Definitions

Implications for Syntax, Semantics, and the Language of Thought

Annabel Cormack

Outstanding Dissertations in Linguistics

OUTSTANDING DISSERTATIONS IN

LINGUISTICS

edited by

LAURENCE HORN
YALE UNIVERSITY

This page intentionally left blank

DEFINITIONS

IMPLICATIONS FOR SYNTAX, SEMANTICS, AND THE LANGUAGE OF THOUGHT

ANNABEL CORMACK

Routledge
Taylor & Francis Group
New York London

First published by Garland Publishing, Inc.

This edition published 2011 by Routledge:

Routledge
Taylor & Francis Group
711 Third Avenue
New York, NY 10017

Routledge
Taylor & Francis Group
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN

Library of Congress Cataloging-in-Publication Data

Cormack, Annabel, 1939–
Definitions : implications for syntax, semantics, and the language of thought / Annabel Cormack.
p. cm. — (Outstanding dissertations in linguistics)
A slight revision of the author's thesis (Ph. D.)—University College, London, 1989.
Includes bibliographical references and index.
ISBN 0-8153-3131-2 (alk. paper)
1. Lexicography. 2. Definition (Logic) 3. Grammar, Comparative and general—Syntax. 4. Thought and trhinking. 5. Semantics. I. Title. II. Series.
P327.C67 1998
413'.028—dc21

98-28259

Contents

Preface vii

Acknowledgments xi

1 Theoretical Preliminaries 3

1.0 Introduction 3
1.1 A little simple syntax 6
1.2 Relevance theory 30
1.3 The 'language of thought' hypothesis 36
1.4 Model theoretic semantics 44
1.5 The formal interpretation of definitions 56
Notes 77

2 Dictionary Definitions 83

2.0 Introduction 83
2.1 Simple nouns 85
2.2 Simple adjectives 94
2.3 Elimination of simple nouns and adjectives 113
2.4 Dictionary definitions—elimination and complements 130
2.5 Problem definitions 151
Notes 156

3 Text Definitions 163

3.0 Introduction 163
3.1 Characteristics of text definitions 165
3.2 Interpretation of simple definitions 174
3.3 Complex definitions with 'if' or 'when' 195
Notes 222

4 Syntax: Conclusions, Problems, and Speculations 231

4.0 Introduction: argument structure in syntax and semantics 231
4.1 Distinctions between adjectives: c-selection and learnability 236
4.2 Syntactic mechanisms for θ-roles 262
Notes 291

5 The Language of Thought: Findings, Problems, and Speculations 301

5.0 Points of departure 301
5.1 Variables 302
5.2 The expressive power and the syntax of the language of thought 305
5.3 LF as the sole input to interpretation? 310
5.4 Conceptual adicity, s-selection and c-selection 313
5.5 How are definitions interpreted and used? 317
5.6 Conclusions 325
Notes 326

Bibliography 329

Index 339

Preface

> What! you have had wings for ten years, and you haven't flown yet!
> Poincaré, *Science and method*; addressed to the 'logisticians'

Apart from this preface itself, what follows is essentially my 1989 Ph.D. thesis, with the omission of the old preface and abstract, and the provision of an index. I have updated bibliographical entries, and added those works mentioned here. The only other changes are typographical, and the correction of typing errors.

As I look at it now, although much is out of date, and some analyses are unsatisfactory, there is also much that has not been discussed elsewhere, or not discussed from the perspectives I employed. In particular, I think that many of the questions I have raised, and some of my conclusions, warrant further work, which I hope may be stimulated by this publication.

The focus is on definitions, which were culled both from a variety of dictionaries (chapter 2), and from various text books (chapter 3). I consider on the one hand, the syntax and semantics of the definition itself, and on the other, how the information conveyed by the definition can be put to use in constructing an item in the mental lexicon. Both kinds of definition pose problems of interpretation and usability, involving semantics and pragmatics, and the text definitions in particular raise syntactic problems as well. I attacked the data using, and as necessary adapting, the theoretical tools of Principles-and-Parameters syntax (Chomsky 1986a), Fodor's (1975) Language of Thought hypothesis (backed up by model theoretic semantics) for meaning, and Relevance theory for pragmatics (Sperber and Wilson 1986). The first chapter discusses relevant elements of each of these tools. It was I think the confrontation of a corpus of data with all these theories simultaneously which was so exciting, and productive of interesting questions and hypotheses.

The analyses raise numerous new questions, and suggest some answers. It is argued that because phrases, as occurring in dictionary definitions, can be understood in isolation, 'NP movement' has to be reanalyzed as transmission of θ-roles. In chapter 4, these ideas are applied to a variety of adjectives which take propositional

complements. I discuss in detail not only what lexical specification is needed to distinguish the various syntactic options for each kind of adjective, but how the child can determine what the specification for the lexical item is. For adjectives such as *ready*, I argue that the external argument obtains one θ-role directly (by virtue of external selection by the head adjective) and one indirectly (by virtue of θ-transmission from an internal position). The θ-transmission from the internal position corresponds to either raising, or *tough*-movement, with correspondingly different interpretations for *ready*.

The main interest of the text definitions is the interplay between object language and metalanguage. Most such definitions are to be construed as statements about the meanings of words of English, but the definiendum is used, not just mentioned, within the definition, despite the fact that its meaning is not known previously. It is argued that for such definitions to be understood, the syntax of the Language of Thought must be close to that of Natural Language in specifiable ways. For example, semantic types must be common to the two languages. There is also considerable involvement of pragmatics, with some interesting consequences, particularly to ensure that the interpretation conforms to the requirements on a well-formed definition.

Both syntax and semantics have developed considerably in the last ten years, with more syntacticians paying attention to semantic issues. It is obvious that had I been addressing the same questions now, my answers would often have been different. In particular, the discipline of the Minimalist program (Chomsky 1995) would have been salutary in limiting the options apparently available.

There has been a fair amount of work which is directly pertinent to my concerns in this thesis. I mention here the most relevant of this. I should have mentioned Bierwisch and Kiefer (1970) in relation to the general problems raised by definitions. In the discussion of section 1.4, it would have been useful to have been able to refer the reader to Larson and Segal (1995), and Heim and Kratzer (1998). Fodor's Language of Thought hypothesis, and in particular the status of its semantics, has been the subject of much discussion by philosophers including Fodor himself. Carruthers (1996) considers much of this. Jackendoff (1992, chapter 2), places Conceptual Semantics in relation to model theoretic semantics and Internalist theories of meaning (and see Zwarts and Verkuyl 1994). Grimshaw 1990 and Williams 1994 are major discussions of argument structure and its relation to syntax. The

content of the lexicon has been investigated under various approaches, from the computational to the conceptual: Boguraev and Briscoe (1989), Pustejovsky (1995), Levin (1993), and Jackendoff (1990) all discuss substantial portions of the lexicon in some detail. Dowty's (1991) discussion of θ-roles is relevant to my section 2.4.3. Cinque's (1990) paper on ergative adjectives is relevant to section 4.1.2, and Stowell (1991) to the discussion of *stupid* in section 4.1.3. Rothstein (1991) discusses the (absence of) projection of minor categories (compare my 'quasi-projections' in section 1.1.2). Chomsky and Lasnik (1993) argue that PRO has Case (see my section 4.1.2). Jacobson's 1990 paper *Raising As Function Composition* is important in relation to my notion of θ-movement (my section 2.2.2 and chapter 4). Bošcović (1994) argues that there can be some movement into θ-positions, and see also Vikner (1988) who invokes the 'addition' of θ-roles for some modals (relevant to section 4.1.3). Hornstein (1996), and Manzini and Roussou (1997) independently argue that PRO can be dispensed with (as I argue for heads such as *able* in section 4.1.3), and assimilate Control to Raising. The latter also argue, as I do, that Raising itself is not to be seen as the movement of a noun-phrase, but arises from a chain involving rather some licensing of the relevant noun-phrase (for Manzini and Roussou, this licensing arises from Aspectual projections).

I have pursued some of the concerns myself. In particular Cormack and Breheny (1994) discuses projection of minor categories, relating to section 1.1.2, and uses covert conjunction as an alternative to the 'adjunction is conjunction' analysis of section 2.1.3; and Cormack and Smith 1994 introduce the notion of a 'nil' semantic θ-role, which obviates the need for an empty specifier position, and further exploit the idea of covert conjunction. Cormack 1996 captures the θ-movement of chapters 2 and 4 using combinators instead of variables and coindexing; but I have yet to find a satisfactory analysis of the problems with *stupid* type adjectives first raised by Postal (1974). Cormack (forthcoming) explores the 'murky' status of specifiers mentioned in section 1.1.2.

I had intended to revise the thesis, but have been prevented by illness; I am grateful to Garland for accepting it in its present form. I would like to thank my two Ph.D. examiners, Deirdre Wilson and Elisabet Engdahl, for stimulating discussion, and for encouragement and advice. and Justin Cormack for proof reading and typographical expertise.

This page intentionally left blank

Acknowledgments

It pleases me to be able to thank those people who have one way or another enabled me to write this thesis. First, this is the appropriate time to thank Martin Davies (of Stirling University) and Dave Bennett (of SOAS), who persuaded me when I was a mathematics teacher that what I was trying to do was linguistics. Secondly, I would like to thank those who helped to turn me into a linguist; in particular Ruth Kempson at SOAS, and Deirdre Wilson and Neil Smith at UCL. Thirdly, I am grateful to the DES for financial support for nearly three years, and to the University of London for allowing time, and to its libraries for supplying books. Fourthly, I would like to thank everyone else: fellow students, visiting scholars, conference acquaintances, friends; linguists (including in particular Robyn Carston and Rita Manzini), psychologists, philosophers—for discussion, stimulation and enlightenment; my colleagues at Reading for making it possible for me to have a sabbatical term; the authors of the books and articles I have read (especially those whom I can no longer identify as the sources of my ideas). Those categories are not mutually exclusive. Finally, I thank Neil Smith, my supervisor. From the heady delight of his first year lectures to the awful illumination of his questions about this work, he has been a most excellent teacher. He has set standards which I fail to reach, but at the same time his encouragement has kept me from despair. I am grateful for his patience, kindness and generosity, and for all he has taught me. It has been a pleasure and a challenge to write for one who combines a limitless intellectual curiosity with a demand for responses which are interesting to questions which are significant; and so I offer Neil this inadequate thesis.

This page intentionally left blank

Definitions

This page intentionally left blank

CHAPTER ONE

Theoretical Preliminaries

> It is in expounding the first principles that we must avoid too much subtlety, for there it would be too disheartening, and useless besides. We cannot prove everything, we cannot define everything, and it will always be necessary to draw on intuition.
>
> Poincaré, *Science and Method*

1.0 Introduction

> I have often had occasion to criticise definitions which I advocate today. These criticisms hold good in their entirety; the definitions can only be provisional, but it is through them that we must advance.
>
> Poincaré, *Science and Method*

It appears that there are such things as definitions. They can be found in dictionaries, and they can be found in text books. They are useful and used: manifestly, these definitions have some special function to do with learning the uses or meanings of words. On the other hand, there is considerable doubt as to the status of definitions. It is said that few or no words are definable; that meanings are not decomposed into parts; that there are not special things properly called definitions, but just ordinary uses of language.

In this thesis, the naive idea of a definition will be pursued, in the expectation that the attempt to show how, syntactically and semantically, there could be such things, will shed some light on the darker areas of doubt, and will test the powers of the explanatory theories we have available.

I had better say straight away that I am not going to say anything about lexicography. That we assess whether one definition is better or worse than another seems to me to dispose of the objection that words are not definable. Perfection seldom is attainable.

At worst, if definitions as such turned out to be illusory, we could regard a so-called definition simply as a source of particularly

concentrated lexical information, and inquire just what syntactic, semantic, encyclopedic, inferential, associative ... information could be conveyed, and how. Is a putative definition in any way different syntactically or semantically from the same string taken as a general statement? Does the apparent special status arise merely from heuristic considerations? What is the gap between a normal interpretation of a string, and what would be needed if it satisfied the formal criteria for a definition—and can this gap be filled by ordinary pragmatic processes? Such questions can be asked, and to some extent answered, whether or not it is agreed that 'there really are' definitions.

The aim is to investigate the function of definitions within a model of language and language processing based more or less on Fodor's 'Language of Thought' hypothesis, Sperber and Wilson's Relevance theory, and Chomsky's Government and Binding theory, with the aid of Model Theoretic Semantics as necessary. The hope is not only to explain what definitions are used for, but to throw some light on aspects of the three theories.

Almost every model of a grammar contains as one component a lexicon. Here, information concerning the use and meaning of a word is stored; each lexical entry contains phonological, syntactic and semantic information. A definition can be regarded as an utterance (written or spoken) designed to facilitate the acquisition or refinement of a lexical entry. If we further suppose that the lexical item must be associated with an appropriate concept, then presumably the definition must facilitate this association, too.

In what follows, I shall be considering on the one hand, the syntax and semantics of the definition itself, and on the other, how the information conveyed by the definition can be put to use in constructing a lexical entry. Both dictionary definitions, such as (i), and written definitions such as (ii) and (iii) from running text, will be considered:

i *perissodactyl* having an odd number of toes

ii A *proper fraction* is a fraction whose numerator is less than its denominator

iii Any number that can be shown as a rectangular pattern of dots is called a *rectangular number*

There are characteristic formats for definitions, and the very fact that an utterance is construed as a definition makes a difference to its interpretation. In particular, an object-language versus meta-language distinction may be imposed, and special inferences drawn.

I assume Chomsky's modular grammar, so a lexical entry for a verb, for instance, should contain information which directly or indirectly gives the subcategorization potential of the verb, and the theta-roles assigned by the verb. Similarly, assuming that the meaning is given in terms of a Fodorian 'Language of Thought', the semantics for the verb must contain information about the roles of semantic arguments. Either a complex representation of meaning or associated 'Meaning Postulates' must indicate the inferential roles associated with the lexical item. The relation between the syntactic and the semantic (language of thought) forms must be mediated at least partly compositionally; the tools of model theoretic semantics, Montague grammar and its extensions, will be used to explore this relation. The simplest hypothesis would be that categorial selection, θ-role assignment, and language-of-thought argument structure were in correspondence, and that the language of thought itself had a simple compositional semantics with respect to some model. Such simplicity unfortunately, but interestingly, is unrealistic. The question then arises as to what constraints there are on possible failures of correspondence; the examination of definitions considered as sources of information about argument structure of the various kinds suggests particular (partial) answers to the question.

In the following sections the various theories just mentioned are briefly discussed, as is the formal theory of definitions. The possible ways in which definitions might be construed using this explanatory apparatus are then set out. Chapters 2 and 3 apply the theories to data; and force some innovation. Chapters 4 and 5 discuss the theoretical consequences of the analyses offered, raise further questions, and suggest possible approaches to problems.

1.1 A little simple syntax

> Such is the primitive conception in all its purity. It only remains to seek in the different case what value should be given to this exponent in order to be able to explain all the facts.
>
> Poincaré, *The value of Science*

1.1.0 Introduction

I shall be assuming a modular, multi-level, 'principles and parameters' theory of universal grammar, more or less a simplified version of what appears in Chomsky's (1986a) *Knowledge of Language*. Aside from the setting of the parameters, it is argued that the knowledge we have of the grammar is innate; what has to be learned is the contents of the lexicon.

The sketch below is not intended to be expository; it is here merely to remind the reader of those parts of the theory which I shall make use of; because my data is relatively simple, much has been omitted and simplified.

1.1.1 Levels of representation

Syntactic structure is primarily a vehicle for representing relationships between elements, where those relationships form the basis for our understanding of the meaning relationships between the semantic reflexes of those elements. On this view, the elaborateness of the grammar is largely a consequence of the elaborateness of the relationships which must, by hypothesis, be coded in the level of LF, in relation to the relative paucity of the relationships encoded directly in the input string. The postulation of S-structure, with its potential for syntactic structure and empty categories, adds information so to speak to the PF string of words. But it would lose this information in an infinitude of ambiguity were it not for the constraints imposed by the modules of the grammar. The grammar can be seen as functional to the extent that the constraints contribute to the construction of LF, on the assumption that it is the structures of LF that bear semantic interpretation.[1]

The most obviously semantic information coded in LF is the distribution of θ-roles (thematic roles) in relation to predicates—which

NP takes agent and which patient role, for instance, in relation to the predicate *kiss*, in *John kissed Mary* or *John was kissed by Mary*. We find that θ-roles are distributed to chains, and that the chains need Case; Case is assigned under government, so the conditions on chains and on government are contributing essentially to interpretation. This is so even of a condition applying just at D-structure, since there are a number of principles ensuring that what is possible at one level is closely related to what is possible at another, in a particular analysis. Move-α itself, the projection principle and the uniformity principle in Case theory are such principles. Those parts of the grammar which are most directly related to semantic interpretation, such as θ-role assignment, are discussed in section 1.1.9 below.

The various modules of the grammar, and other overriding principles, act as well-formedness conditions on the several levels of representation. Each level of representation consists (as far as I am concerned) of lexical items, morphemes and features, and other items of syntactic vocabulary (such as brackets and labels) in some fixed linear order. The levels are connected with each other by transformations. The available transformations are movement (move-α), deletion, and insertion (including the insertion of indices). The process of comprehending a sentence involves constructing four levels of representation, LF, PF, S-structure and D-structure, all of which are as far as possible properly connected and satisfy the various well-formedness conditions applicable to them. This formulation exonerates the language processor from any responsibility to construct intermediate levels, say between D-structure and S-structure. Chomsky (1987) has argued that a proof that D-structure is not just a notational convenience depends on finding a situation where S-structure does not in itself bear all the evidence for D-structure and for a proper relationship between the two levels; an alien element has obliterated some trace, perhaps. Such a situation would require that the grammaticality is ascertained by checking the licensing of S-structure in relation to some intermediate representation, and the licensing of the intermediate representation in relation to D-structure. In this way the ESSENTIALLY transformational nature of 'move-α' between these levels could be demonstrated. For myself, I hope that convincing arguments are not forthcoming, for two reasons. One is that the 'notational convenience' and simplicity arguments that one would be left with if no

such demonstration is forthcoming seem to me to offer possibilities for further exploration of notation and computation; and secondly the temporal ordering of movements presupposed (first you move this to here and then you move that to there) seem to be intuitively at least incompatible with the shift to licensing rather than rules which is otherwise pervasive in the syntactic parts of the grammar. At worst, it might be possible to annotate S-structure further so that the missing information relating to the intermediate representation was present. If the sentence is not fully grammatical, it may all the same be possible to construct four representations such that just one or two conditions have been violated; then it may still be acceptable and understood.

In the familiar inverted T- (or Y-) diagram, S-structure is the central node, to which the other levels are connected. In the direction of PF, we move to realization; I shall have nothing to say about this, and in general will simply assume (inaccurately) that PF gives a surface string, the lexical items of S-structure in the order of S-structure, but without any brackets, empty categories, indices or other ornamentation (if this were correct, PF would be obtained from S-structure using just deletion). LF on the other hand is more informative than S-structure, in that further movement and more indexing has taken place; it is the interface level for semantic interpretation, and so must represent the totality of the syntactic contribution to meaning. D-structure is not an interface level (unless as suggested in Chomsky 1990 it is the interface with the lexicon), but is logically prior to S-structure in that S-structure is transformationally derived from D-structure (perhaps entirely by move-α). Intuitively, D-structure is the level at which grammatical relations are seen in their 'pure' state; but within the theory the levels are individually defined by the well-formedness conditions set on them, and collectively by their proper relation to each other in each particular derivation.

I want to characterize the three grammatical levels rather generally, before discussing in a little more detail some of the modules and principles of the grammar.

D-structure is the simplest level, because the totality of symbols available for forming strings at this level does not include indices. The requirements of some of the modules (binding and bounding) are then vacuously met. The projection principle and perhaps θ-theory apply to all the syntactic levels. X-bar theory is usually considered to be

proprietary to D-structure. I think that the conditions indicated define D-structure. Movement is plausibly considered to be not assumed (or used) unless forced—this would clearly reduce computational options usefully, if translated into a processing model. From D-structure to S-structure, movement is characteristically forced by the imposition at S-structure of Case theory (accounting, as is familiar, for passive and raising structures *i.e.* NP-movement), and for English, a '*wh*-scope requirement' (accounting for the fronting of non-exempt *wh*-phrases; to be discussed below).[2] Other movement is forced by the inability of various features or morphemes to stand alone without a lexical host (accounting for anything like affix-hopping or V-movement). I think it is these conditions which identify S-structure, provided we assume there is only forced movement. The problem is to characterize S-structure rigorously, given the existence of further movement at LF. The movement leading to LF (which I shall assume is restricted to QR) is forced by the requirement that every QNP (quantified noun phrase) binds a syntactic variable;[3] for languages like Japanese without the *wh*-scope requirement on S-structure, *wh*-movement would be forced at LF by this condition. If we do not assume that movement is forced, then something must actually prevent QR from taking place at S-structure. Possibly the relevant restriction is that landing sites for S-structure movement must be base-generated (not produced by adjunction).

Other modules impose well-formedness conditions on S-structure and LF: θ-theory and the projection principle already mentioned, and those modules referring particularly to indexing relations (bounding theory and binding theory). Chomsky argues that there are no parameters to be fixed for LF. We may expect that S-structure and LF share most of their properties, especially if what S-structure is is parametrized, where the relevant parameters serve to single out some particular level between D-structure and LF. Chomsky (1990) suggests that S-structure may be considered as the solution to the equations imposed by the three peripheral (interface) levels. Taken definitively, this would lead us to postulate that any conditions apparently applying just to S-structure within syntax must in fact belong to PF—Case theory, perhaps, then.

The description in terms of well-formedness conditions suggests a rather negative existence for grammatical strings: they are what falls through if everything possible has been dropped through a set of sieves.

Rather more positively, we can conceive of structures as consisting of various elements (sub-structures) each of which must be *licensed*. We might expect and require that every element should be licensed with respect to each module. Alternatively, a single structure is seen differently by each module, so that which elements are visible varies according to the module. For instance X-bar theory may be blind to all except brackets and category labels: lexical content might not be visible to this module. The notion of *visibility* has been introduced and used so far just for θ-role assignment: a chain is not visible for θ-role assignment unless it has Case. In effect, the chains that the θ-module can see are not identical to the chains that say the binding module can see. In particular, what is visible for θ-theory is conditional on the dispositions of the Case-module.[4] I think that the notion of visibility may be usefully extended further.

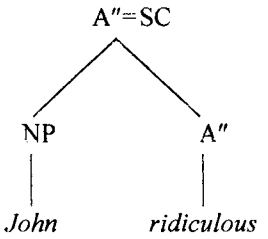
1.1.2 Categorial Structure

X-bar theory is a general theory of the internal construction of maximal projections, or if you like, a theory of the categories related to a head. Assuming a three level analysis, we have licensed COMPLEMENTS, which are sisters of X-zero under X′; and a SPECIFIER, which is a sister of X′ under X″. Complements and specifiers—and indeed any other independently introduced categories—must be maximal projections. (I shall frequently refer to complements as arguments, whether or not they are NPs, since the semantic role of a complement is to be an argument.)

It is clear that we need as well some theory of the distribution of maximal projections, unless this is all. And this is not all—there are missing structures for predication, and structures for modification, at least. Both these can be accommodated as adjunction structures; so we need a general theory of adjunction. For instance, predication in small clauses may be analyzed like this (following Chomsky 1986b):

1 i I consider [$_{SC}$ John ridiculous]

ii

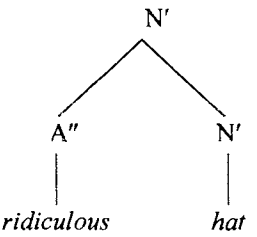


Here, the NP *John* is adjoined to the left of the A″ *ridiculous*, and this sisterhood permits the predicate *ridiculous* to be predicated of its subject (external argument) *John*.

A typical modification structure is given for (2i) by (2ii):

2 i [her [$_{N'}$ ridiculous hat]]

ii

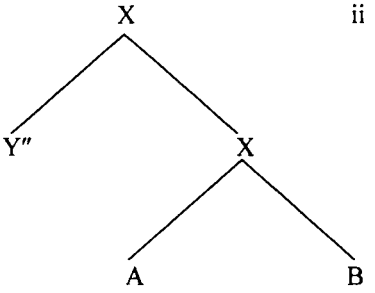


Assuming that these two structures are correct, we may take it that base-generated adjunction to at least X″ and X′ level is possible. Let us assume that base-generated adjunction of a Y″ category is possible to a category of any level, including X-zero, for the moment.

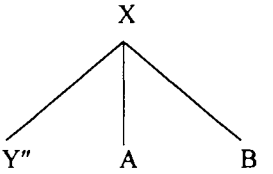
Base-generated adjunction, and the subject-status of the NP in (1) might suggest that all subjects should be found in adjoined rather than specifier positions; Koopman and Sportiche (1985) and Manzini (1989) argue that subjects are indeed adjoined to VP, but only for some languages. These subjects will be moved to the pre-I position. The status of specifiers in general seems to me to be rather murky (as it is bound to, to someone with sympathy for extended categorial grammar), but for the most part I shall take what is offered by the standard theory. In chapter 4 I make some particular suggestions regarding specifiers.

It is important to note that an adjunction structure can be seen by any particular module with more or less structure (see section 1.1.6). That is, given (i), for instance, a module may effectively 'see' rather (ii):

3 i



ii



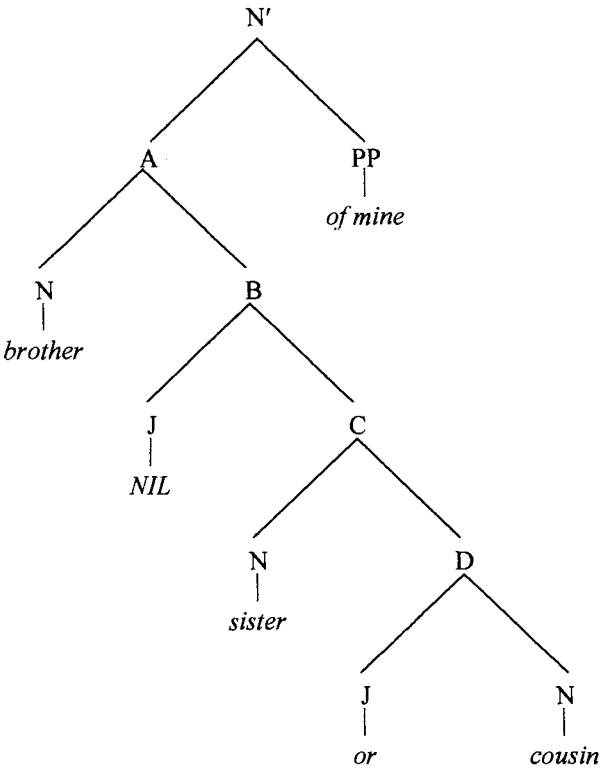
Intuitively, in (ii), the two instances of 'X', which are identical in (i), are actually identified.

We have not yet accommodated coordination and subordination. The arguments for constituent coordination seem to me to be overwhelming, now that we have semantics for this. (Non-constituent coordination is another matter, but fortunately it does not arise here.) Gazdar *et al.* (1985) argue for a combination of binary branching and multiple flat structures (see also Sag *et al.* 1985). They provide for examples like *no [man, woman or child]*. This would usually be considered to involve N or N′ conjunction, and hence necessarily non-maximal categories. However on the DP analysis, what was an N′ becomes an N″. But we may have instead examples such as

4 i no [[brother, sister, or cousin] of mine]

Here the conjuncts must be non-maximal, if we take the *of mine* to be the complement of the conjuncts. Consider then the structure:

4 ii



Gazdar *et al.* would have a flatter structure, with A = B = C, here. It is clear that this cannot be construed as falling under X′ theory, as it stands, since the N nodes are non-maximal, nor under an adjunction theory where adjuncts must be maximal projections. The whole functions as an N, so for most purposes certainly, that is what A should be. On the other hand, so far as argument structure in the semantic sense is concerned, one might suppose that A was a projection of a head 'NIL', and C of the head *or* (both of category J for conJunction). We would suppose then that the whole was a J″. This conflict can be dealt with neatly in GPSG using feature-transmission. The feature system of GPSG is further developed than that of GB theory, and there is probably more than one line that might be pursued in adapting the GPSG analysis to a GB theory, assuming it to be correct. I shall make

some tentative suggestions, which could be extended by adapting more of the GPSG solution.

If A, B, C and D are all N, we would have a non-standard adjunction structure; suppose this is licensed by the nonstandard 'heads' NIL and *or* which are of a defective category J. This category is so defective that it is relegated to the status of feature rather than category in the presence of any 'proper' category. Thus, for instance, node A is [N, J″], where the category visible out of the two is for almost all purposes the N. We have to allow this defective category to take "arguments" which are not maximal projections. Thus B is [N, J′], C is [N, J″], and D is [N, J′]. Let us call a defective projection like that of J a 'quasi-projection', with J the quasi-head. We may note that by category, A = B = C = D, so that the structure is capable of functioning as if it were a flat structure.

It is possible to conjoin items of distinct syntactic categories, and GPSG makes provision for this using an 'underspecified category'(see Sag *et al.* 1985). The need for this shows up in definitions like

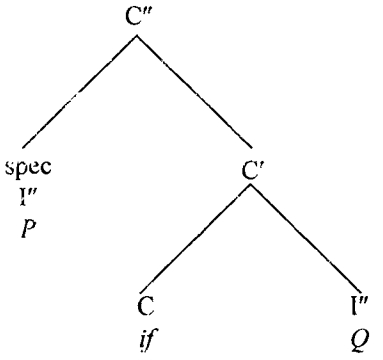
5 *phlogistic* of, like, or containing phlogiston

where we have categories P, A and V conjoined.

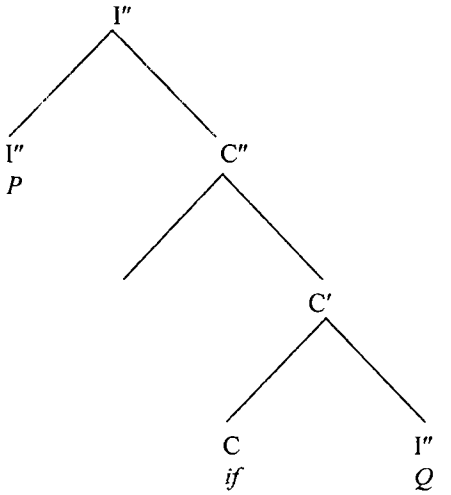
It may be that there are other 'minor' categories which would yield to being treated as quasi-projections. Some possibilities are discussed in later sections. We would expect them all to be semantically operators, given their yielding to full categories.

Finally, we need a structure for subordination, and in particular for connectives like *if*. There are good grounds for supposing that the connective is not like the conjunctions just discussed. There are various alternatives within what has been provided already and some of these are discussed in relation to definitions in chapter 3. For a sentence consisting of "P if Q", at least the following structures would appear to be licensed under the assumptions of X′ theory and adjunction theory:

6 (*if* as head of C-projection, I″ in SPEC of CP)



7 (adjunction of C″ to I″)



In chapter 3, section 3.3.1, I shall argue that (7) is the structure for subordination in the sentences which will concern us.

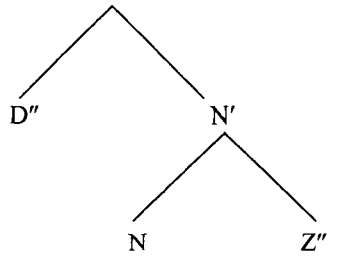
Chomsky (1986a, 161) states that the level at which X-bar theory applies is D-structure, because movement may produce structures which do not conform to the X-bar schemata. He also states that D-structures do conform to the X-bar theory. I have suggested above

that we need base-generated adjunction (and the quasi-projections for coordination). If we require that D-structures are licensed by one of these schemata, then we will find that S-structure and LF will fall under the same schemata, so that something like Emonds' (1976) structure preserving principle is again operating. Let us suppose then that structures at all levels are licensed by the X-bar theory with the extensions indicated above.

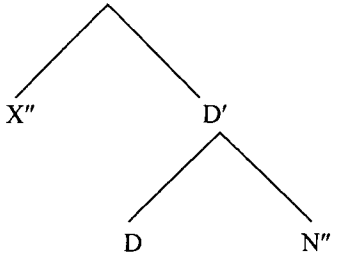
1.1.3 Categories and features

As can be seen above, I am assuming that we have the full projections of the categories for I and C, as put forward in Chomsky 1986b. However, for expository convenience I shall sometimes refer to the I″ as VP (but not as V″) in the old style. There are the other major categories: A, V, P and N. With respect to this last, there has been a recent suggestion within GB theory (Abney 1987) that the head of a noun-phrase should be the determiner, so that we have a determiner-phrase, D″, as the top node, with D taking N″ as its complement, instead of N″ as top node with (presumably) D″ in specifier position.

8 i



ii



Note that if D″ in (i) and N″ in (ii) are fully expanded, the two trees produce directly the same string of potential categories *viz* (iii):

8iii X″ D Y″ N Z″

Of course, these are bracketed differently under the two hypotheses. Until it becomes relevant, mainly in chapter 4, I shall use the more familiar NP structures of (i), referring to the maximal projection as NP rather than as N″.

I have suggested a minor quasi-projection J in the previous section, for conjunctions. There are a number of categories such as those needed for the various traditional adverb categories that I have not discussed at all.

The system of features assumed within GB theory is informal. I shall assume anything I need to assume, borrowing from GPSG (see Gazdar *et al.* 1985). For example, it is commonly taken that some verbs such as *ask* may admit or require an embedded question as complement. If we take it that this is stipulated by assigning a feature to the subcategorization specification, then the complement may be CP[+WH]. This feature may be realized legitimately on either the specifier of CP (as a *wh*-phrase), or on the head of CP (as either a *wh*-complement such as *if*, or for root clauses by the presence of a V[+AUX] in C). I shall also assume (following GPSG) that lexical items may be used as features on complements, and that this requires that such an item appears in the SPEC or Head of the complement. For example, the adjective *opposite* may have a complement NP[P[*to*]]. Since *to* itself requires an NP argument, the structure induced is a PP, [*to* NP]; I assume that the whole of the structure associated with the feature *to* is invisible for the purposes of checking the gross subcategorization for an NP.

1.1.4 Movement

Movement is characterized by 'move-α', a shorthand for the idea that any category α may be moved—in principle, to any position. In fact, movement seems to be restricted to zero-level and maximal projections. X-zero movement must be 'head to head' movement; little use will be made of this in what follows.[5] Maximal projections, which in the most studied cases are NPs, move in three distinct ways. There is 'NP movement', movement to an A-position; and A-bar movement. A-bar movement is either to an existing but empty slot designated for maximal projections, such as specifier of C″; or it is adjunction to some maximal projection. A-positions are defined as positions to which a theta role could in principle be assigned (were the relevant lexical item suited to doing so) (Chomsky 1986a, 80; this has always seemed to me to be a somewhat odd definition—there is some discussion in my chapter 4). What is accounted for are the structures of passive, and

raising. Movement to the specifier slot of the C-projection has replaced movement to COMP as the characterization of *wh*-movement. Adjunction is involved in QR, the movement of quantified noun-phrases out of A-positions at LF. As noted above, none of this movement produces structures of a kind distinct from those needed at D-structure.

1.1.5 Indices

Further structure is given to representations by indices. The main indexing which will concern us is the coindexing of NPs under anaphoric dependency and the indexing consequent on movement.

The former is brought into being by the free indexing of every NP impossible coindexings are ruled out by the binding theory and (possibly outside the grammar) by inappropriate effects such as gender mismatch. Direct dependency (loosely referred to as coreference) gives rise to 'bound variable' interpretations of pronouns, for instance. Less direct dependency can account for the status of indefinite NPs in 'donkey sentences' (discussed section 3.5.1), and possibly for *wh*-NPs left *in situ* in multiple *wh*-questions.

The latter is an automatic concomitant of movement (or is definitive of it): a moved element leaves behind it a coindexed *trace*. The trace is usually an empty element, but possibly it may be realized as a resumptive pronoun. An element in an A-position and its traces form a chain, with (usually) a Case marked head and a θ-marked terminal position. These chains are said to form abstract representations of their heads (Chomsky 1986a, 96). An element that moves to an A-bar position changes its status: it becomes an operator (a binder) rather than an argument. The operator and its traces form an A-bar-chain; the term chain is used either just for A-chains, or for both kinds; and quite often to mean maximal chain in some obvious sense, though a subpart of a chain is a chain too (Chomsky 1986a, 96–98).

1.1.6 Structural relations

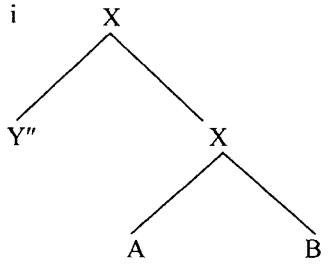
There are a number of structural relations which hold between a particular element and some domain or set of elements: the relation may hold jointly by virtue of properties of the particular element and

the structural configuration or other properties of its neighborhood. One such relation is that of domination. They are usually seen as defined relations—as part of the way we can describe things. If we were to include the notions of barriers (Chomsky 1986b) here, then we would be giving part of what otherwise falls under a module—the bounding module, mostly. Since most of what I talk about is fairly simple syntactically, I shall keep to the earlier formulations, and not use the apparatus of barriers.

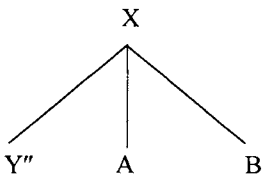
The simplest domain is the c-command domain of an element. C-command (from 'Constituent-command') was introduced by Reinhart (1976) in the course of accounting for the possibility of the 'bound variable' reading of pronouns (as in *Everyone lost his temper*). For an element, say α, this is defined to be the set of elements dominated by the first branching node above α, but (usually) excluding those elements dominated by α. In terms of bracketed strings, and ignoring the possibility of non-branching nodes, in the string $[_\gamma[_\alpha \ldots] [_\beta \ldots]]$, the node α c-commands β and everything in it; and the node β c-commands α and everything in it. It is thus closely related to the standard notion of logical scope: if α were an operator, then β would be its logical scope, and *vice versa*.[6] If what is dominated by α is NOT excluded, then in the string given, the 'domain' of α (or β) will be γ.

A variant of c-command is 'm-command' ('m' for 'maximal'). The definition (Chomsky 1986b, 8) follows that of c-command except that "first branching node above α" is replaced by "first maximal projection dominating α." However, matters are confused by the fact that the definition of 'dominate' may also be modified. The modifications affect only adjunction structures, where it appears that different modules of the grammar need distinct versions of 'dominate'. In an adjunction structure like (9i) below, the two nodes labeled 'X' are considered to be two 'segments' of a single category X.

9 i



ii



Following suggestions by May (see May 1985), Chomsky (1986b, 7) defines 'dominate' so that "α is dominated by β only if it is dominated by every segment of β." Under this definition, Y″ is not dominated by X in (9i), whereas A is. The category X is asymmetric with respect to these two. The relation for the other pattern is given by 'exclusion' (ibid., 9): "α excludes β if no segment of α dominates β." Thus in (9i), Y″ does 'not exclude' A, and A also does 'not exclude' Y″. Hence the relation between Y″ and A is identical with respect to the category X, and the effect is just as if the structure were as in (9ii)—the relation of 'non-exclusion' effectively flattens adjunction structures. May uses m-command and the modification of domination in his characterization of the logical scope potential of NPs in LF structures. This is discussed in section 3.5.1.

There are two other important relations: those of government and those of binding. The relation of government enters into the licensing conditions of more than one module. Government is a relation between a category and a set of other elements defined partly in terms of m-command. I follow here the definition of Chomsky (1986a, 162, but with the terminology of m-command). The category must be either a lexical category or one of their projections (N, V, A, P, NP, VP are listed by Chomsky) or the element AGR of INFL. One of these elements, a category α say, governs a maximal projection X″ if and only if α and X″ m-command each other. Then by definition α also governs the specifier and head of X″. The effect will be for instance that heads govern their complements, that a subject is governed by AGR, that when an NP is governed then so is its determiner in the specifier position (or the SPEC of DP).

Binding is an asymmetric relation between two categories, rather than between an element and a domain. An element α binds an element β if and only if the two are coindexed and α c-commands β. An element which is not bound is free; an element is bound in a particular domain if it has a binder in that domain.

1.1.7 Binding, bounding, control and the ECP

Binding theory, bounding theory, control theory and the ECP all impose well-formedness conditions on structures with indices, and *a fortiori*, traces.

Binding theory is the module dealing with the permitted relations between an NP binder and an NP bindee where the latter is in some sense anaphorically dependent on the former. It thus imposes well-formedness conditions on chains, on operator binding of a variable, on the interpretation of pronouns and anaphoric expressions. The theory is sensitive to the status of the bindee. In Chomsky (1986a, 166ff.) the three principles of the Binding Theory are given as:

(A) An anaphor is bound in a local domain
(B) A pronominal is free in a local domain
(C) An r-expression is free (in the domain of the head of its chain)

Anaphors include reflexive and reciprocal pronouns; pronominals include other pronouns. An approximation to the required local domains for (A) and (B) is the minimal governing category for the bindee in question. A governing category for α is a maximal projection containing both a subject and a lexical governor for α. The notion of r-expression is clear at the center but not always clear elsewhere. The central cases of r-expressions are referential expressions like *John*, *the child*, and variables (a trace bound by an operator). If all the conditions apply just to LF, then any quantified NP such as *every rabbit* will have been moved out of its A-position, (becoming an operator), so just these two kinds of r-expression will arise. If expletives are deleted at LF, then all overt NPs will be accounted for. Of the empty categories, *wh*-trace is a variable, and an r-expression, as noted. NP-trace is an anaphor; PRO is standardly both an anaphor and a pronominal, from which it follows that it cannot be governed.

Control theory deals with the circumstances under which PRO (the empty subject of gerundive and infinitival clauses) may be anaphorically dependent on some antecedent. I shall simply take it that there is such a theory; semantic interpretation makes it clear when there is anaphoric dependency, since otherwise we have the 'arbitrary' interpretation as in *It is forbidden* [PRO *to go there*].

Bounding theory puts restrictions on the kind or number of category boundaries which can intervene between a binder and a bindee. There are a number of principles enunciated, but they will not be discussed here since I have little (probably nothing) to say concerning this part of the theory. Bounding theory is vacuously

satisfied at D-structure, must apply to S-structure, and is possibly supposed to be met at LF.

The ECP is described by Chomsky (1986a) as imposing "certain narrow identification conditions" on empty categories, specifically traces. In its simplest formulation it states that an empty category must be 'properly governed', where proper government means θ-government or antecedent-government (Chomsky 1986b). Antecedent government is simply government by a coindexed antecedent, so that a trace can fail to be antecedent governed only if there is some barrier to government of the trace by its nearest antecedent. But in this situation, the trace might instead be θ-governed, that is, governed by an element that θ-marks it; because it is a complement for instance. It is subjects and adjuncts that may fail to leave properly governed traces, then.

1.1.8 Case

Case is an abstract property assigned to NPs (and perhaps to SPEC and head inside); *Case theory* deals with how it is assigned. "If the category *a* has a Case to assign, then it may assign it to an element that it governs" (Chomsky 1986a, 187). Inherent Case is assigned at D-structure and realized at S-structure by morphological Case marking (residual in English) or alternatively at least for possessive Case by an empty preposition (*of*). Items of the lexical categories N, A, and P, as well as V, are (sometimes) capable of assigning Case, with for instance *picture* assigning genitive Case marked by *of* to *John* in [*picture* [*of John*]]. Inherent Case assignment by a category α can only be to an element θ-marked by α (at D-structure); Case realization can only be on an element governed by α (at S-structure). Structural Case is assigned by V, and by INFL if it has the feature [+AGR], and it is assigned at S-structure, with a requirement of adjacency (in English). It is not essentially associated with θ-marking, but as with inherent Case, it can only be assigned and realized under government. Nominative Case is assigned by INFL; V assigns objective (accusative) Case, normally to an adjacent complement NP. Exceptional Case marking, as in *John believes* [*her to be clever*], is accounted for, provided *believe* governs *her*, by the fact that *believe* DOES have objective Case to assign, as can be seen in [*John believes that fact*].

The condition that any lexical NP has to have Case, the 'Case filter' has been replaced by a condition on chains, which stipulates that a chain is not visible for θ-marking unless it has Case. The chain has Case if it has a Case-marked element in it—which is usually the head. This reformulation has the important consequence that there is no requirement on a non-argument NP to have Case, since it does not get θ-marked (morphologically, it must appear with a default Case or a Case acquired by agreement with the NP to which it is adjoined) (see Chomsky 1986a, 95). It appears that argument CP chains also need Case.

The visibility condition must clearly apply just to S-structure, or anyway not to D-structure, since no subject will have Case at D-structure. There is a uniformity condition, too, which is to apply at S-structure, although its function is to relate D-structure and S-structure inherent Case marking. This stipulates that "If α is an inherent Case-marker, then α Case-marks NP if and only if α θ-marks the chain headed by NP" (Chomsky 1986a, 194). The Case-marking in question is Case realization derived from Case-assignment at D-structure, and θ-marking depends on there being a θ-position in the chain, where this position is that assigned a θ-role at D-structure.

1.1.9 Semantic structure

The semantic component of the grammar (as opposed to semantics outside the grammar) deals with the recognition and distribution of such categories as predicate and argument, operator and variable, adjunct and modifier. It has to be related to the structures and categories already set up.

Let us start with the notions of predicate and argument. These are intended to correspond at least approximately to those same notions from predicate calculus. Heads, as defined within X-bar theory, are predicates in some sense, and they take as internal arguments their complements. A head together with its complements forms a maximal projection and this may itself be a predicate if it (still) has a θ-role to assign.[7] For example, a preposition usually requires an internal NP argument, and a prepositional phrase like *in the garden* is capable of being predicated of an NP such as *John* in an intuitively obvious way. The capacity of a head to take or require arguments is given in the

lexicon, by means of a specification of what θ-roles the head disposes, and whether they are internal or external. Note that the current syntactic terminology usually uses 'predicate' only for maximal projections.

We may assume that the external role is specified compositionally so that it does in fact get assigned by the maximal projection of the head, though because the relation of head to external argument is normally transparent, we speak of it loosely as an external argument of the head. The problematic cases concern idioms, for instance, where the actual role is not necessarily predictable from the head, although the fact of θ-marking capacity is so predictable (consider *kick the bucket* as compared with *kick the doorstep*). There are also somewhat exceptional cases where an NP, which is normally an argument rather than a predicate, can function as a predicate, as in *I think* [*John a fool*], which is similar in meaning and structure to *I think* [*John foolish*]. This requires special provision, semantically, and also syntactically in order to account for the fact that the argument NP (*John*) receives a θ-role.

What counts as an argument depends on what level of representation we are talking about. I will confine the discussion to NP arguments for the moment. At D-structure, any NP except an expletive may count as an argument. At S-structure and at LF, it is an A-chain that is an argument, where the chain is an abstract representation of the NP at its head. Trivially, we could see the NPs of D-structure as one-member chains. NPs in A-bar position are excluded from being arguments, then, but a chain headed by the trace bound by such an NP is an argument—the trace is a variable, and an argument. Since expletives are not subject either to *wh*-movement or to QR, there will not be any problem here. At LF, it is only chains headed by simple 'referential' expressions that are arguments—this is one way of forcing QR for quantified NPs. The permitted simple expressions will be definite NPs like *the rabbit* or *John*, and variables.

The well-formedness condition connecting θ-roles (from the lexicon) and arguments (as they appear in a structure) is the θ-criterion. A position to which a θ-role is assigned is a θ-position. The θ-criterion states that a chain has at least one and at most one θ-position (Chomsky 1986a, 133, 135), (where as mentioned above, the arguments are chains, including the one-member chains of D-structure). The θ-criterion must hold at LF (but see the Projection Principle below), and perhaps holds at other levels. The formulation above permits a

single NP to receive two θ-roles at the single θ-position (Chomsky 1986a, 97). Consider

10 John left the room angry

John is the argument of two separate predications, one by the VP *left the room*, and one by *angry*, by each of which a θ-role is assigned, but there is a single θ-position involved. There is a uniformity condition stipulated on theta marking: if an element is θ-marked by any potential θ-marker, then it is θ-marked by every potential θ-marker (Chomsky 1986a, 97). This is to rule out structures like

11 *John seems [that it is raining] angry

The Projection Principle states that the subcategorization properties of a lexical item must be projected to every syntactic level (Chomsky 1981, 31). There is also a rule connecting subcategorized complements and θ-roles: subcategorization entails θ-marking. Consequently, θ-roles associated with internal arguments will always be realized, whereas on this formulation of the Principle, those relating to external arguments need not be. Stronger versions of the Principle have been proposed: we may for instance require that all θ-roles are projected.

Non-head predicates are constrained to find an argument through a general statement to the effect that predicates must have arguments. In this context, if a phrase IS an argument, it is absolved, even though it might be capable of being a predicate.[8] For instance, it is possible in principle for a verb to take a VP argument (Chomsky 1982, 18 footnote 13), although VPs are normally predicates. Conversely, every argument must be the argument of some predicate (head or otherwise). If predicates have arguments, then even though subjects are in principle optional, a non-argument VP will have to have an external argument if it has an external θ-role to assign. But the grammar requires all clauses to have subjects—that is, clausal non-argument VPs have syntactic subjects, even if these are empty or expletive *i.e.* are not arguments). This requirement is usually tacked onto the projection principle, with the two together being called the EPP (extended projection principle).

The fact that variables are arguments means that the construal of a clause with a variable in it is in the Tarskian style, rather than in the Russellian 'open sentence' style where variables are merely 'place-holders'. Semantically, such a clause is a proper sentence. The necessity