypothesis” that will be distinguished as the “Strict” Competence Hypothesis. The claim will be that the theory of grammar defended in the first part adheres more closely to this very restrictive hypothesis than the alternatives do. This position has recently been opposed by Stabler [81], and his criticisms will also be considered in the second part of the paper.

PART I: COMPETENCE

§2 COMPETENCE AND CONSTITUENCY

The artificial languages that we design ourselves, such as predicate logic or the programming language Lisp, exhibit a very strong form of a “rule-to-rule”

relation between syntax and semantics as they are defined in the textbook or reference manual. The rule-to-rule hypothesis as it is used here can be stated as follows:

- (1) THE RULE-TO-RULE ASSUMPTION: syntactic rules and syntactic constituents are paired with semantic translations.

The fact that syntax functions so straightforwardly as the handmaiden of semantics in these languages is so reasonable and desirable that it might be expected to transcend all particulars of function and content. It is reasonable to expect the natural system to exhibit the same property because it is hard to imagine any evolutionary pressure that would force it to be otherwise. Indeed, there is at least one identifiable force that can be expected to work positively to keep them in line. It arises from the fact that children have to learn the language, apparently on the basis of quite unsystematic presentation of positive instances alone. Since even such trivial classes of grammars as finite state grammars are not learnable from mere exposure to positive instances of the strings of the language (cf. Gold [40]), and since there is pretty clear evidence that no more explicit guidance is provided by adults (cf. Brown and Hanlon [13]), some other source of information, “innate” in the sense that it is available to the child prior to the acquisition of any specific language, must guide them. As has often been pointed out, the most promising candidate a priori is semantic interpretation or the related conceptual representation.² It is likely that this cognitive representation includes such grammatically relevant notions as actual and potential participants and properties of events, and the attentional focus of other conversational participants, as well as the phenomenal reflections of the physical situation.³

¶2.1 CONSEQUENCES OF THE RULE-TO-RULE ASSUMPTION

²In the context of modern linguistics, the suggestion goes back at least to Chomsky [20, p.59] and Miller [61]. Of course it is a much older idea. See Pinker [71] for a review of some proposed mechanisms, and see Gleitman [39] for some cogent warnings against the assumption that such semantic representations have their origin solely in present perception and the material world in any simple sense of that term.

³However inadequate our formal (and even informal) grasp on the child’s prelinguistic conceptualisation of the situation, there can be no doubt that it includes such aspects, for even non-linguistic animals have that much.

A number of consequences follow. The first, which follows from the expectation of transparency between syntax and semantics, is so strong and so utterly uncontentious that no theory of grammar has failed to observe it in spirit, though it is probably true to say that none have so far succeeded in following it to the letter. From the claim that syntax and semantics are rule-to-rule, it immediately follows that the syntactic entities that are combined by a syntactic rule must be semantically interpretable. (Otherwise, they could not be combined by the semantic interpretation of the rule). It follows that syntactic rules can only combine or yield *constituents*.

This condition, which has been called “The Constituent Condition on Rules”, has been a central feature of Generative Grammar from its earliest moments. It frequently surfaces in that literature in the guise of the “Structure Dependency” of grammatical rules. It is also the notion that is embodied in the “Proper Analysis” condition on transformations of Chomsky [22] (written in 1955). The most illuminating and ambitious endorsement of the principle is to be found in [22, p.210-211] (chapters written in 1956) where the following four criteria are offered as tests for grammatical constituency:

1. [Susceptibility to] the rule for conjunction
2. [Resistance to] intrusion of parenthetical expressions
3. Ability to enter into transformations
4. [Cooccurrence with] certain intonational features [associated with intonational phrases]

These criteria are very cautiously advanced, carefully surrounded with qualifications, and the subsequent discussion is deliberately designed to demonstrate that some of them at least raise as many questions as they answer. Nevertheless, there is an implicit claim of great boldness, however programmatically stated. If these operations are tests for constituency, it can only be because they are rules of grammar, subject to the constituent condition.

The bulk of *The Logical Structure of Linguistic Theory*, and most work in Generative Grammar since, mainly bears on the claim relative to the third criterion, concerning transformational rules of movement (and their modern equivalents and alternatives). By and large, such rules have been

scrupulously true to the constituent condition. It has proved much more difficult to make good on the implicit claim with respect to the remaining three phenomena. Theories of coordination, intonation, and (insofar as there are any) parentheticalisation have generally been forced at some point to compromise the constituent condition. The present work should be viewed as an attempt to vindicate the original claim.

Before proceeding to examine this claim in further detail in the next section, it is worth noting that two further consequences follow from the assumption of a rule-to-rule relation between syntax and semantics. First, the assumption also implies that the *only* linguistic entities that have interpretations are constituents. This consequence is again entirely uncontroversial, and virtually all theories of competence have adhered to it (insofar as they have involved an explicit semantics at all). However, it will be relevant to the discussion of processing below, for it implies that if we ever find ourselves wanting to argue that the processor has access to a structure that is not a grammatical constituent, such as the combination of a subject and a transitive verb *Harry found*, then we may have violated the strict version of strong competence.

Finally, and slightly more controversially, the rule-to-rule hypothesis, and its justification in terms of its parsimony with respect to the theory of language learning and evolution, imply that syntactic and semantic rules should have the property of “monotonicity”. That is, there can be no rules such as old-style transformations which convert structures which are ill-formed (at either level) into structures which are well-formed. The involvement of any intermediate, uninterpretable level of structure – such as an old-style surface structure – immediately raises the question of how that level of grammar can be learned, compromising the parsimony of the theory. The most straightforward way of interpreting this condition is as implying that the grammar should be “monostratal”, in the sense of having a single level of syntactic structure, directly associated with an interpretation.⁴

⁴To claim that syntax is monostratal is not of course to deny that theories of language need to recognise distinct modules such as phonology, morphology, syntax and semantics. It does on the other hand imply that those levels too should be monotonically related, a point to which we return below.

§3 FRAGMENTS

To what extent do the facts of natural language syntax conform to the predictions set out above? The generative theoretical tradition, as was noted earlier, has had considerable success in accounting for many constructions involving discontinuities between elements that are semantically dependent upon one another. Many such constructions were originally brought within the fold of the constituent condition on rules by the introduction of transformational rules of “movement” of constituents. Such constructions fall naturally into two groups. The first group includes phenomena which can be accounted for entirely in terms of “bounded” dependencies – roughly, dependencies between items which occur within the same clause, as in the following examples:⁵

- (2) a. John expects that Harry will take a bath
- b. John expects Harry to take a bath
- c. John expects to take a bath
- d. Harry is expected to take a bath by John

As Brame [11], [12], and Bresnan [16] were among the first to point out, the clause-bounded nature of these dependencies means that they can be base-generated, or (equivalently) specified in the lexicon, thus bringing them within the domain of the constituent condition without the use of movement as such. The present work will have little more to say about the bounded constructions.

The generative approach has also proved extremely successful in accounting for the phenomenon of unbounded dependency exhibited in relative clauses and topicalisations such as the following, again in terms of movement:

- (3) a. That book, *I expect I shall find*
- b. These articles, *I think that you must have read without understanding.*

In such constructions, elements that are related in the interpretation of the construction, such as the extracted NPs and the verb(s) of which they are

⁵It will be apparent from the examples that the definition of clause here needs further specification.

the objects, can be separated by arbitrarily long substrings and indefinitely much embedding. While the residue of topicalisation or relativisation clause at first glance looks like a non-constituent fragment of a sentence, it can be rather satisfactorily regarded as a constituent of type S, with a special kind of “invisible” or “moved” argument, and thereby be brought within the constituent condition. While it has also been proposed to bring these constructions too within the scope of base-generative and lexical rules (see Gazdar, [38], *passim*), the early theories were so productive of generalisations concerning these constructions that it will be convenient in what follows to continue to describe them in terms of movement, regardless of the fact that the present theory in fact eschews rules of movement entirely.

It has proved much more difficult to account for coordination, parentheticalisation, and phrasal intonation within the confines of the constituent condition. It is worth looking at some data in this connection.

¶3.1 COORDINATION AND PARENTHETICALS

At first glance, there is a striking overlap between the kinds of fragments that result from relativisation and the related topicalising construction, and those that can coordinate. In particular, practically anything that can occur as the residue of leftward movement can be coordinated, as in examples like the following:

- (4) a. A book which *I expect I shall find*,
and *I think that you must have read without really understanding*.
- b. *I expect I shall find*,
but *I think that you must have read without really understanding*,
my novel about the secret life of legumes.

The second example involves rightward movement (again, the term is used descriptively). It is also striking that there is a similarly overwhelming (though not quite so complete) conspiracy between the residues of leftward and rightward movement. That is, most residues that arise from leftward movement can also arise from rightward movement.⁶

However, the fragments that result from coordination are much more

⁶The obvious exception arises in subject extraction.

diverse than those that result from (leftward and rightward) movement. For example:

- (5) a. I *want to try to write*, and hope to see produced,
a movie about the secret life of vegetables.
b. Are you, or have you ever been, a member of the Friends of the Vegetable Film Society?
c. I gave a policeman a flower, and *a postman, a letter*.
d. I want to try to write a novel, and *you, a screenplay*

While considerably less attention has been devoted to parenthetical utterances (but cf. Emonds [31, section II.9], McCawley [60] and Levelt [55]), some similarly unconstrained fragments arise from their intrusion, as in:

- (6) *Are you*, I ask myself, a member of the Friends of the Vegetable Film Society?

The result has been that, while linguistic theories have had some success in accounting for the relative clause construction in terms of devices which reinstate the constituent condition by deriving such fragments from traditional constituents such as *S* via devices like movement [20], indexed “traces” [23], and feature-passing, [38], they have been much less successful in showing that the same devices will account for coordination. Instead, coordination has led to the introduction of rules of deletion to supplement rules of movement. Such rules again attempt to reinstate the constituent condition to the grammar, by deriving the fragments underlying *S*. However, such rules have frequently been forced to compromise the constituent condition. For example, 5c appears to require either the movement or the deletion of a non-constituent, and 5d appears to offer no alternative to the deletion of the non-constituent *want to try to write*. More worrying still, this fragment looks suspiciously like the kind of fragment that is the surface-structural *result* of deletion or movement, as in 5a. This problem has turned out to be extremely resistant to solution.

This result is suprising, because intuitively all of these constructions appear to be related to the semantic notion of *abstraction*, or definition of a *property*. Most obviously, a restrictive relative clause like (a) below seems to correspond to a predicate or property of *being preferred by Mary*. Formally such properties, concepts or abstractions can be conveniently and transparently represented by terms in the λ -calculus like (b):

- (7) a. ... (which) Mary prefers
 b. λx [*prefer'* x *Mary'*]

(For those who are unfamiliar with this notation, the operator λ declares the symbol x to be a variable local to the expression that follows. The expression can therefore be thought of as denoting the property of being “a thing such that Mary prefers it”.) The variable is thus in every way comparable to a parameter or formal variable of a subroutine or function in a computer programming language, and the operator λ can be thought of as defining a function in such a language, mapping entities onto truth values according to whether Mary prefers them or not. (Here as elsewhere in the paper, constants like *prefer'* are used to identify semantic interpretations whose details are not of immediate interest, and a convention of “left associativity” of function application is observed, so that the above formula is equivalent to λx [(*prefer'* x) *Mary'*]). Most current theories of natural language grammar since “standard” transformational grammar (Chomsky [20]) more or less explicitly embody the analogy between relativisation and abstraction over a variable. Thus, the Government-Binding theory explicitly identifies traces as bound variables, while GPSG and its descendants use explicit variable binding via the lambda calculus in the interpretation of the slash categories that mediate long range dependencies (cf. Pollard and Sag [72]). However coordination is much freer than relativisation in the fragments that it generates, and has been resistant to satisfactory explanation in such terms.

¶3.2 INTONATION AND DISCOURSE CONTEXT

Similar fragments abound in spoken language, arising from phenomena associated with prosody and intonation, as well as less well-behaved phenomena like restarts, and the parentheticals discussed earlier. For example, one quite normal prosody for an answer to the following question (a) involving stress on the word *Mary* and a rise in pitch at the end of the word *prefers*, intuitively imposes the intonational structure indicated in (b) by the brackets (stress, marked in this case by raised pitch, is indicated by capitals):⁷

⁷We shall not rely on mere intuition to define this phenomenon in the developments below. But the effect is very strong. It is ironic that one of the first difficulties in teaching introductory syntax is to persuade the students that this is *not* the notion of structure that is relevant to the study of grammar. Of course, in an important sense, they are right.

- (8) a. I know that Alice prefers VELVET. But what does MARY prefer?
 b. (MARY prefers) (CORDUROY).

Such a grouping is orthogonal to the traditional syntactic structure of the sentence.

Intonational structure nevertheless remains strongly constrained by meaning. For example, contours imposing bracketings like the following do not seem to be allowed, as Selkirk [78] has pointed out:

- (9) #(Three cats)(in ten prefer corduroy)

Halliday [46] observed that this constraint, which Selkirk has called the “Sense Unit Condition”, seems to follow from the *function* of phrasal intonation, which is to convey what will here be called “information structure” – that is, distinctions of focus, presupposition, and propositional attitude towards entities in the discourse model. These discourse entities are more diverse than mere noun-phrase or propositional referents, but they do not seem to include such non-concepts as “in ten prefer corduroy”.

Among the categories that they *do* include are what Wilson and Sperber [95] and Ellen Prince [73] have termed “open propositions”. One way of introducing an open proposition into the discourse context is by asking a Wh-question. For example, the question in 8, *What does Mary prefer?* introduces an open proposition, or topic of conversation, corresponding once again to the concept of *a thing such that Mary prefers it*. As Jackendoff [50] pointed out, it is once again natural to think of this open proposition as a functional *abstraction*, and to express it in the notation of the λ -calculus, as in 7, repeated here:⁸

- (10) λx [*prefer'* x *mary'*]

When this function or concept is supplied with an argument *corduroy'*, it *reduces* to give a proposition, with the same function argument relations as the canonical sentence:⁹

⁸The proposal is also implicit in the “structured propositions” approach of Cresswell [28] and von Stechow [80].

⁹Again, function application associates to the left.

(11) *prefer' corduroy' mary'*

It is the presence of the above open proposition rather than some other that makes the intonation contour in 8b felicitous. (That is not to say that its presence uniquely *determines* this response, nor that its explicit mention is necessary for interpreting the response.)

These observations have led linguists such as Selkirk to postulate a level of “intonational structure”, independent of syntactic structure and related to information structure. The involvement of two apparently uncoupled levels of structure in natural language grammar appears to complicate the path from speech to interpretation unreasonably.

It is therefore interesting to observe that the constructions considered in the previous section, whose semantics also seems to be reminiscent of functional abstraction, are also subject to something like a “sense unit condition”. For example, strings like “in ten prefer corduroy” seem to be as reluctant to take part in coordination as they are to be treated as intonational phrases, and, to the extent that they are possible at all, demand strikingly similar contexts of utterance:

(12) #Three cats in twenty like velvet, and in ten prefer corduroy.

Since we have already noted that coordinate constructions constitute another major source of complexity for current theories of natural language grammar, and also offer serious obstacles to computational applications, it is tempting to suspect that this conspiracy between syntax and prosody might point to a unified notion of structure that is somewhat different from traditional surface constituency.

It is slightly surprising that all of these constructions whose semantics is so directly related to the notion of functional abstraction give rise to these strikingly similar kinds of fragment, in which there is no explicit linguistic manifestation of the bound variable itself. It seems odd that an element so central to the semantics of abstraction should be left implicit. (Bound variables are otherwise very directly realised in the syntax, for example in association with quantifiers.)

It is interesting in this connection that there are systems other than the λ -calculus which capture the notion of abstraction, but which eschew the

use of bound variables entirely. These are the combinatory systems invented by Schönfinkel and Curry as a formal foundation for the semantics of the λ -calculus. In such systems (which will be discussed in detail below), terms equivalent to abstractions like the above are built up using a handful of elementary operations for combining functions, such as functional composition. Systems using quite small numbers of combinators can be shown to be equivalent in expressive power to the λ -calculus.

The existence of these systems raises the possibility that alternative theories of grammar can be developed based as directly upon the combinatory applicative systems as the traditional ones are upon the λ -calculus. The body of the present paper presents a theory of grammar in which syntactic rules corresponding to such combinators are used to lend the status of constituents to such substrings *want to try to write* and even *a policeman a flower* in the above examples, without the use of movement or deletion. Such grammars will be seen to provide a unified treatment of a wide variety of syntactic phenomena in natural language, and to explain phenomena of long distance dependency (including relativisation), coordination, and intonation, within the confines of the constituent condition on rules, and in terms of a single principle. That principle is that the function-argument relations that hold in sentences of natural languages are projected from the relations defined in the lexicon by syntactic operations corresponding to combinators, rather than by operations explicitly or implicitly involving bound variables. The paper then goes on to show that the same notion of constituency can be used to explain spoken intonation.

The theory of grammar that results will be characterised by a profusion of alternative surface analyses for any given sense of a moderately complex sentence. Part II of the paper discusses the nature of the human sentence processing mechanism, and argues that under the “strict” version of the strong competence hypothesis, these grammars are compatible with the simplest possible parser, and that they therefore have a further explanatory edge over alternative theories.

§4 COMBINATORY GRAMMARS.

Combinatory Categorical Grammar (CCG, [1], [83]) is an extension of Categorical Grammar (CG, [2], [6]). That is to say that elements like verbs are

associated with a syntactic “category” which identifies them as *functions*, and specifies the type and directionality of their argument(s) and the type of their result. The present paper uses a notation in which the argument or domain category always appears to the right of the slash, and the result or range category to the left. A forward slash / means that the argument in question must appear on the right, while a backward slash \ means that the argument must appear on the left.¹⁰

$$(13) \text{ prefers} := (S \backslash NP) / NP: \textit{prefer}'$$

The category $(S \backslash NP) / NP$ could be regarded as both a syntactic and a semantic object, as in the unification-based categorial grammars of Uszkoreit [91], Wittenburg [97], Zeevat et al. [99] and Pareschi and Steedman [64]. (See [85] for an explicit expression of the present grammars in unification-based terms, uniting syntax and semantics in a single category.) However, it will be convenient for present purposes to separate the semantic and syntactic types in the notation. In the present paper, an expression identifying the translation of a category appears to its right, separated by a colon. It is of course the translation which determines the grammatical or functional role of the first argument to be that of the object, and the second to be the subject.

Such functions can combine with arguments of the appropriate type and position by rules of functional application, written as follows:

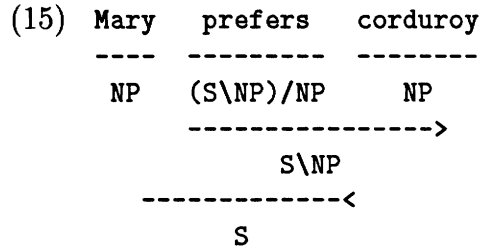
(14) *The Functional Application Rules:*

$$\begin{array}{ll} a. & X/Y : F \quad Y : y \quad => \quad X : Fy \quad (>) \\ b. & Y : y \quad X \backslash Y : F \quad => \quad X : Fy \quad (<) \end{array}$$

Such rules are both syntactically and semantically rules of functional application, as is indicated by the interpretations that appear to the right of the colon for each category.¹¹ They allow derivations like the following:

¹⁰The reader is warned that other superficially similar notations are used by some of the other authors referred to here.

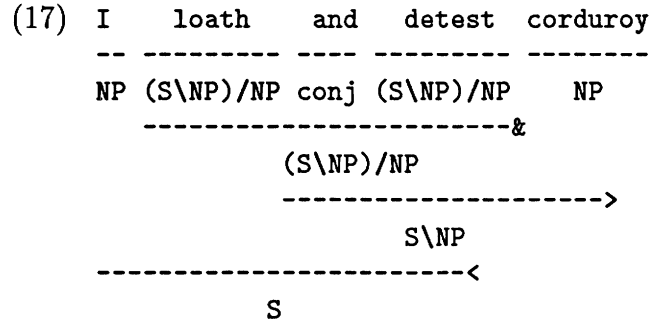
¹¹Again, it is helpful for the present expository purposes to explicitly identify the semantics, but the semantic annotations are, strictly speaking, redundant, since the categories themselves can be regarded as both syntactic and semantic objects, as they are for example in [85].



The syntactic functional types are identical to the semantic types of their translations, apart from directionality. This derivation therefore also builds a compositional interpretation, which we will write *prefer' corduroy' mary'*, using a convention of “left associativity” of functional application. Of course such a “pure” categorial grammar is context-free.

Coordination might be included in CG via the following rule, allowing any constituents of like type, including functions, to form a single constituent of the same type, and thereby to take part in derivations exactly analogous to the above:¹²

(16) $X \text{ conj } X \Rightarrow X$



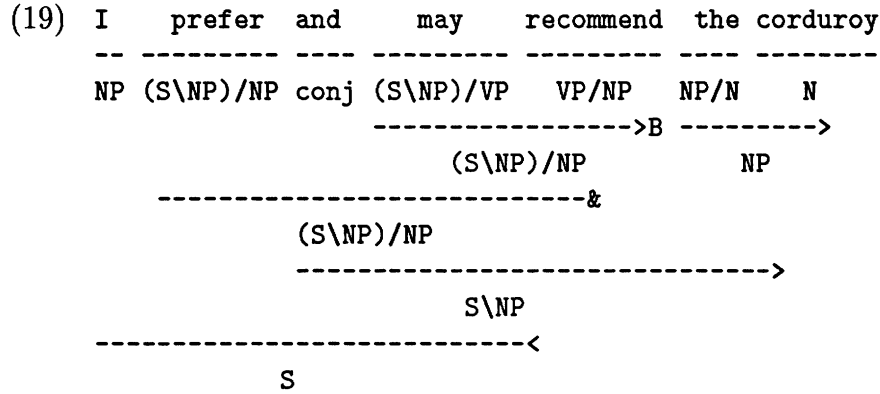
In order to allow coordination of contiguous strings that do not constitute constituents, CCG generalises the grammar to allow certain operations on functions related to Curry’s combinators [29], [79]. For example, functions may *compose*, as well as *apply*, under the following rule

¹²Such a rule is in fact a simplification, and a more linguistically adequate account is presented elsewhere (cf. [85]), capturing the fact that conjunctions in English are proclitic, and associate to the rightmost conjunct.

(18) *Forward Composition:*

$$X/Y : F \quad Y/Z : G \Rightarrow X/Z : \lambda x [F(Gx)]$$

The most important single property of combinatory rules like this is that they have an invariant semantics. This one composes the interpretations of the functions that it applies to, as is apparent from the right hand side of the rule.¹³ Thus sentences like *I prefer, and may recommend, corduroy* can be accepted, via the following composition of two verbs (indexed as >B, following Curry's use of the identifier **B** for the composition combinator) to yield a composite of the same category as a transitive verb. Crucially, composition also yields the appropriate interpretation, assuming that a semantics is also provided for the coordination rule.



It should be noted that this rule will potentially allow certain non-conjoinable sequences to compose and coordinate. For example, under the assumption that determiners are *NP/N*, we could derive the following, by composing a transitive verb with a determiner:

¹³Once again, the explicit identification of the semantics of this and all subsequent rules, using the notation of the λ -calculus, is only used for expository clarity. The categories themselves are complete syntactic and semantic entities, and it is functional composition itself, embodied in the rule relating the categories themselves, that is the primitive of the theory, not the λ operator.

(20) *I must cook a, and eat the, potato.

 S/VP VP/NP NP/N conj VP/NP NP/N N
 ----->B ----->B
 VP/N VP/N
 -----<&>

Such examples could be excluded by imposing a condition on the forward composition rule, forbidding the variable Y to be instantiated as NP.¹⁴ However, the increased acceptability of related examples like the following suggests that the problem is not purely syntactic:

- (21) a. I will cook three, and eat two, potatoes.
 b. I want cooked, and he wants uncooked, potatoes.
 c. I will paint a picture of, and write a novel about, the potato.

It rather seems to arise from the good sense or otherwise of the concept that arises in the interpretation of the composition across the NP boundary. In some of the derivations that follow, such a constraint will tacitly be assumed, without any particular commitment to where precisely in the grammar it arises.¹⁵

Combinatory grammars also include type-raising rules, which turn arguments into functions over functions-over-such-arguments. These rules allow arguments to compose, and thereby take part in coordinations like *I dislike*, and *Mary prefers*, *the corduroy*. They too have an invariant compositional semantics corresponding to the combinator known to (some) combinatory logi-

¹⁴Such a condition would follow automatically and without stipulation if the category NP in determiners and other nominal categories were replaced throughout the lexicon by type-raised categories, as envisaged in the discussion of type raising below.

¹⁵To the extent that this constraint does indeed arise from the conceptual anomaly of the interpretation associated with compositions like $\#[eat\ the]_{VP/N}$ in comparison with ones like $[eat\ two]_{VP/N}$, a theory like the present one which will actually provide such an interpretation via the composition rule may be more helpful than one which does not. Note that if the NP constraint is regarded as arising from lexical type raising, then exceptions like the above must be regarded as arising from further raising the VERB over the raised NP category.

cians as \mathbf{T} , which ensures that the result has an appropriate interpretation.¹⁶ For example, the following rule, indexed $>\mathbf{T}$, allows the conjuncts to form as below (again, the remainder of the derivation is omitted):

(22) *Subject Type-raising:*

$$NP : x \Rightarrow S/(S \backslash NP) : \lambda P [Px]$$

(23)

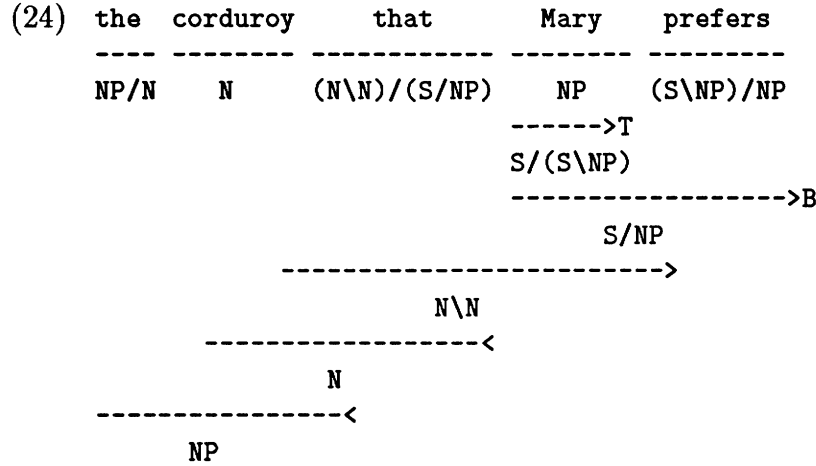
I	dislike	and	Mary	prefers	...
\mathbf{NP}	$(S \backslash \mathbf{NP})/\mathbf{NP}$	\mathbf{conj}	\mathbf{NP}	$(S \backslash \mathbf{NP})/\mathbf{NP}$	
$\text{-----}>\mathbf{T}$			$\text{-----}>\mathbf{T}$		
$\mathbf{S}/(S \backslash \mathbf{NP})$			$\mathbf{S}/(S \backslash \mathbf{NP})$		
$\text{-----}>\mathbf{B}$			$\text{-----}>\mathbf{B}$		
\mathbf{S}/\mathbf{NP}			\mathbf{S}/\mathbf{NP}		
$\text{-----}\&$					
\mathbf{S}/\mathbf{NP}					

There is an option here as to whether we consider type-raising to be an operation of active syntax, like composition, or as a rule of the lexicon or of morphology. In the latter case, of course, not only nominative NPs like *I*, but also uncased NPs like *Mary*, and even articles like *the* would have to bear additional categories like $S/(S \backslash NP)$, $(S/(S \backslash NP))/N$, and so on. As far as the parser is concerned, it is likely that this alternative is desirable.¹⁷ For present purposes, however, it makes no difference, and it is notationally less wearisome to think of type-raising as a syntactic rule, merely noting that the alternative exists.

It will probably be obvious at this point that the addition of type raising and composition to the theory of grammar immediately provides everything we need in order to account for leftward extractions in relative clauses, on the further assumption that relative pronouns bear a lexical category $(N \backslash N)/(S/NP)$ – a function from fragments like *Mary prefers* into noun modifiers which is itself closely related to a type-raised category:

¹⁶This combinator is also known as \mathbf{C}_* . Again, the explicit semantics is purely mnemonic.

¹⁷It is also possible that the grammar itself may require this interpretation of type-raising – cf. [62].



It should be similarly obvious that the theory immediately predicts that leftward and rightward extraction will be unbounded:¹⁸

- (25) a. I think that Mary prefers, and I know that you dislike, corduroy
b. The corduroy which I think that Mary prefers

The subject type-raising rule is a special case of a more general rule which can be written as follows:

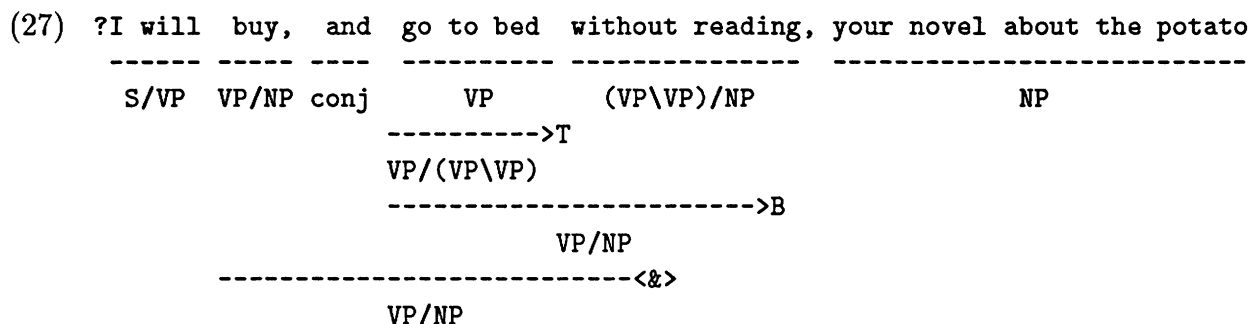
- (26) *Forward Type Raising:*
- $$X : x \Rightarrow T / (T \backslash X) : \lambda P [Py]$$

The symbol T is a (polymorphic) variable standing for any category that the grammar permits.¹⁹

As in the case of the forward composition rule 18, such a free type-raising rule threatens to overgeneralise. For example, it potentially permits VPs to raise over adjunct categories, to allow derivations like the following:

¹⁸See the earlier papers and [89] for details, including remarks concerning ECP and Coordinate Structure constraints, and on pied piping.

¹⁹The generalised form of composition used in the other papers referred to here (see later footnote) allows grammars to include unboundedly many categories. However, as Rooth and Partee [77], [65] pointed out within a different framework, there are some obvious ways of restricting type raising which eliminate this problem. The most obvious one is to adopt the tactic mentioned at a number of points above, of type raising nominal categories over a fixed set of categories in the lexicon, rather than in syntax.



Such examples are generally held to be unacceptable, and to violate the island status of adjuncts. They are certainly less acceptable than similar examples involving “parasitic” extraction, discussed in [83]:

(28) I will buy, and burn without reading, your novel about the potato.

However, the marginal acceptability of 27, and of related examples discussed by Chomsky [24, p. 72], as well as a number of other subjacency-related constraints, such as Ross’s Complex NP constraint, suggests that what is wrong with 27 is again a matter of the good sense of the predicate *go to bed without reading*. In the examples that follow, appeal will be made to a parallel with similar constraints upon intonational constituents, without any commitment as to the source of such constraints.

At this point, it is natural to ask what further rules are permitted by the theory, and whether the degrees of freedom that they imply are required elsewhere in the grammar of English and other languages. This question has been extensively discussed in [83], [85], where it is argued that the possible rules are limited by three Principles called *Adjacency*, *Directional Consistency*, and *Directional Inheritance*. The first of these simply amounts to the assumption that purely local combinatory rules, as opposed to long-range rules of movement, abstraction, or indexing over variables, will in fact do the job. The second principle, that of Directional Consistency, prohibits combinatory rules which contradict the directional specifications of the lexicon, such as the following version of functional application:

(29) * $X \backslash Y : F \quad Y : y \Rightarrow X : Fy$

The proof of this claim lies in showing that the rules of combinatory grammar can be made sensitive to intonation contour, which limit their application in spoken discourse. We must also show that the major constituents of intonated utterances like 8b, under the analyses that are permitted by any given intonation, correspond to the information structure of the context to which the intonation is appropriate, as in (a) in the example 8 in section 3.2. This demonstration will be quite simple, once we have established a notation for intonation contours.

¶5.1 INTONATION AND CONTEXT

I shall use a notation which is based on the theory of Pierrehumbert [68], as modified in more recent work by Selkirk [78], Beckman and Pierrehumbert [7], [69], and Pierrehumbert and Hirschberg [70]. I have tried as far as possible to take my examples and the associated intonational annotations from those authors. The theory proposed below is in principle compatible with any of the standard descriptive accounts of phrasal intonation. However, a crucial feature of Pierrehumbert’s theory for present purposes is that it distinguishes two subcomponents of the prosodic phrase, the *pitch accent* and the *boundary*.²⁰ The first of these tones or tone-sequences coincides with the perceived major stress or stresses of the prosodic phrase, while the second marks the righthand boundary of the phrase. These two components are essentially invariant, and all other parts of the intonational tune are interpolated. Pierrehumbert’s theory thus captures in a very natural way the intuition that the same tune can be spread over longer or shorter strings, in order to mark the corresponding constituents for the particular distinction of focus and propositional attitude that the melody denotes. It will help the exposition to augment Pierrehumbert’s notation with explicit prosodic phrase boundaries, using brackets. These do not change her theory in any way: all the information is implicit in the original notation.

Consider for example the prosody of the sentence *Mary prefers corduroy* in the following pair of discourse settings, which are adapted from Jackendoff [50, pp. 260]:

²⁰I am suppressing the distinction between the intonational phrase proper, and what Pierrehumbert and her colleagues call the “intermediate” phrase, which differ in respect of boundary tone-sequences.

(32) Q: Well, what about the CORduroy? Who prefers THAT?

A: (MARY) (prefers CORduroy).
 H* L L+H* LH%

(33) Q: Well, what about MARY? What does SHE prefer?

A: (MARY prefers) (CORduroy).
 L+H* LH% H* LL%

In these contexts, the main stressed syllables on both *Mary* and *corduroy* receive a pitch accent, but a different one. In the former example, 32, there is a prosodic phrase on *Mary* made up of the pitch accent which Pierrehumbert calls H*, immediately followed by an L boundary. There is another prosodic phrase having the pitch accent called L+H* on *corduroy*, preceded by null or interpolated tone on the words *prefers*, and immediately followed by a boundary which is written LH%. (I base these annotations on Pierrehumbert and Hirschberg's [70, ex. 33] discussion of a similar example.)²¹ In the second example 33 above, the two tunes are reversed: this time the tune with pitch accent L+H* and boundary LH% is spread across a prosodic phrase *Mary prefers*, while the other tune with pitch accent H* and boundary LL% is carried by the prosodic phrase *corduroy* (again starting with an interpolated or null tone).²²

The meaning that these tunes convey in these contexts is intuitively very obvious.²³ As Pierrehumbert and Hirschberg point out, the latter tune seems to be used to mark some or all of that part of the sentence expressing information that the speaker believes to be *novel to the hearer*. In traditional terms, it marks the "comment" – more precisely, what Halliday called the "rheme". In contrast, the L+H* LH% tune seems to be used to mark some or all of that part of the sentence which expresses information which in traditional terms is the "topic" – in Halliday's terms, the "theme". For present

²¹We continue to gloss over Pierrehumbert's distinction between "intermediate" and "intonational" phrases.

²²The reason for notating the latter boundary as LL%, rather than L is again to do with the distinction between intonational and intermediate phrases.

²³I do not of course intend to claim that these are the *only* meanings that these tunes can convey.

purposes, a theme can be thought of as conveying *what the speaker assumes to be the subject of mutual interest*, and this particular tune marks a theme as *novel to the conversation as a whole*, and as standing in a contrastive relation to the previous theme. (If the theme is not novel in this sense, it receives *no* tone in Pierrehumbert's terms, and may even be left out altogether.)²⁴ Thus in 33, the L+H* LH% phrase including this accent is spread across the phrase *Mary prefers*.²⁵ Similarly, in 32, the same tune is confined to the object of the open proposition *prefers corduroy*, because the intonation of the original question indicates that preferring corduroy *as opposed to some other stuff* is the new topic or theme.

It follows that the position of the pitch accent in the phrase has to do with a further orthogonal dimension of information structure *within both theme and rheme*, corresponding to *the interesting bit* of either information unit. It is tempting to call this part the "focus" of the theme or rheme, but I shall avoid the term and follow Halliday in calling it "new" information, in contrast to the "given" information accompanied by the null tone. The term "new" is perhaps not entirely satisfactory, but it is intended to suggest that the part of the theme that is marked in this way is *new to the speaker*, while the part of the rheme that is so marked is believed by the speaker to be *new to the hearer*. This usage is illustrated in the following example:

(34) Q: I know that Mary's FIRST degree is in PHYSICS.

But what is the subject of her DOCTORATE?

A: (Mary's DOCTORATE)		(is in CHEMISTRY)	
L+H* LH%		H* LL%	
Given	New	Given	New
Theme		Rheme	

Here the theme is *Mary's doctorate*, where the head noun is emphasised because it stands in contrast to another of her qualifications. The rheme is that it *is in chemistry*, where chemistry is emphasised in contrast to another subject.

²⁴Here I depart slightly from Halliday's definition. The present proposal also follows Lyons [56] in rejecting Halliday's claim that the theme must necessarily be sentence-initial.

²⁵An alternative prosody, in which the contrastive tune is confined to *Mary*, seems equally coherent, and may be the one intended by Jackendoff. I believe that this alternative is informationally distinct, and arises from an ambiguity as to whether the topic of this discourse is *Mary* or *What Mary prefers*. It too is accepted by the rules below.

¶5.2 COMBINATORY PROSODY

The L+H* LH% intonational melody in example 33 belongs to a phrase *Mary prefers ...* which corresponds under the combinatory theory of grammar to a grammatical constituent, complete with a translation equivalent to the open proposition $\lambda x[(\textit{prefer}' x) \textit{mary}']$ introduced by the question in that example *What does Mary prefer?* The combinatory theory thus offers a way to derive such intonational phrases, using only the independently motivated rules of combinatory grammar, entirely under the control of appropriate intonation contours like L+H* LH%.

In fact we might think of doing this in one of two ways. We might either think of annotating combinatory rules, for example imposing a restriction forbidding composition across an intonational phrase boundary. Or we might exploit the advantages of categorial grammar more directly, as follows.²⁶

¶5.3 CONSTITUENCY AND INTONATION.

We will begin by defining the two pitch accents as functions of the following types:²⁷

$$\begin{aligned}
 (35) \quad L+H^* &:= \textit{Theme}/Bh \\
 H^* &:= (\textit{Utterance}/\textit{Theme})/bl \\
 H^* &:= (\textit{Utterance} \backslash \textit{Theme})/Bl
 \end{aligned}$$

These categories define the two pitch accents as functions over boundary tones into the two major informational types, *Theme* and *Rheme*, where the latter category is itself a function from themes into utterances.²⁸ In the case of the H* pitch accent, there are two categories, one combining with an intermediate phrase boundary to yield a *forward* category of the latter kind, the other combining with an intonational phrase boundary to yield a

²⁶This section is a summary of the proposals in [86] and [87].

²⁷The following rules constitute a minor revision of [86], prompted by Steven Bird's [10] observation of an overgeneralisation arising from the conflation of intonational and intermediate phrases in the earlier account.

²⁸The choice of the rheme rather than the theme as the "head" of the prosodic utterance is to some extent arbitrary, but is motivated by the observation that the rheme is the obligatory member of the pair.

backward version. It will be convenient to refer to the backward and forward category collectively as “the rheme category”.

We define the various boundary tones as arguments to these functions, as follows:²⁹

$$\begin{aligned}
 (36) \quad LH\% &:= Bh \\
 LL\% &:= Bl \\
 L &:= bl
 \end{aligned}$$

Finally, we accomplish the effect of interpolation of other parts of the tune by assigning the following category to all elements bearing no tone specification, which we will represent as the null tone \emptyset :³⁰

$$(37) \quad \emptyset := X/X$$

X is a category that can match any category, importantly including X/X .³¹ It will therefore introduce a considerable amount of non-determinism to the prosodic side of the grammar. However, this will turn out to be strictly necessary: the null tone is very ambiguous in present terms.

The prosodic combinatory rules include forward and backward functional application. They also include the following very restricted version of forward functional composition:

$$\begin{aligned}
 (38) \quad &\textit{Forward Prosodic Functional Composition:} \\
 &X/Y \ Y/Z \Rightarrow X/Z \\
 &\text{where } Y \in \{Bh, Bl, bl\}
 \end{aligned}$$

²⁹An alternative grammar, which would be closer to Pierrehumbert & Beckman 1989, and which might also be more directly compatible with Pierrehumbert and Hirschberg’s proposals for the compositional assembly of discourse meanings from more primitive elements of meaning carried by each individual tone, would be obtained by assigning pitch accents the category of functions from *phrasal* tones into intermediate phrases marked as theme, rheme, etc, and assigning the boundary tones the category of functions from intermediate to intonational phrases, similarly marked. However, the precise details of such an alternative depend on some imponderables in the original theory concerning the precise position and nature of the phrasal tone itself (see Pierrehumbert & Beckman 1989:236-237 and an earlier footnote).

³⁰I am grateful to Steven Bird for first suggesting a related category for the null tone.

³¹In a unification-based realisation such as that sketched in Steedman 1990a, X would be realised as the equivalent of a unique Prolog variable, distinct from that in any other instance of the null tone.

The restriction is required because the whole point of the prosodic categories is to PREVENT composition across the theme/rheme boundary. It will become apparent below that, given the categories chosen above, the only occasion on which composition is required is when X/Y is a pitch accent – that is, a function over a boundary. Another important consequence is that two null tones of type X/X can only combine by application, not composition.

Syntactic combination can now be made subject to intonation contour by the following simple restriction:

- (39) THE PROSODIC CONSTITUENT CONDITION: Combination of two syntactic categories via a syntactic combinatory rule is only allowed if their prosodic categories can also combine (and *vice versa*).

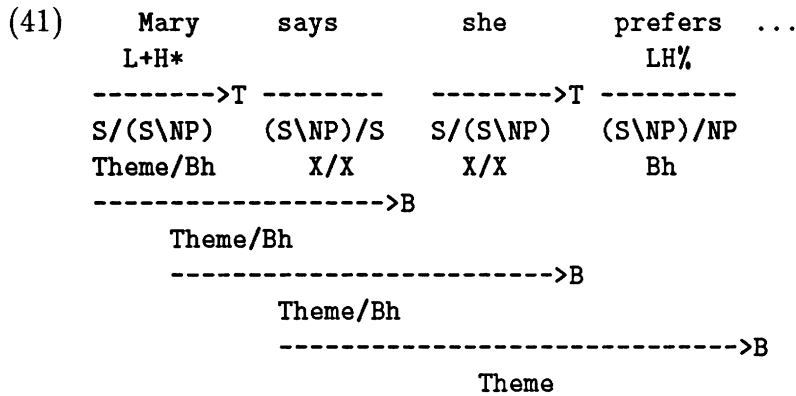
(The prosodic and syntactic combinatory rules need not, and usually will not, be the same).

This principle has the sole effect of excluding certain derivations for spoken utterances that would be allowed for the equivalent written sentences. For example, consider the derivations that it permits for example 33 above. The rule of forward composition is allowed to apply to the words *Mary* and *ate*, because the prosodic categories can combine (by functional application):

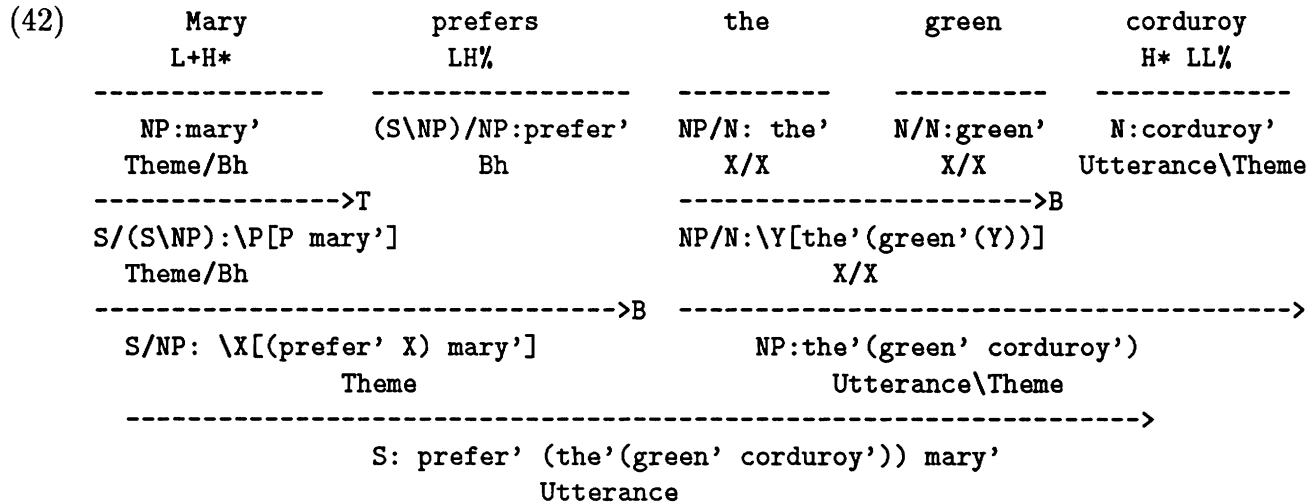
- (40)
- | | | |
|--------------------------------------|-------------------------|-----|
| Mary
L+H* | prefers
LH% | ... |
| ----- | | |
| NP:mary'
Theme/Bh | (S\NP)/NP:prefer'
Bh | |
| ----->T | | |
| S/(S\NP): \P[P mary']
Theme/Bh | | |
| ----->B | | |
| S/NP: \X[(prefer' X) mary']
Theme | | |

The category X/X of the null tone allows intonational phrasal tunes like L+H* LH% tune to spread across any sequence that forms a grammatical constituent according to the combinatory grammar. For example, if the same question *What does Mary prefer?* evokes the response *MARY says she prefers CORduroy*, then the tune will typically be spread over *Mary says she prefers*

... as in the following (incomplete) derivation, in which much of the syntactic and semantic detail has been omitted in the interests of brevity:



Such non-standard constituents, marked as Theme, can take part in such derivations as the following:



The division of the utterance into an open proposition constituting the theme and an argument constituting the rheme is appropriate to the context established in 33. Moreover, the theory permits no *other* division into a theme and a rheme for this intonation contour.

All the other possibilities for combining these two contours in a simple sentence are shown elsewhere [86] to yield similarly unique and contextually appropriate interpretations.

Sentences like the above, including marked theme and rheme expressed as two distinct intonational/intermediate phrases are by that token unambiguous as to their information structure. However, sentences like the following, which in Pierrehumbert's terms bear a single intonational phrase, are much more ambiguous as to the division that they convey between theme and rheme:

- (43) (I read a book about CORduroy)
H* LL%

Such a sentence is notoriously ambiguous as to the open proposition it presupposes, for it seems equally appropriate as a response to any of the following questions:

- (44) a. What did you read a book about?
b. What did you read?
c. What did you do?

Such questions could in suitably contrastive contexts give rise to themes marked by the L+H* LH% tune, bracketing the sentence as follows:

- (45) a. (I read a book about)(CORduroy)
b. (I read)(a book about CORduroy)
c. (I)(read a book about CORduroy)

It seems that we shall miss a generalisation concerning the relation of intonation to discourse information unless we extend Pierrehumbert's theory very slightly, to allow prosodic constituents resembling *null* intermediate phrases, without pitch accents, expressing unmarked themes. Since the boundaries of such intermediate phrases are not explicitly marked, we shall immediately allow all of the above analyses for 43. Such a modification to the theory can be introduced by the following rule, which nondeterministically allows constituents bearing the null tone to become a theme:

- (46) *Null Theme Promotion Rule:*
 $X/X \Rightarrow Theme$

The rule is nondeterministic, so it correctly continues to allow a further analysis of the entire sentence as a single Intonational Phrase conveying the

Rheme. Such an utterance is the appropriate response to yet another open-proposition establishing question, *What happened?*

The grammar exemplified in the derivation 42, above, implicitly embodies an explanation of the second dimension of information structure identified above, Halliday's contrast between "given" and "new" information. The derivation that is given there for the rheme *the green corduroy* is only one of two alternatives permitted by that intonation contour, represented by the following bracketings:

- (47) a. (the green)(CORduroy)
b. (the) (green CORduroy)

However, this ambiguity, like the ambiguity inherent in the null theme, is also a genuine one. The intonation contour in question does not distinguish between the information structures appropriate to responses to the following two questions:

- (48) a. Does Mary like the green corduroy or the green velvet?
b. Does Mary like the green corduroy or the red velvet?

That is, it does not distinguish whether *green* is part of the given or part of the new. This is a subtle ambiguity, but its reality is clear from the existence of the intonation contour exhibited in the following exchange, for which the grammar only permits the second of these derivations, a fact that is consistent with the intuition that here it is the entire N-phrase *green corduroy* that is marked as being new and interesting:

- (49) Q. Does Mary like the green corduroy or the red corduroy?
A. (MARY prefers)(the (GREEN corduroy))

It follows from this analysis that the related utterance with an unmarked theme is extremely ambiguous as to its information structure, since there are several ways of splitting it up into a theme and a rheme, and there may be several ways of further dividing each of these into given and new:

- (50) (Mary prefers the green CORDUROY).
H* LL%

However, this ambiguity appears to be correctly constrained by the grammar. That is, the following does not appear to be a possible information structure, any more than it is a possible syntactic structure. (Relativisation and Coordination reveal that the sentence is in violation of an NP constraint.)

(51) *(Mary prefers the green)_{Theme}(CORDUROY)_{Rheme}

With the generalisation implicit in the above rules, we are therefore in a position to claim that the structures demanded by the theory of intonation and its relation to contextual information are the same as the surface syntactic structures permitted by the combinatory grammar. Moreover, because constructions like relativisation and coordination are more limited in the derivations they require, often forcing composition, rather than permitting it, it follows that anything which can coordinate can be an intonational constituent, and *vice versa*. It also follows that anything which can be the residue of relativisation can be an intonational constituent. These claims are discussed further in [86] and [87].

According to the theory sketched in this section, the pathway between phonological form and interpretation is much simpler than most recent phonological studies would suggest. Phonological form maps via the rules of combinatory grammar directly onto a surface structure, whose highest level constituents correspond to intonational constituents, annotated as to their discourse function. Surface structure therefore subsumes intonational structure. It also subsumes information structure, since the translations of those surface constituents correspond to the entities and open propositions which constitute the topic or theme (if any) and the comment or rheme. These in turn reduce via functional application to yield canonical function-argument structure, or “logical form”.

There may be significant advantages for theories of processing in such a theory of grammar. Most obviously, where in the past parsing and phonological processing have tended to deliver conflicting structural analyses, and have had to be pursued independently, they now are seen to be in concert. That is not to say that intonational cues remove all local structural ambiguity. Nor should the problem of recognising cues like boundary tones be underestimated, for the acoustic realisation in the fundamental frequency F_0 of the intonational tunes discussed above is entirely dependent upon the rest

of the phonology – that is, upon the phonemes and words that bear the tune. It therefore seems most unlikely that intonational contour can be identified in isolation from word recognition.³²

³²This is not necessarily a bad thing. The converse also applies: intonation contour effects the acoustic realisation of words, particularly with respect to timing. It is therefore likely that the benefits of combining intonational recognition and word recognition will be mutual.

PART II: PERFORMANCE

While the above account is necessarily brief, a number of properties of the combinatory theory of grammar should be clear at this point. In order to account for coordination, unbounded dependency, and intonational structure, strictly within the confines of the constituent condition on rules, we are forced to take a view of surface structure according to which strings like *Mary prefers* and *thinks that Mary prefers* are constituents in the fullest sense of the term. It follows that they must also be possible constituents of non-coordinate sentences like *Mary prefers corduroy*, and *Harry thinks that Mary prefers corduroy*. For moderately complex sentences there will in consequence be a large number of non-standard alternative derivations for any given reading.³³ The nature of the combinatory rules guarantees that all such derivations deliver the same function argument relations as more traditional derivations (which are of course also allowed by the combinatory grammar).

There is a temptation to reject this claim out of hand, on the grounds that it is at odds with much linguistic received opinion, and that if linguists know nothing else, they must know what a surface constituent is. However, the argument from authority does not stand up. On many tests for constituency, for example the list originally proposed by Chomsky and cited at the start of Part I, the combinatory theory does better than most. And other criteria for constituency, such as susceptibility to movement transformations, are confounded with *semantic* constituency, as Fodor *et al.* [33] have pointed out, an observation that merely underlines the wisdom of Chomsky's warning against relying too much on intuitions about meaning in the pursuit of syntactic form.

The temptation to reject the proposal on the basis of parsing considerations is similarly ill-founded. We have already noted that the presence of such semantic equivalence classes of derivations engenders rather more non-determinism in the grammar than we may have been aware of. While this fact implies that problem of writing parsers is a little harder than we might have expected, it is clear that this non-determinism really is a property of

³³The interested reader is referred to the earlier papers for further discussion, and to Weir and Joshi [94], Hepple and Morrill [45], and Vijay-Shankar and Weir [92] for results on parsing, automata-theoretic power and worst-case parsing complexity.

the language, and will be encountered by any theory with the same coverage with respect to coordination and intonational structure. We may also note that natural grammars show no sign whatsoever of a pressure to minimise non-determinism elsewhere in the grammar. There is therefore no reason whatsoever to doubt the competence theory on these grounds.

The only conclusion that we can draw from this grammatical non-determinism is that the mechanism for coping with it must be very powerful. The next part of the paper will argue that the most important device for dealing with non-determinism in the human processor is a process of eliminating partial analyses whose interpretation is inconsistent with the discourse context. The paper will also claim that combinatory grammars are particularly suited to the use of this tactic.

§6 ANATOMY OF A PROCESSOR

All language processors can be viewed as made up of three elements. The first is a grammar, which defines how constituents combine to yield other constituents. The second is a non-deterministic algorithm for applying the rules of the grammar to a string. The third is an oracle, or mechanism for resolving non-determinism.³⁴ The oracle decides *which* rule of grammar to apply at points in the analysis where the non-deterministic algorithm allows more than one rule to apply. The following sections discuss these elements in turn.

¶6.1 GRAMMAR

The strong competence hypothesis as originally stated by Bresnan and Kaplan assumes that the grammar that is used by or implicit in the human sentence processor is the competence grammar itself. It is important to be clear that this is an assumption, not a logical requirement. The processors that we design ourselves (such as compilers for programming languages) quite often do not exhibit this property. There is no logical necessity for the structures involved in processing a programming language to have anything to do with the structures that are implicated by its competence grammar –

³⁴The division of labour in processing between a non-deterministic algorithm and an oracle is not always made, particularly in implementations. However, all processors can be viewed in this way.

that is, the syntactic rules in the reference manual that are associated with its semantics. The compiler or interpreter can parse according to a quite different grammar, provided that there exists a computable homomorphism mapping the structures of this “covering grammar” onto the structures of the competence grammar. If the homomorphism is simple, so that the computational costs of parsing according to the covering grammar plus the costs of computing the mapping are less than the costs of parsing according to the competence grammar, then there may be a significant practical advantage in this strategem. For this reason, it is quite common for compilers and interpreters to parse according to a weakly equivalent covering grammar, mapping to the “real” grammar via a homomorphism under concatenation on a string representing the derivation under the covering grammar. For example, the strategem tends to be used in programming language compilers, when a parsing algorithm that is desirable for reasons of efficiency demands grammars in a normal form that is not adhered to by the grammar in the reference manual (cf. Gray and Harrison [41], and Nijholt [63]). Such a situation also arises in artificial parsers for natural languages, when it is desired to use a top-down or table-driven algorithm which is ill-suited to the left recursive rules which commonly occur in natural grammars (cf. Kuno [54] for an early example). As Berwick and Weinberg [8, esp. p.78-82] have noted, there is therefore no logical necessity for the structures involved in human natural language processing to have anything to do with the structures that are implicated by its competence grammar.

Nevertheless, similar considerations of parsimony in the theory of language evolution and language development to those invoked earlier might also lead us to expect that, as a matter of fact, a close relation will turn out to hold between the competence grammar and the structures dealt with by the psychological processor, and that it will in fact incorporate the competence grammar in a modular fashion. One reason that has been frequently invoked is that language development in children is extremely fast. This speed suggests that it proceeds via the piecemeal addition, substitution and modification of individual rules and categories of competence grammar. Any addition of, or change to, a rule of competence grammar will not in general correspond to a similarly modular change in a covering grammar. Instead, the entire ensemble of competence rules will typically have to be recompiled into a new covering grammar. Even if we assume that the transformation

of one grammar into another is determined by a language-independent algorithm, and can be computed each time at negligible cost, we have still sacrificed parsimony in the theory, and increased the burden of explanation on the theory of evolution. In particular, it is quite unclear why the development of either of the principal components of the theory in isolation should confer any selective advantage. The competence grammar is by assumption unprocessable, and the covering grammar is by assumption uninterpretable. It looks as though they can only evolve as a unified system, together with the translation process. The evolution of such a system is likely to be harder to explain than that of a more directly competence-based system. Indeed the first thing we would have to explain is why a covering grammar was necessary in the first place. The reference grammars of programming languages and the competence grammars of natural languages have syntaxes that are ill-suited to parsing with our favourite algorithms because they are constrained from outside by our own requirements. It is we who find Greibach Normal Form tedious, and find grammars with left recursive rules congenial, forcing the use of covering grammars by some artificial processors. It is quite unclear what comparable external force could have the effect of making natural grammars similarly ill-matched to the *natural* sentence processor.³⁵

It is important to note that the strong competence hypothesis as stated

³⁵That processor might conceivably require grammars to be in some normal form. However, provided that the normal form is a class of grammars of the same automata-theoretic power that the semantics of the language requires (and therefore of the same power as the competence grammar), we would expect that normal form to simply be a characteristic of the grammars we observe. In other words, we would view it as a (processing-based) constraint on the form of the competence grammars that actually exist. If on the other hand we are to entertain the possibility that the requirement for a covering grammar might arise from the fact that the mechanisms that have access to the outside world are for some reason of a lesser automata-theoretic power than the competence grammar, then the evolutionary claims become even more far-fetched. It seems inevitable that the mapping between analyses under the two grammars must become more complex. The problem of incremental language learning becomes correspondingly more complex. So of course does the problem mentioned above, of evolving the two systems in lock step. Indeed, we have to ask ourselves how these two systems which by assumption have completely different automata-theoretic character could begin to talk to one another in the first place. We have to ask ourselves whether it would not be simpler for evolution to bring the processor more in line with the requirements of competence grammar – after all, it has already come up with such a mechanism once, in the form of the interpreter for the competence semantics.

by Bresnan and Kaplan imposes no further constraint on the processor. In particular it does not limit the structures built by the processor to fully instantiated constituents. However, the present paper proposes a “strict” version of the competence hypothesis, which imposes this further condition.³⁶ The reasoning behind this strict version is again evolutionary. If in order to process sentences we need more than the grammar itself, even a perfectly general “compiler” that turns grammars into algorithms dealing in other structures, then the load on evolution is increased. Similar arguments for the need for the grammar and processor to evolve in lockstep mean that a theory that keeps such extras to the minimum wins.

This version of the competence hypothesis has the effect of generalising the constituent condition to cover the processor. The claim is that the constituents that are recognised in the grammar (and their interpretations) will be the only structures that the processor will give evidence of. Anything else whatsoever that we are forced to postulate is an extra assumption, and will require an independent explanation if it is not to count against the theory. Of course, such an explanation may be forthcoming. Almost no-one would think it a shortcoming to have to postulate a language-independent mechanisms to map grammar rules onto Earley-style dotted rules. Such a mechanism might plausibly have been independently evolved for other functions involving hierarchical control. The postulation of the apparently more special purpose table-driven LR parsing mechanism would presumably be seen by most people as a more serious violation of this strict condition upon the second component of the processor, the non-deterministic algorithm, which we will examine next.

¶6.2 THE ALGORITHM

If we believe that the natural processor must incorporate the competence grammar directly, what more must it include? According to the assumptions with which this section began, it must include a non-deterministic algorithm that will apply the rules of the grammar to accept or reject the string, together with some extra apparatus for simultaneously building a structure

³⁶This stronger condition is somewhat unhelpfully referred to in Steedman [84] simply as “the strong competence hypothesis”. The version defined there and above is stricter than Bresnan and Kaplan’s version.

representing its analysis. Provided that the competence grammar is monos-tratal and monotonic, this structure can be the semantic translation itself, rather than a strictly syntactic structure. Under this view, (which has been commonplace in computational linguistics since the ATN), a syntactic derivation is simply a trace of the way in which this interpretable structure was built.

The processor must also include an oracle (dealt with in the next sub-section) to resolve non-determinism in the grammar for the algorithm. A theory will be successful to the extent that both of these components can be kept as minimal and as language-independent as possible. For this reason, we should be very careful to exclude the possibility that either the algorithm or the oracle covertly embeds rules of a grammar other than the competence grammar.

There are of course a great many algorithms for any given theory of grammar. The alternatives most commonly proposed by computational linguists either work top-down and depth-first through the rules of the grammar, or work bottom-up from the string via the rules, or employ some mixture of the two. For obvious reasons, all such algorithms usually work from the earliest elements of the sentence to the last, or (for the present orthography) leftmost-first.

Such algorithms require (at least) a store, in addition to the competence grammar and a mechanism for eliminating non-determinism. The nature of this store depends upon the automata-theoretic class of the grammar. For context-free grammars it is a push-down store, or stack. For the classes of grammars we shall deal with below, it is a generalisation of the same device. The question that we must ask under the strict competence hypothesis is, how little more can we get away with? In particular, can we get away with nothing more than the theoretical minimum? That is, the stack, plus a control which does nothing except apply rules of grammar to the string and the stack, subject to the adjudication of the oracle, and dealing in no other structures whatever.

It has frequently been observed that the competence hypothesis in this strict form, under which the processor is allowed the absolute minimum of rules and structures in addition to those already implicit in the grammar, seems to favour bottom up-processing, for two reasons. The first reason is

that (leftmost-first) top-down processors need extra modifications in order to deal with (left-)recursive rules, which are abundant in natural languages in the form of adjunction constructions, which engender non-terminating recursion in the processor.³⁷ There are two basic ways of dealing with the problem. One is to add special watchdogs that detect and interrupt endless loops. The other is to use a covering grammar, as discussed above. Both tactics require further assumptions and thus place a greater burden upon the evolutionary account. Hence, in the absence of any further information, the top-down processor seems more costly in evolutionary terms than the bottom up processor, which works for all grammars.

The second reason for favouring bottom-up processing is that the bottom-up parser essentially requires nothing *more* than the minimal automaton that the grammar requires in automata theoretic terms. In the context-free case, this is a simple push-down stack. By assumption, each syntactic rule is paired with a semantic rule defining the interpretation of the parent category in terms of those of the offspring categories. We have also noted that a processor can perfectly well yield such an interpretation in lieu of traditional syntactic structure. It follows that we need add no more than a mechanism (such as term unification) to do the equivalent of binding variables. For example, consider a left-to-right bottom-up analysis of the sentence *Thieves love watches* according to the following context free grammar:

$$(52) \quad \begin{array}{l} S : pred\ subj \rightarrow NP : subj \quad VP : pred \\ VP : verb\ obj \rightarrow V : verb \quad NP : obj \end{array}$$

(The colons associate syntactic categories with their semantic translations). The minimal non-deterministic algorithm says that the parser can either “shift” words from the string to the stack, with their grammatical category showing, or “reduce” the topmost items on the stack according to one of the rules of the grammar, replacing those items by the entity specified on the left-hand-side of the rule. This algorithm allows an analysis in which three successive shifts first results in stack in a, below:

³⁷Of course, rightmost-first top-down processors are vulnerable to the even more prevalent right-recursive constructions.

(53)

<i>NP : watches'</i>
<i>V : love'</i>
<i>NP : thieves'</i>

a.

<i>VP : love' watches'</i>
<i>NP : thieves'</i>

b.

<i>S : love' watches' thieves'</i>

c.

The top two items can then reduce via the second rule of the grammar 52, and be replaced by a VP *love watches*, as in b, whereupon the top two items can reduce to an S.

By contrast, a top-down parser needs a little more in order to analyse this sentence according to the same grammar. This is easiest to see by first considering the recogniser in isolation, then seeing how it would use the same trick of building an interpretation structure directly. In mid analysis, after the words *thieves love* have been encountered, the recogniser's stack must hold information specifying the control structure, or history of recursive rule invocations, as opposed to the simple stack of complete constituents used by the bottom up processor. That is to say that the stack must contain a note of all incomplete categories of which the verb *love* is an element, together with a note of what further categories are predicted by the non-deterministic algorithm to occur in the rest of the string for the completion of each one. We might represent this information as follows, following Earley-style notation and writing *X.Y* to mean "we need a *Y* to complete this *X*":

(54)

<i>VP.NP</i>
<i>S.VP</i>

We can augment this recogniser in order to make it build structures, using the same mechanism as before for building interpretation structures via graph unification. Then the control stack at the same point will look like this:

(55)

<i>VP:love' x.NP:x</i>
<i>S: pred thieves'.VP:pred</i>

It is easy to see that once we accept a further object, *NP : watches'*, the variable *x* can be bound, and the VP will be completed as *VP : love' watches'*.

This in turn can be used to bind the variable *pred* and popped from the stack, yielding the same result, $S : \textit{love}' \textit{watches}' \textit{thieves}'$.

To accomplish this result, we have had to add more apparatus than we needed for the bottom-up algorithm. The entities on the control stack were not confined to grammatical constituents. We needed some extra apparatus to derive entities like $S.VP$ from rules like those in 52, and to manipulate them during a derivation. Indeed, such entities seem suspiciously akin to new structures or grammatical categories, which we might think of as “an S lacking a VP ”, that is, the function from VP to S that in categorial grammar would be recognised as the type-raised category $S/(S \setminus NP)$. The structures that the processor deals in are no longer limited to the constituents licensed by the grammar, and require an extra mechanism to derive them from that grammar. The bottom-up algorithm appears somewhat less costly in evolutionary terms.

We noted earlier that the addition of such a mechanism is a particularly harmless violation of strict competence. A similar kind of mechanism could conceivably have evolved for other ends, such as the hierarchical control of planned actions.³⁸ In this sense, the earlier problem of recursive rules is a more telling reason for rejecting the top-down alternative.

In other respects, as Kimball [53] and Frazier and Fodor [37] have pointed out, the details of real grammars do not seem to be as one would expect from grammars designed for use with a (leftmost-first) bottom-up parser, under the strict competence assumption. All leftmost-first processors, when faced with right-branching constructions must, as the previous examples illustrate, simply put things on the stack until the constituent on the right branch is complete. However, the top-down algorithm is predictive in nature. When the lexically ambiguous word *watches* is encountered, this algorithm can be regarded as predicting an NP, via the stacked category $VP.NP$, a VP “looking for” an NP. Since *no* analysis permitted by the non-deterministic algorithm and the grammar predicts an intransitive verb at this point in the string, the verb *watches* can be ignored. In particular, the fact that this verb can potentially combine with the previous word *love* (which could be an NP) to yield an S is ignored, since that S is not predicted by the grammar either.

³⁸Of course, the inverse claim, that planning is parasitic upon the language faculty, would be equally (if not more) plausible.

The bottom up algorithm, by contrast, is not predictive in this way. Both analyses of the substring *love watches* are accepted by the non-deterministic algorithm. This property threatens to make the job of the oracle harder, since there is a danger of building many structures which play no part in any successful analysis. Unlike the earlier observations concerning the mismatch between (leftmost-first) top-down parsing and left branching, this is not a problem for the algorithm itself, but for the mechanism that copes with non-determinism, so we will delay further discussion of the problem until after some preliminaries in the next section.

¶6.3 AMBIGUITY AND NON-DETERMINISM

Nothing about the expected close relation between syntax and semantics entails that the mapping should be unambiguous, despite the fact that the grammars for artificial languages we design ourselves typically permit only the most local of nondeterminism, because they are designed for use as formal calculi. Expressions in natural languages seem to be remarkably free with ambiguity, both of the global and local variety, as in the following famous examples;

- (56) Flying planes can be dangerous
- (57) a. Have the students taken the exam?
 b. Have the students take the exam!

In the latter example, from Marcus [57], the substring *have the students* is (syntactically and semantically) locally ambiguous, in the sense that a processor cannot immediately know which rule of grammar to apply after encountering the words *have the students*. Human beings seem to be remarkably adept at resolving such ambiguities, which are astonishingly profuse in natural language.³⁹

Probably for the same reason, we do as a matter of fact tend to design our artificial languages in ways which make their symbols “locally” ambiguous,

³⁹This profusion of ambiguity in fact seems to suggest that there is a pressure to keep the lexicon small by comparison with the number of elementary concepts. Such a pressure could reflect a need to keep spoken words from taking too long to utter, in comparison to the accompanying thought processes.

either in terms of which rule of syntax should apply, or in terms of which rule of semantics should apply. (An example of the latter is the “overloading” of an operator like $+$ to denote distinct operations applying to integers, reals and complex numbers.) The one lesson that we can derive from our experience with artificial languages and processors like compilers is that *such ambiguities must be resolvable quickly and locally if the computational complexity of processing is to be contained*.

In order to facilitate this requirement, programming languages are invariably carefully designed so that local ambiguity can be resolved immediately, either syntactically by examining the next symbol in the string, or semantically by examining the types of functions and arguments (as in the case of overloading above). However, natural language shows no sign of any such constraint from within grammar. For example, while the locally ambiguous substring *Have the students ...* in the above example is disambiguated by the next word *take/taken*, an indefinite amount of further linguistic material may intervene between the ambiguous substring and the disambiguating information, as when the sentences begin *Have the students who were late with the homework ...*, *Have the students who were late with the homework that I set last Monday ...*, and so on. This apparent non-determinism in the grammar is an anomaly that requires some further explanation, for if we allow the ambiguities to proliferate, then the costs of maintaining the alternatives will explode. Indeed, as Marcus points out, we must be able to eliminate all but some bounded number of alternative paths, on the basis of purely local evidence, since there is no evidence that processing load increases as a worse than linear function of sentence length. We will call the device that eliminates non-determinism, and decrees which rule of the grammar should be invoked at any point in the derivation, an “oracle”. However this device works, it is clear that it must be very effective in order to deal with the degree of non-determinism that natural grammars exhibit. Moreover, as noted earlier, it must also be entirely language-independent, if it is not to compromise the parsimony and modularity of the theory of the processor.

Most existing accounts of the human sentence processing mechanism have assumed that local ambiguity resolution is based on structural criteria, such as parsing “strategies” (Fodor et al. [33]; Kimball [53]), structural preferences (Frazier, [35]), rule orderings (Wanner [93]); lexical preferences (Ford et al. [34]), or lookahead (Marcus [57]). Such accounts have been claimed to explain

a wide range of sentence processing phenomena, the most spectacular of which is undoubtedly the identification by Bever [9] of the “garden path phenomenon” – that is, the existence of sentences like the following, for which a local ambiguity is *mis*-resolved in a way that makes a perfectly grammatical sentence unanalysable:

(58) The horse raced past the barn fell

However, such accounts have generally been characterised either by empirical shortcomings or by proliferation of devices and degrees of freedom in the theory (see for example the exchange between Frazier and Fodor and Wanner in [37], [93]). In particular, since the earliest stages of the work, it has been clear that all phenomena to do with human parsing were extremely sensitive to the influence of semantics and especially referential context. Bever himself noted a difference in the strength of the garden path effect in minimal pairs of sentences analogous to the following, an effect which he attributed to the differing pragmatic plausibility of analysing the initial noun phrases as a subject of the ambiguous verb/participle *sent*.⁴⁰

(59) a. The doctor sent for the patient arrived

b. The flowers sent for the patient arrived

There were a number of computational proposals for how this effect might work according to a “weak” interaction between syntax and semantics, via a filtering process of comparing rival partial analyses on the basis of their success or failure in referring to entities in the model or discourse context (cf. Winograd [96]). In particular Crain et al. [27] and Altmann et al. [4] proposed a criterion for selecting among analyses called the Principle of Parsimony, which can be stated as follows:

(60) THE PRINCIPLE OF PARSIMONY: the analysis whose interpretation carries fewest unsatisfied presuppositions will be preferred.

These authors use the term “presupposition” in the “pragmatic” sense of Stalnaker and Lewis, and explain this principle in terms of the associated

⁴⁰I owe this modification of Bever’s original example to Michael Niv.

notion of “accomodation” of unsatisfied presuppositions. They point out that the two analyses of sentence 58 as beginning with a simple NP *the horse* and a complex NP *the horse raced past the barn* differ in the number of horses that are presupposed to exist in the model, and in whether it is presupposed (via the modifier) that some agent raced one of them. They argue that contexts which because of previous mention of a horse, or some horses and some racing, already support one or the other set of presuppositions will favour the related analysis at the point of ambiguity, and thereby either induce or eliminate the garden path for this sentence under the principle of parsimony. Crucially, they also argue that the empty context, in which *no* horses and no racing have been mentioned, will favour the simplex NP analysis, because it carries fewer presuppositions, and is therefore easiest to accomodate. The principle accordingly predicts a garden path in the empty context.

The vast majority of early psycholinguistic experiments have used empty contexts, and therefore failed to control adequately for this effect. However, experiments by Carroll *et al.* ([18]), Tannenhaus ([90]), Marslen-Wilson *et al.* ([59]), Crain *et al.* [27]), and Altmann *et al.* ([3], [4]) have produced growing evidence that effects of semantics, and especially of referential context, are extremely strong. Indeed it is now the case that all theories of performance admit some such component, in the form of a “thematic processor” (cf. Frazier [36]) or the equivalent, and the only question is whether anything *else* besides this potentially very powerful source of ambiguity resolution is actually required. (See the exchange between [26] and [88]). For, if interpretations are available at every turn in sentence processing, then there is every reason to suppose that the local syntactic ambiguities which abound in natural language sentences may be resolved by taking into account the appropriateness to the context of utterance of those interpretations, even when the rival analyses are in traditional terms incomplete. Indeed, the possibility that human language processors are able to draw on the information implicit in the context or discourse model seems to offer the only mechanism powerful enough to handle the astonishing profusion of local and global ambiguities that human languages allow, and to explain the fact that human language users are so rarely aware of them. Such a selective or “weak” interaction between syntactic processing and semantic interpretation is entirely modular, as Fodor has pointed out ([32, pp.78,135]).

If interpretation in context is the basis of local ambiguity resolution, then a number of further properties of the parser follow. The felicity of an interpretation with respect to a context is not an all-or-none property, comparable to syntactic well-formedness. Utterances are often surprising – indeed, they are infelicitous if they are *not* at least somewhat novel in content. It follows that evaluation in context can only yield information about the *relative* good fit of various alternatives. We should therefore expect the parser to use a tactic known as “beam-search”, whereby at a point of local ambiguity, all alternative analyses permitted by the grammar are proposed in parallel. The associated interpretations are then evaluated in parallel, readings that fail to refer or are otherwise implausible being discarded or disfavoured in favour of others that are consistent with what is known, along the lines suggested earlier. The parsing process then proceeds with the best candidate(s), all others being discarded or interrupted.

On the assumption that the number of alternative analyses that can be maintained at any one time is strictly limited, we can also assume that the process of semantic filtering occurs very soon after the alternatives are proposed. It should at least be completed before the next point of local ambiguity, for otherwise we shall incur the penalties of exponential growth in the number of analyses. Given the degree of non-determinism characteristic of natural grammars, this means that the interplay of syntactic analysis and semantic adjudication must be extremely intimate and fine-grained. Since most words are ambiguous, semantic adjudication will probably be needed almost word-by-word.

For example, consider example 59, repeated here:

- (61) a. The doctor sent for the patient arrived
 b. The flowers sent for the patient arrived

The garden path effect in a is reduced in b, because flowers, unlike doctors, cannot send for things. The very existence of a garden path effect in a suggests that this knowledge must be available early. If the processor were able to delay commitment until the end of the putative clause *the flowers sent for the patient*, then it would have got to the point of syntactic disambiguation by the main verb, and there would be no reason not to expect it to be able to recover from the garden path. It follows that to explain the lack of such

an effect in b, we must suppose that the interpretation of the incomplete proposition *the flowers sent ...* must be available in advance of processing the PP, and that it is the fact that it has no extension that causes the garden path analysis to be aborted.

However, the proposal to resolve non-determinism by appeal to such interpretations immediately appears to lead to a paradox. If the processor resolves non-determinism in mid-sentence, more or less word by word, on the basis of contextual appropriateness of interpretations, then those interpretations must be available in mid-sentence, also more or less word-by-word. However, under the rule-to-rule hypothesis and the strict version of the competence hypothesis, only *constituents* have interpretations, and only constituents are available to the processor. Now, there is no particular problem about constructing a grammar according to which every leftmost string is a constituent, so that processing can proceed in this fashion incremental processor. Any left-branching grammar provides an example. For such grammars, the assumption of a rule-to-rule compositional semantics means that, for each terminal in a left-to-right pass through the string, as soon as it is syntactically incorporated into a phrase, the interpretation of that phrase can be provided. And since the interpretation is complete, it may also be evaluated – for example, if the constituent is a noun or a noun phrase, then its extension or referent may be found.

A right-branching CF grammar, on the other hand, does not have this property for left-to-right processors: In the absence of some further apparatus going beyond rule-to-rule processing and rule-to-rule semantics, all comprehension must wait until the end of the string, when the first complete constituent is built, and can be interpreted. Until that point, any processor which adheres to the strict competence hypothesis must simply pile up constituents on the stack. It therefore seems that we should expect the languages of the world to favour left branching constructions, at least wherever incremental interpretation is important for purposes of resolving non-determinism. However, the languages of the world make extravagant use of *right*-branching constructions – the crucial clause in 61, *The flowers sent for the patient*, being a case in point. The availability of an interpretation for what is in traditional terms not a constituent, namely *The flowers sent ...* therefore contradicts the strict competence hypothesis.

In this connection, it is interesting that the combinatory theory makes such fragments as *The doctor/flowers sent* available in the competence grammar as constituents, complete with an interpretation, and comparable in every way to more traditional constituents like the clause and the predicate. To the extent that the empirical evidence, say from the experimental comparison of the garden path effect in minimal pairs of sentences like 61, leads us to believe that interpretations are available to the processor for such fragments, it follows that the present theory of grammar gives rise to a simpler account of the processor, without compromising the strict version of the strong competence hypothesis.

It is important to be clear that this problem for traditional right-branching grammars is independent of the parsing algorithm, and applies to bottom-up and top-down algorithms alike, so long as they adhere to the strict competence hypothesis. The top-down algorithm still has the apparent advantage of being syntactically predictive, as we noted earlier, following Kimball [53] and Frazier and Fodor [37].⁴¹ However, neither algorithm of itself will allow an interpretation to be accessed for the leftmost substring, in advance of them being combined into a constituent. Therefore neither algorithm unaided will allow semantic filtering as a basis for the oracle within right-branching constructions.

To say this much is not of course to deny that incremental interpretation is possible for right-branching grammars if they do not adhere to the strict competence hypothesis in this extreme form. Pulman [74, p.212-213] proposes a bottom-up shift reduce processor which includes a rule (“Clear”) that collapses stacked dotted categories such as a subject *the flowers* and a

⁴¹The usual response to this problem is to incorporate some top-down information in a basically bottom up algorithm. Kay [52] has proposed to do this via “reachability tables”. Alternatives are LR and “left-corner” parsing. Such devices of course constitute more or less serious violations of the strict version of the strong competence hypothesis. However, it is worth noting that the combinatory grammar advocated in Part I can be seen as embedding a certain amount of this information in the grammatical categories themselves. For example, a type raised subject $S/(S \setminus NP)$ expresses the fact it is the left corner of an S, and that S is “reachable” from it. Resnik [75] has pointed out that the addition of functional composition to such parsers, either in a left corner parser following Pulman [74], or by implication in the grammar, may have certain desirable effects upon short-term memory requirements for the parser.

potential complement-taking verb *sent*, thus:⁴²

$$(62) \frac{\boxed{VP:send-for' x .PP:x}}{\boxed{S:pred flowers' .VP:pred}} \Rightarrow \boxed{S:send-for' x flowers' .PP x}$$

However, this operation is clearly nothing more than functional composition. The theory sketched in Part I proposes such an operation as a component of competence grammar. It therefore predicts this operation in the processor, even under the strict competence hypothesis, rather than requiring it as an extra stipulation and thereby violating strict competence.⁴³

Similarly, we can make the earlier sketch of a top-down (non-deterministic) parsing algorithm incrementally interpretive by unifying semantic terms on the stack in advance of syntactic combination, yielding the following state for the control stack after the words *The flowers sent* have been encountered in the example, thereby potentially making available the partially interpreted category *S:send-for' x flowers'*, which might be used by some computational demon or watchdog to reject the analysis on the basis of violating a selectional restriction:

$$(63) \frac{\boxed{VP:send-for' x .PP:x}}{\boxed{S:pred flowers' .VP:pred}} \Rightarrow \frac{\boxed{VP:send-for' x .PP:x}}{\boxed{S:send-for' x flowers' .VP:send-for' x}}$$

Again, this operation amounts to stipulating composition as a rule of the processor, this time in the semantics. The selection restriction-detecting demon is also clearly in violation of strict competence hypothesis.

There is in fact no sense in which a parser using a right-branching grammar under strict competence could conceivably access interpretations for substrings that are not constituents. To be absolutely clear that this is the case let us suppose that we parse with a pure categorial grammar, and are therefore given lexical type-raising, so that NPs can be paired with semantic roles in the proposition. Let us consider the question of whether the anomaly of the active verb sense of *the flowers sent* ... can be detected in advance of processing *to the patient*. Let us assume a context that contains

⁴²The notation is adapted.

⁴³Pulman also has type-raising as a rule of processing, under the name “Invoke”. See Haddock [43], [44] for proposals for a more realistic incremental semantics than the toy version assumed here.

exactly two propositions, namely that a doctor sent some flowers to a patient (say *send – for' flowers' patient' doctor'*), and the flowers smelt nice (*smell' nice' flowers'*). Imagine that the parser is the simplest bottom-up parser, that must entertain all possible categories for the NP *the flowers*, and that there are just three of these, the categories corresponding to a subject and an object, and the category involved in the modifier analysis that will ultimately succeed, if the garden path can be avoided.

For each of these categories, a processor could without violating strict competence find all possible predications over the referent that are supported by the database, and associate that set with the category, as a constraint upon its interpretation. Since NPs are type-raised categories, we can restrict this set to predications of a particular theta-role. If this set is empty, the processor could conceivably use this information to suspend further investigations of this path.⁴⁴ This much would be entirely in keeping with the strict competence hypothesis, since we have dealt in nothing but grammatical constituents.

However, it is not going to help us to detect the anomaly of the spurious sentence #*The flowers sent for the patient* in the example 61b before the end of the clause, thereby escaping the garden path. The set of propositions that are compatible with *the flowers* as a subject includes *smell' nice' flowers'* and nothing else. This set might conceivably be viewed as a constraint on possible outcomes of the parse on this path.⁴⁵ Similarly, the path that starts with the assumption that *the flowers* is an object contains the single proposition *send – for' flowers' patient' doctor'*. (This latter path will not in fact

⁴⁴Of course, to say that the detection of anomaly amounts to no more than the absence of an extension in the discourse model is a caricature. In real life, it is considered rather bad form for an utterance to provide no novel information. A more realistic example would have to allow for the accomodation of novel predications into the model, in which case no strong conclusion could be drawn from the mere absence of a compatible extension *as yet*. Nevertheless, on occasion the context may in fact rule out all available analyses, causing an analysis to fail early, in advance of sentence internal disambiguating material. The possibility of such premature interrupts does mean that we can never rule out any theory of competence grammar from mere psycholinguistic evidence of access to interpretations. However, the primacy of competence over performance is not at issue.

⁴⁵Again, this is a caricature. We certainly cannot assume that the actual sentence will correspond to the fact we already know – in fact, we ought to be able to assume the reverse, if our interlocutor is being cooperative.

yield even a syntactic analysis corresponding to this proposition, of course).

When the word *sent* is processed, it is similarly conceivable that we may notice that the model supports a single proposition concerning this relation, which for the sake of argument, we will again allow ourselves to regard as some kind of constraint on possible outcomes. However, we can do nothing further whatsoever with the information we have associated with the putative subject and the verb without violating strict competence. If we were allowed to intersect the sets we might observe that the intersection was empty. But we can draw no conclusion from this fact until we know that the subject and the verb belong to the *same* clause or proposition.⁴⁶ The only legitimate source for this information is the grammar. However, the grammar we have chosen does not know about subjects and verbs. It only knows about subjects and predicates. The analysis must therefore continue. When the processor encounters the words *for the patient*, it will find them consistent with the information in the context, and will complete the predicate. It will then find that the predicate can syntactically combine with the subject, at which time the anomaly can be detected.

However, at this point it is too late, as we can tell from the presence of a garden path effect in 61a. So we have failed to achieve incremental interpretation. Notice also that this is an absurd way to run a parser. In the toy example, extensions were very small sets. However, for realistic models they will be very large. Moreover, since they are constructed without benefit of syntactic information, we must potentially maintain vast numbers of such sets, most of which will never correspond to any syntactic analysis at all. (For example, merely because *the flowers* is locally categorised as a subject, we had to construct the set of all propositions with the referent as an object, despite the fact that no analysis will ever need the information.)

Stabler [81], criticising an earlier version of this proposal,⁴⁷ has argued that the present claims are in error, and that incremental interpretation of right-branching constituents is in fact possible without violating the strong competence hypothesis in the strict sense used here. This claim does not go through, and it is worth considering why.

⁴⁶In English, it happens that they must be. However, if we take advantage of this fact, we will be in violation of strict competence. In Japanese, which is verb-final and relative clause-initial, the assumption would break down.

⁴⁷Steedman [84].

First, it is important to be clear that Stabler technically does not adhere to strong competence in the strict sense. It is clear that he is assuming a weaker definition of the competence assumption, although he avoids an explicit definition.⁴⁸ In particular, in his first worked example, [81, p.226], he binds the variable *Subj* in the interpretation of the sentence to the interpretation of the actual subject, via partial execution in the rule *I1* (p.208). This is only possible because he is using the grammar as a predictive parser. He uses this information to identify the fact that since the context only includes one predication over this subject, that must be the one that is to come, under a similar caricature of incremental evaluation to that used above.

However, we have seen that this much is merely information that could legitimately have been built into the grammar itself, in the form of type raising. (In fact this analogy is effectively embodied in his second example using an LR parser, although the details here are less clear.)⁴⁹ As in the example offered above, Stabler has not actually handled any interpretations that correspond to non-constituents. It is therefore more important to ask whether his processor in all other respects adheres to the strict hypothesis by avoiding interpretation of non-constituents of the *The flowers sent* variety entirely, or whether it violates the hypothesis by covertly constructing interpreted objects which are not merely constituents according to the competence grammar, either in the form of Earley-style dotted categories or in the form of partially instantiated semantic interpretations.

Curiously, since his paper is ostensibly addressed to a predecessor of the present proposal, and despite the fact that he has technically allowed strict competence to be compromised, all of Stabler's actual examples take the first tactic. Thus in his exegesis of the (right-branching) sentence *The joke is funny* there is no sense in which there ever exists an interpretation of the non-constituent *The joke is*, or indeed anything comparable to *The flowers sent* (see Stabler [81, p.215 and 232]).⁵⁰ Were he to address the question raised by example 61, his parser would offer no help. The fact is that Stabler's processor is neither consistent with the strict competence hypothesis, nor incrementally interpretative in the sense argued for here.

⁴⁸See p.233, note 1.

⁴⁹The use of LR tables is of course another technical violation of strict competence.

⁵⁰Stabler's note 3 suggests that he has not noticed that this is the central issue.

The resolution of the paradox of incremental interpretation does not lie with Stabler’s parser, but rather with the observation that competence grammar itself seems to be telling us that strings like *Mary prefers* and *The flowers sent* are in some sense constituents. It follows that we can retain the strict version of the strong competence hypothesis, and to continue to require the grammar to support incremental interpretation, if we also take on board the combinatory theory of grammar. This theory offers a broader definition of constituency, under which more substrings, and in particular more left prefixes, are associated with interpretations as a matter of grammar. The interpretations of such non-standard constituents can therefore be used to compare rival analyses arising from non-determinism on the basis of their fit to the context, without violating the strict competence hypothesis.

§7 CONCLUSION

The paper began by stating some uncontroversial assumptions, in the form of the rule-to-rule assumption, and the competence hypothesis, and deducing the even more widely accepted constituent condition on rules of competence grammar. Having noted the difficulties presented by coordination and intonation viz-a-viz the constituent condition, the paper went on to advance an alternative view of competence grammar under which the paradoxical constructions could be seen to conform to that condition after all. In Part II, the consequences for performance under a “strict” version of the competence hypothesis were derived.

The competence theory in question implies the strongest possible relation between surface structure, intonation structure and information structure. This property of the theory in itself has significant implications for processing under the strict competence hypothesis. The isomorphism between syntactic structure and intonational structure means that simply-structured modular processors which use both sources of information at once should be easier to devise. Such an architecture may reasonably be expected to simplify the problem of resolving local structural ambiguity in both domains.

However, we have noted that a considerable amount of non-determinism remains in the grammar, both for spoken and written language. The properties of the grammar are consistent with the suggestion that the basis for the oracle that renders the process as a whole deterministic is the incre-

mental availability of semantic interpretations. The generalised notion of constituency that is engendered by the combinatory rules ensures that leftmost substrings are potentially constituents with interpretations, subject of course to the limitations of the grammar and any further information that may be available from intonation. Such a theory of grammar may therefore have an added advantage of parsimony, in being compatible with such a processor without compromising the strict competence hypothesis.

Indeed, we can stand this argument on its head. If we believe that the parser has to know about interpretations corresponding to strings like *the flowers sent for ...*, and we identify such interpretations with the notion of abstraction, then the advocate of traditional grammar must ask themselves why their grammar does *not* endorse such useful semantic concepts as grammatical constituents.

The claim is strengthened by the observation that the residual non-determinism in the grammar of intonation, arising from the use of the null tone, as discussed in connection with example 43, is confined precisely to those occasions on which the theme is believed by the speaker to be entirely known to all parties, and to be recoverable by comparison of the interpretation of a (usually leftmost) substring with the contextually contextually open proposition. It would be surprising if the mechanism for disambiguating written language were very different from its ancestor in the processor for spoken language.

It is of course unlikely that we will ever know enough about the biological constraints to evaluate the assumptions on which the “strict” version of the competence hypothesis is based with any certainty. In the absence of such certainty, we must beware of falling into the error of evolutionary Panglossism. However, it is in order to speculate upon its implications for the theory as a whole, for the following reason.

Competence grammar and performance mechanism are in the end a package deal. Any claim about competence grammar is also a claim about the entire package. The theory of evolution is not mocked, and one day, a reckoning will be demanded. On that day, the linguistic theories that have survived the ordeal of descriptive adequacy will be judged not merely on their purity and parsimony as theories of competence, but on their explanatory value as part of the package. We have already noted that all theories will require

something more, in the form of a language-independent mechanism for resolving local ambiguity, or grammatical non-determinism, together with a language-independent algorithm and automaton. But if a theory of competence requires much more than that, or if that mechanism in turn implicates a notion of structure that is not covered by the competence grammar, then those assumptions will weigh against it. If there is another theory that requires fewer such assumptions, perhaps even *no* further assumptions beyond the mechanism for resolving non-determinism and the minimal bottom-up algorithm, say because it has a different notion of surface syntax, then the scales may tilt in its favour.

That day is not yet upon us, but it is good on occasion to meditate upon one's latter end. If it is true that the principal responsibility for local ambiguity resolution lies with incremental interpretation at extremely fine grain, then any theory which does not make similar assumptions to CCG about constituency in the competence grammar will, as we have seen above, have to make some strikingly similar structures available to the processor, complete with interpretations. Such additional assumptions would not in any sense be inconsistent with the competence theory itself. However, they compromise the strict version of the competence hypothesis. To the extent that the combinatorial theory achieves the same result without any additional assumptions, it may be preferred.

REFERENCES

- [1] Ades, Anthony and Mark Steedman: 1982, 'On the Order of Words', *Linguistics and Philosophy*, 4, 517-558.
- [2] Ajdukiewicz, Kazimierz: 1935, 'Die syntaktische Konnexität', *Studia Philosophica*, 1, 1-27. English translation in Storrs McCall (ed), *Polish Logic 1920-1939*, Oxford University Press, pp. 207-231.
- [3] Altmann, Gerry: 1988, 'Ambiguity, Parsing Strategies, and Computational Models' *Language and Cognitive Processes*, 3, 73-98.
- [4] Altmann, Gerry and Mark Steedman: 1988, 'Interaction with Context During Human Sentence Processing' *Cognition*, 30, 191-238
- [5] Bach, Emmon: 1976, 'An extension of Classical Transformational Grammar', in *Problems in Linguistic Metatheory*, Proceedings of the Conference at Michigan State University.
- [6] Bar-Hillel, Yehoshua: 1953, 'A Quasi-arithmetical Notation for Syntactic Description', *Language*, 29, 47-58.
- [7] Beckman, Mary, and Janet Pierrehumbert, 1986, 'Intonational structure in Japanese and English', *Phonology Yearbook*, 3, 255-309.
- [8] Berwick, Robert and Amy Weinberg: 1984, *The grammatical Basis of Linguistic Performance*, MIT Press, Cambridge MA.
- [9] Bever, Tom G: 1970, 'The Cognitive Basis for Linguistic Structures', in J. Hayes, (ed.), *Cognition and the Development of Language*, Wiley, New York NY.
- [10] Bird, Steven: 1991, 'The Intonational Phrase in Sign-based Grammar', Workshop on Declarative Perspectives on the Syntax-Prosody Interface, Utrecht, May 1991.
- [11] Brame, Michael: 1976, *Conjectures and Refutations in Syntax*, Elsevier/North Holland, New York NY.

- [12] Brame, Michael: 1978, *Base Generated Syntax*, Noit Amrofer, Seattle WA.
- [13] Brown, Roger and Camille Hanlon: 1970, 'Derivational Complexity and Order of Acquisition in Child Speech' in John Hayes (ed.), *Cognition and the Development of Language*, Wiley, New York. 155-207.
- [14] Bolinger, Dwight: 1985. 'Two views of Accent', *Journal of Linguistics*, 21, 79-124.
- [15] Bolinger, Dwight: 1989. *Intonation and its Uses*, Stanford University Press.
- [16] Bresnan, Joan: 1978, 'A Realistic Transformational Grammar', in M. Hallé, J. Bresnan, and G. Miller (eds.), *Linguistic Structure and Psychological Reality*, MIT Press, Cambridge, MA.
- [17] Bresnan, Joan (ed.): 1982, *The mental Representation of Grammatical Relations*, MIT Press, Cambridge, MA.
- [18] Carroll, J. and Tom Bever: 1978, 'The Perception of Relations', in William J. M. Levelt and Giovanni Flores d'Arcais (eds.), *Studies in the Perception of Language*, Wiley, New York NY.
- [19] Chomsky, Noam: 1957, *Syntactic Structures*, Mouton, the Hague.
- [20] Chomsky, Noam: 1965, *Aspects of the Theory of Syntax*, MIT Press, Cambridge MA.
- [21] Chomsky, Noam: 1970, 'Deep Structure, Surface Structure, and Semantic Interpretation', in D. Steinberg and L. Jakobovits, *Semantics*, CUP, Cambridge, 1971, 183-216.
- [22] Chomsky, Noam: 1973, *The Logical Structure of Linguistic Theory*, Chicago University Press, Chicago IL.
- [23] Chomsky, Noam: 1975, *Reflections on Language*, Pantheon, New York NY.
- [24] Chomsky, Noam: 1982, *Concepts and Consequences of the Theory of Government and Binding*, MIT Press, Cambridge MA.

- [25] Chomsky, Noam and Morris Hallé, 1968, *The Sound Pattern of English*, Harper and Row, New York.
- [26] Clifton, Charles and Fernanda Ferreira: 1989, 'Ambiguity in Context', *Language and Cognitive Processes*, 4, in press.
- [27] Crain, Stephen and Mark Steedman: 1985, 'On not being led up the garden path: the use of context by the psychological parser', in D. Dowty, L. Karttunen, and A. Zwicky, (eds.), *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*, ACL Studies in Natural Language Processing, Cambridge University Press, 320-358.
- [28] Cresswell, Max: 1973, *Logics and Languages*, Methuen, London.
- [29] Curry, Haskell and Robert Feys: 1958, *Combinatory Logic*, North Holland, Amsterdam.
- [30] Dowty, David: 1988, 'Type raising, functional composition, and non-constituent coordination', in Richard T. Oehrle, E. Bach and D. Wheeler, (eds.), *Categorial Grammars and Natural Language Structures*, Reidel, Dordrecht, 153-198.
- [31] Emonds, Joseph: 1976, *A Transformational Approach to English Syntax*, Academic Press, New York NY.
- [32] Fodor, Jerry: 1983, *The Modularity of Mind*, MIT Press/Batsford, Cambridge MA.
- [33] Fodor, Jerry, Tom Bever, and Merrill Garrett: 1974, *The Psychology of Language*, McGraw-Hill, New York NY.
- [34] Ford, Marilyn, Joan Bresnan, and Ronald M. Kaplan: 1982, 'A Competence-based Theory of Syntactic Closure', in Joan Bresnan (ed. *The Mental Representation of Grammatical Relations* MIT Press, Cambridge MA.
- [35] Frazier, Lyn: 1978, *On Comprehending Sentences*, PhD dissertation, University of Connecticut, distributed by Indiana University Linguistics Club.

- [36] Frazier, Lyn: 1989, 'Against Lexical Generation of Syntax', in William Marslen-Wilson, (ed.), *Lexical Representation and Process*, MIT Press, Cambridge MA. 505-528.
- [37] Frazier, Lyn, and Janet Fodor: 1978, 'The Sausage Machine: a New Two-stage Parsing Model', *Cognition*, 6, 291-325.
- [38] Gazdar, Gerald: 1981, 'Unbounded Dependencies and Coordinate Structure', *Linguistic Inquiry*, 12, 155-184.
- [39] Gleitman, Lila: 1991, 'The Structural Sources of Verb Meanings', *Language Acquisition*, Incomplete reference.
- [40] Gold, E.M: 1967, 'Language Identification in the Limit', *Information and Control*, 16, 447-474.
- [41] Gray, J. and M. Harrison: 1972, 'On the Covering and Reduction Problems for Context-free Phrase Structure Grammars', *Journal of the Association for Computing Machinery*, 19, 385-395.
- [42] Grosz, Barbara and Candy Sidner, 1986, 'Attention, Intention and the Structure of Discourse', *Journal of Computational Linguistics*, 12.
- [43] Haddock, Nick, 1987, 'Incremental Interpretation and Combinatory Categorical Grammar', *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, Milan, 661-663.
- [44] Haddock, Nick, 1989, *Incremental Semantics and Interactive Syntactic Processing*, Ph.D Dissertation, University of Edinburgh.
- [45] Hepple, Mark and Glyn Morrill, 1989, 'Parsing and Derivational Equivalence', *Proceedings of the 4th Conference of the European Association for Computational Linguistics*, Manchester, April 1989, 10-18.
- [46] Halliday, Michael: 1967, *Intonation and Grammar in British English*, Mouton, The Hague.
- [47] Halliday, Michael, 1967, 'Notes on Transitivity and Theme in English, Part II', *Journal of Linguistics*, 3, 199-244

- [48] Halliday, Michael, 1970, 'Language Structure and Language Function", in John Lyons (ed.), *New Horizons in Linguistics*, Penguin.
- [49] Hirschberg, Julia and Janet Pierrehumbert: 1986, 'Intonational Structuring of Discourse', *Proceedings of the 24th Meeting of the Association for Computational Linguistics, New York*, 136-144.
- [50] Jackendoff, Ray: 1972, *Semantic Interpretation in Generative Grammar*, MIT Press, Cambridge MA.
- [51] Karttunen, Lauri: 1986, 'Radical Lexicalism', in Mark Baltin and Tony Kroch, (eds.), to appear, University of Chicago Press, (incomplete reference).
- [52] Kay, Martin: 1980, 'Algorithm Schemata and Data Structures in Syntactic Processing, CSL-80-12, Xerox PARC.
- [53] Kimball, John: 1973, 'Seven Principles of Surface Structure Parsing in Natural Language', *Cognition*, 2, 15-47.
- [54] Kuno, Susumo: 1966, 'The Augmented Predictive Analyser for Context-free Languages: Its Relative Efficiency', *Communications of the Association for Computing Machinery*, 9, 810-823.
- [55] Levelt, Willem: 1989, *Speaking*, MIT Press/Bradford, Cambridge MA.
- [56] Lyons, John: 1977, *Semantics, Vol.II*, CUP, Cambridge.
- [57] Marcus, Mitch: 1980, *A Theory of Syntactic Recognition for Natural Language*, MIT Press, Cambridge MA.
- [58] 'Artificial Intelligence A Personal View' *Artificial Intelligence*, 9, 37-48.
- [59] Marslen-Wilson, William, Lorraine K. Tyler and Mark Seidenberg: 1978, 'The Semantic Control of Sentence Segmentation', in William J. M. Levelt and Giovanni Flores d'Arcais (eds.), *Studies in the Perception of Language*, Wiley, New York NY.
- [60] McCawley, James: 1982 'Parentheticals and Discontinuous Constituent Structure', *Linguistic Inquiry*, 13, 91-106.

- [61] Miller, George: 1967: 'Project Grammarama', in G. Miller, (ed.) *The Psychology of Communication*, Penguin Books, Harmondsworth, 124-181.
- [62] Morrill, Glyn: 1988, *Extraction and Coordination in Phrase Structure Grammar and Categorical Grammar*, unpublished Ph.D thesis, University of Edinburgh.
- [63] Nijholt, Anton: 1980, *Context-free Grammars: Covers, Normal Forms, and Parsing*, Springer Verlag, Berlin.
- [64] Pareschi, Remo, and Mark Steedman, 1987, 'A lazy way to chart parse with categorical grammars', *Proceedings of the 25th Annual Conference of the ACL, Stanford*, July 1987, 81-88.
- [65] Partee, Barbara and Mats Rooth: 1982, 'Generalised Conjunction and Type Ambiguity', Ms, University of Massachusetts, Amherst.
- [66] Pereira, Fernando: 1985, 'A New Characterisation of Attachment Preferences', D. Dowty, L. Karttunen, and A. Zwicky, (eds.), *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*, ACL Studies in Natural Language Processing, Cambridge University Press, 307-319.
- [67] Pereira, Fernando, and Martha Pollock: 1989. 'Incremental Interpretation', ms. ATT Bell Labs and SRI International. (Incomplete reference.)
- [68] Pierrehumbert, Janet: 1980, *The Phonology and Phonetics of English Intonation*, Ph.D dissertation, MIT, distributed by Indiana Linguistics Club, Bloomington, IN.
- [69] Pierrehumbert, Janet, and Mary Beckman: 1989, *Japanese Tone Structure*, MIT Press, Cambridge MA.
- [70] Pierrehumbert, Janet, and Julia Hirschberg, 1987, 'The Meaning of Intonational Contours in the Interpretation of Discourse', ms. Bell Labs.
- [71] Pinker, Steven: 1979, 'Formal Models of Language Learning', *Cognition*, 7, 217-283.

- [72] Pollard, Carl, and Ivan Sag: 1983, 'Reflexives and Reciprocals in English: and An Alternative to the Binding Theory', in M. Barlow et al. (eds.) *Proceedings of the 2nd West Coast Conference on Formal Linguistics*, Stanford CA, 189-203.
- [73] Prince, Ellen F. 1986. On the Syntactic Marking of Presupposed Open Propositions. Papers from the Parasession on Pragmatics and Grammatical Theory at the 22nd Regional Meeting of the Chicago Linguistic Society, 208-222.
- [74] Pulman, Steve: 1986, 'Grammars, Parsers, and Memory Limitations', *Language and Cognitive Processes*, 197-226.
- [75] Resnik, Philip: 1992, 'Left-corner Parsing and Psychological Reality' ms. Dept. of CIS, University of Pennsylvania.
- [76] Rooth, Mats: 1985, *Association with Focus*, unpublished PhD dissertation, University of Massachusetts, Amherst.
- [77] Rooth, Mats, and Barbara Partee, 1982, 'Conjunction, Type-ambiguity, and Wide-scope "or"', *Proceedings of the 1st West Coast Conference on Formal Linguistics*, ed. by Dan Flickinger et al., 353-62, Stanford CA.
- [78] Selkirk, Elisabeth: *Phonology and Syntax*, MIT Press, Cambridge MA.
- [79] Smullyan, Raymond: 1985, *To Mock a Mockingbird*, Knopf, New York.
- [80] von Stechow, Arnim: 1989, 'Focussing and Backgrounding Operators', Fachgruppe Sprachwissenschaft der Universität Konstanz, Arbeitspapier Nr. 6.
- [81] Stabler, Ed: 1991, 'Avoid the Pedestrian's Paradox', in Robert Berwick, S. Abney, and C. Tenny (eds.), *Principle-based Parsing*, Kluwer, Dordrecht. 199-238.
- [82] Steedman, Mark, 1985, 'Dependency and Coordination in the Grammar of Dutch and English', *Language*, 61, 523-568.
- [83] Steedman, Mark: 1987. 'Combinatory grammars and parasitic gaps', *Natural Language and Linguistic Theory*, 5, 403-439.

- [84] Steedman, Mark: 1989, 'Grammar, Interpretation and Processing from the Lexicon', in W. Marslen-Wilson (ed.), *Lexical Representation and Process*, MIT Press/Bradford, Cambridge MA. 463-504.
- [85] Steedman, Mark: 1990, 'Gapping as Constituent Coordination', *Linguistics and Philosophy*, 207-263.
- [86] Steedman, Mark: 1991, 'Structure and Intonation', *Language*, **67**, 260-296.
- [87] Steedman, Mark: 1991, 'Surface Structure, Intonation, and "Focus"', in E. Klein and F. Veltmann, (eds.) *Natural Language and Speech: Proceedings of the Symposium, ESPRIT Conference, Brussels, Nov. 1991*, Kluwer, Dordrecht. 21-38.
- [88] Steedman, Mark and Gerry Altmann: 1989, 'Ambiguity in Context: a Reply', *Language and Cognitive Processes*, 4, 105-122.
- [89] Szabolcsi, Anna: 1987, 'On Combinatory Categorical Grammar', *Proceedings of the Symposium on Logic and Language, Debrecen*, Akadémiai Kiadó, Budapest, 151-162.
- [90] Tanenhaus, Michael: 1978, *Sentence Context and Sentence Perception*, Ph D thesis, Columbia University.
- [91] Uszkoreit, Hans: 1986, 'Categorical Unification Grammars'. In *Proceedings of the 11th International Conference on Computational Linguistics*, Bonn, August, 1986, pp187-194.
- [92] Vijay-Shanker, K. and David Weir: 1990, 'Polynomial Time Parsing of Combinatory Categorical Grammars', *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, Pittsburgh, June 1990, 1-8.
- [93] Wanner, Eric: 1980, 'The ATN and the Sausage Machine: Which One is Baloney?' *Cognition*, 8, 209-225.
- [94] Weir, David, and Aravind Joshi: 1988, 'Combinatory Categorical Grammars: Generative Power and Relation to Linear CF Rewriting Systems', *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, Buffalo, June 1988, 278-285.

- [95] Wilson, Dierdre and Dan Sperber:1979, 'Ordered Entailments', *Syntax and Semantics, XI, Presupposition*, Academic, New York, 299-323.
- [96] Winograd, Terry, *Understanding Natural Language*, Edinburgh University Press.
- [97] Wittenburg, Kent B. 1986. *Natural Language Parsing with Combinatory Categorical Grammar in a Graph-Unification Based Formalism*. D.Phil Dissertation, U. of Texas at Austin.
- [98] Wittenburg, Kent: 1987, 'Predictive Combinators: a Method for Efficient Processing of Combinatory Grammars', *Proceedings of the 25th Annual Conference of the ACL, Stanford*, July 1987, 73-80.
- [99] Zeevat, Henk, Klein, E. and Calder, J. (1987) 'An Introduction to Unification Categorical Grammar'. In N. Haddock et al. (eds.), *Edinburgh Working Papers in Cognitive Science, 1: Categorical Grammar, Unification Grammar, and Parsing*.
- [100] Zipf, George Kingsley: 1936, *The Psycho-biology of Language: an Introduction to Dynamic Philology*, Routledge, London.