Combinatory Categorial Grammar

Combinatory Categorial Grammar
An Introduction

Mark Steedman

DRAFT 2.0 June 16, 2022

The SOMESUCH Press
Philadelphia, PA
Edinburgh, UK

This book was set in Times Roman by the author using LaTeX 2ε, and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Steedman, Mark.
Combinatory Categorial Grammar: An Introduction / Mark Steedman
p. cm.
Includes bibliographical references and index.
ISBN 000-0-000-00000-0 (alk. paper)
Grammar, Comparative and general—Syntax. 2. Grammar, Comparative and general—Semantics. I. Title.
P299.Q3S74 2012
415′.7—dc23
2017000000

10 9 8 7 6 5 4 3 2 1

Contents

Preface xv

Chapter 1
Introduction 1

1.1 The Problem of Language Diversity 1

1.2 The Problem of Child Language Acquisition 4

1.3 The Problem of Ambiguity 10

1.4 The Problem of Discontinuous Constituents 11

1.5 Some Complex Discontinuities 15

1.6 Pronominal Anaphora and Coreference 20

1.7 Some Distracting Anaphoric Coordinate Constructions 20

1.8 Explaining Discontinuity 25

Part I
Categories, Combinators, and Case

Chapter 2
Categorial Grammar 31

2.1 The Categorial Lexicon 31

2.2 Combining Categories I: Application 33

2.3 CCG and the Minimalist Program 43

Chapter 3
Case 47

3.1 Tense Morphology and Agreement 47

3.2 Case and Morpho-Lexical Type-raising 49

3.3 Modal Case 53

3.4 Case and Case Agreement 55

3.5 Stacked Case 56

Chapter 4
Composition 59

4.1 Combining Categories II: Composition 60

4.2 Discussion 66

Chapter 5
Word-Order 69

5.1 Scrambling in Japanese 69

5.2 Germanic scrambling and crossing dependencies 72

5.3 Discussion 79

Chapter 6
Intonation Structure 81

6.1 The Autosegmental-Metrical account of English Intonation Structure 83

6.2 Intonation and Implicature 87

6.3 Intonation Structure in Child Language Acquisition 91

6.4 Making CCG derivation Specify Intonation Contour 92

6.5 Intonation and "*In situ*" *wh* and Topicalization" in English 94

6.6 Discussion 96

Part II
The Basic Constructions

Chapter 7
The Lexicalized Constructions 101

7.1 Subject-Auxiliary Inversion 101

7.2 *Neg*-placement 103

7.3 The Passive 104

7.4 Raising Etc. 105

7.5 Light Verbs 118

7.6 Reflexive Binding 119

7.7 Discussion 122

Chapter 8

Are All Constructions Lexically Governed? 125

8.1 Dative alternation 125

8.2 The "Unaccusative" alternation 126

8.3 The Causative Alternation 128

8.4 The Resultative alternation 129

8.5 Some More Idiosyncratic Constructions 131

8.6 Discussion 136

Chapter 9

***Wh*-Constructions** 141

9.1 *Wh*-questions 143

9.2 Indirect questions 146

9.3 Topicalization 146

9.4 Relativization 148

9.5 Bare relatives 152

9.6 Embedded Subject Extraction 155

9.7 Island Constraints 162

9.8 Preposition Stranding and Relativization 165

9.9 On "Remnant Movement" 168

9.10 *Tough*-movement 170

9.11 Pied-Piping and *Wh-In-Situ* 173

9.12 Celtic relativization 175

9.13 Discussion 179

Chapter 10

Parasitic *Wh*-dependencies 181

10.1 Combining Categories III: Substitution 181

10.2 Discussion 189

Chapter 11

Symmetry and Asymmetry in Left and Right Extraction 191

11.1 Across-the-Board Extraction 192

11.2 Right-edge restrictions 196

11.3 Asymmetric islands 198

11.4 Ross's Argument/Adjunct Cluster Coordination Asymmetry 199

11.5 Preposition Stranding and Coordination 203

11.6 Multiple and Correlative Coordination 206

11.7 Gapping and Stripping 209

11.8 On the distinction between Across-the-Board and Parasitic Extraction 210

11.9 Discussion 212

Chapter 12

Combinatory Minimalism 215

12.1 Binding and Coreference 217

12.2 Grammar without Action-at-a-Distance 219

12.3 The Place of CCG in an Extended Language Hierarchy 220

12.4 Descriptive adequacy of CCG 224

12.5 Explanatory Adequacy of CCG 227

12.6 On the Notion “Possible Lexical Category” 228

12.7 Envoi 235

Part III

Semantics and Anaphora—See 2018 md. for changes

Chapter 13

Quantifier Scope 239

13.1 The Universals 240

13.2 Eliminating the Existentials using Generalized Skolem Terms 241

13.3 Coordination and Quantifer Scope Alternation 244

13.4 Embedded Subjects and Intermediate Scope 246

13.5 Distributivity 247

13.6 Other uses of Skolem indefinites 249

13.7 Generics 255

13.8 *Same* and *different* 256

13.9 Discussion 257

Chapter 14
Anaphora 259

14.1 Pronominal Anaphora 259

14.2 Donkey pronouns 260

14.3 Logophoric Anaphora 261

14.4 Null Anaphora 263

14.5 Discussion 264

Chapter 15
Conclusions 265

Part IV
Appendices: CCG in Application

Appendix A
The Evolutionary Emergence of Language 269

A.1 Views of Language Evolution 269

A.2 Plans and the Structure of Mind 272

A.3 Conclusion 278

Appendix B
Child Language Acquisition 281

B.1 Semantic Bootstrapping 284

B.2 The Proposal 285

B.3 Syntactic Bootstrapping 289

B.4 The experiment (Abend, Kwiatkowski, Smith, Goldwater, and Steedman, 2017) 292

B.5 Against Parameters 295

B.6 A More Realistic Lexicon 297

B.7 Smoothing and Generalization 300

B.8 Limiting Overgeneralization 301

B.9 Grounding Semantics in Use and Interaction 303

B.10 Conclusion 304

Appendix C
Natural Language Processing 307

C.1 The Grammar 307

C.2 The Algorithm 310

C.3 The Oracle 311

C.4 Discussion 315

Appendix D
Towards a Form-Independent Semantics 317

D.1 Decompositional Lexical Semantics 318

D.2 Semantic Primitives as Hidden Variables 319

D.3 Temporal Semantics 323

Appendix E
The Devil's Dictionary 329

References 333
Index 387

Preface

This book is intended as an introduction to Combinatory Categorial Grammar (hereafter, CCG) as a linguistic theory. CCG is a theory of grammar that has been devised with the aim of keeping syntax and semantics as simple and as closely linked as possible, consistent with truth to the linguistic facts. It has also been developed with the secondary aim of keeping the theory as close as possible to the psychological and practical computational mechanisms that map sentences into meanings, and meanings into sentences, and that enable the child or computer program to learn the grammar for any human language from exposure to meaningful sentences in contexts that it understands.

In order to achieve these aims, as in any scientific endeavour, it is important to keep the degrees of freedom in the theory as few as possible in comparison to the degrees of freedom in the data that we seek to explain. That is to say that we need to make our theory expressive enough to capture the patterns of relations between sound and meaning that are exhibited in the languages of the world, but not so expressive as to also be able to capture patterns that we have good emprical reasons to believe will never occur.

Since nobody actually knows for certain what the totality of those patterns actually is, CCG has necessarily been developed inductively, over a period of many years, in interaction anmong linguists, psychologists, and computer scientists. Over that period, the theory has, like other contemporary linguistic theories, gone through a number of changes in notation, from the earliest, developed with Tony Ades and Anna Szabolcsi in the early 1980s, in which lexical entries for categories like verbs and determiners expressed valency but not directionality, and in which the basic inventory of rule-types that still applies was developed, through a hybrid notation in which lexical categories captured both valency and language-specific directionality, but in which the rules still bore language-specific restrictions, to the present version, in which, following Mark Hepple and Jason Baldridge, among others, all language-specific information lies in the lexical entries, and the combinatory rules are free and universal across all languages.

The major linguistic results of the theory concern unbounded dependencies such as those found in relative clauses, the fragmentary constituents that arise under coordination, the similar fragments that can appear as intonational phrases, and the interaction of scope of quantification and negation with all of the above constructions. Its gradual evolution means that these results have been presented over the years in various notations which, while underlyingly similar, are sufficiently different as to possibly cause confusion.

It therefore seems time to recast the original results, together with some new ones that have emerged in the process, in a single unified notation, and in a simpler, sometimes less detailed presentation, specifically intended for linguists working in other frameworks, and for computer scientists and psycholinguists interested in natural language processing who are encountering theoretical linguistics for the first time.

The book is specifically intended for undergraduate and graduate students at an early stage in their linguistic studies. Since our guiding principle is that syntax must be developed in tandem with semantics, it might best be be thought of as a course on the Syntax-Semantic Interface, following either an introduction to syntax (in any framework) to at least the level of identifying the major parts-of-speech and constructions, or an introduction to logic or formal semantics to at least the level of the predicate calculus.

However, the book is also intended to act as a free-standing introductory text, and to that end, I have included some extremely gentle exercises, and a glossary of basic linguistic terms, which I hope to offend no-one by following Ambrose Bierce in calling "The Devil's Dictionary", since some of the terms therein are used in the literature in confusing and contradictory ways, which I have tried to reconcile.

The book is divided into four parts. After an introductory chapter defining the problem, Part I, "Categories, Combinators, and Case", consists of chapter 2, concerning pure categorial grammar (CG) and the lexicon, chapter 3, concerning the central role played in the theory by case in the form of type-raising, chapter 4 concerning the fundamental role of fuction composition in coordination, chapter 5, concerning the nature of word-order variation in configurational languages, and chapter 6, on the relation of intonation structure to syntactic derivation. Part I could be used for a short course on the basics of CCG. Part II consists of chapters 7 and 8 on the lexically headed constructions, chapters 9 and 10 on the *Wh*-constructions, and chapter 11 on symmetry and asymmetry between *Wh*-extraction and right node-raising and other rightward extractions. Chapter 12 draws comparisons with analyses in other lingustic frameworks, particularly G/HPSG, TAG, LFG, and the Chomskyan Minimalist Program. Parts I and II, perhaps skipping chapters 8 and 10, constitute a self-contained introductory course on syntactic aspects of CCG that could be covered in a term or a semester. To that end, brief exercises have been included throughout

Part III, "Semantics and Anaphora" is more demanding in terms of the formal semantics involved. It consists of chapter 13, on anaphora and coreference, and chapter 14, on quantification. Chapter 15 is a brief conclusion to the whole thing. These chapters could be omitted from an introductory course, or could be included as a basis for a one semester more advanced seminar.

Part IV, "Appendices", consists of brief chapters on various aspects of linguistic performance that can safely be ignored by those whose interests are purely in theoretical linguistics. Appendix A concerns the computational problem of parsing. Appendix B concerns the problem of discovering a form- and language independent semantics. Appendix C considers the problem of Language Acquisition. Appendix D speculates concerning the emergence of language in the process of biological Evolution. Appendix E is the glossary.

The theory developed here began in joint work with Tony Ades and Anna Szabolcsi over 40 years ago. Thanks to Cem Bozşahin for joint explorations of linguistic diversity and the foundations of Combinatory Linguistics over a number of years, and to John Torr for guidance through the literature of the minimalist program. Help of various kinds was given along the way by Paul Atkinson, Miriam Butt, Geoff Pullum, Rob Truswell, and Bonnie Webber.

Earlier versions under various different titles formed the basis for a class at the LOT Winter School at the University of Tilburg, in January 2016, the ESSLLI Summer School, Bolzano in August 2016, the LSA Summer Institute, Lexington KY in July 2017, and the NASSLLI Summer School at USC in Los Angeles in 2022. My thanks to the students there and at Penn and Edinburgh for helping me by their criticism towards a better formulation of these ideas.

The work was supported in part and at various stages by EPSRC grants GR/M96889, GR/R02450, GR/R82838, and GR/S22509, EU IP grant FP6-2004-IST-4-27657 PACO-PLUS, ERC Advanced Fellowship 249520 GRAM-

PLUS, EU IP grant EC-FP7-270273 XPERIENCE, ARC Discovery Grant 160102156, and ERC H2020 Advanced Fellowship GA 742137 SEMANTAX.

Chapter 1

Introduction

1.1 The Problem of Language Diversity

At first glance, the languages of the world seem astonishingly diverse. For example, English has a huge and constantly expanding vocabulary of often highly specialized nouns—that is, words referring to kinds of artefact, like “towel”—which can freely be compounded to form new nouns denoting new kinds, such as “towel rack”. Unlike French, nouns can even be used as verbs denoting a characteristic event afforded by objects of that kind, as in “towel it dry”. In contrast, while Navajo has nouns for certain “natural kinds” like *asdz´ą´ą*, “woman” and *łóó’*, “fish”, many Navajo nouns are nominals formed from a verb denoting what Gibson (1977) called an *affordance* of the object in question—that is, an event that its presence makes possible. So *dily´įhí*, “lead” is literally “that which melts”, derived from the verb *dily´įh*, “melts” and the nominalizing enclitic *-í*, “the one who/that”. Similarly, *bee’ádít’oodí*, “towel”, is glossed as “that with which one wipes oneself”.

This process of deverbal noun formation is in Navajo comparably productive to the English formation of denominal verbs. *Bee ’ádít’oodíbąąh dah náhidiiłtsos*, “towel-rack” is glossed as “that upon which one repeatedly puts flat flexible things with which one wipes oneself” (Young and Morgan, 1980).[1]

This “synthetic” character of the language is pervasive. The transcript of a Navajo radio talk-show, of which the topic under discussion was the music of Hootie and the Blowfish, reveals that the participants had no problem in translating “Hootie” as a proper name and “fish” as *łóó’*. However, a long discussion was necessary to establish exactly who was blowing exactly what and with what effect before it was possible to come up with the equivalent of

1. The enclitic *-ł-tsos* on the verb *náhidii-*, “put up, hang” is the iterative aspectual form of a classifier for flat flexible objects, which in Navajo appears as part of the verb stem, rather than the noun.

"fish which inflates itself" as the Navajo translation of "blowfish".

Other languages, such as Hopi (Whorf, 1946; Whorf and Carroll, 1956), differ from English in almost entirely lacking the elaborate system of tenses, moods, and aspects (which are even more elaborate in Navajo than in English), and which we think of (mistakenly, as is argued in appendix A) as denoting time. Instead, Hopi has a similarly complex system of evidential markers, distinguishing the speaker's grounds for making their statement, such as witnessing it or receiving the information at second-hand, via hearsay.

This divergence does not mean that Hopi speakers think about time in a radically different way from speakers of English and Navajo, or that the latter think about blowfish entirely differently from English and Hopi speakers. Despite their apparent diversity of surface forms, there is every indication that the meaning representation or semantics that underlies English, Navajo, Hopi and all other languages (to which semantics of course we have no direct access) is essentially the same. That is, as assessed in terms of their ability to get up in the morning on time and keep appointments, the Hopi seem to think about time in very much the way we do, as Whorf himself pointed out (1950). Likewise, Navajo speakers seem to think about towels and blowfish in much the same way as English speakers.

Such differences as do seem to exist in the way that speakers of different languages conceptualize the world seem to be very slight and quite transient. Rosch (1974) reports a study by Greenfield and Childs of a language with only two color terms. After making sure that the native-speaker subjects could distinguish visually between white, pink, orange, and red, they were asked to copy a design of red and white drinking straws. The results differed according to whether, when previously asked to name the colors, subjects had produced different descriptive terms for them. Those who had used the same word to refer to both pink and white, and/or for both red and orange, tended to mix those pairs in copying the design, whereas those who had had used different compound terms for the colors tended to use only pure red and white.

Similarly, studies of the course of children's linguistic and cognitive development such as Bowerman and Choi (2001) and Gopnik (2001) show somewhat different sequences of development of spatial concepts, and of object categorization versus causal reasoning about events across languages, both being correlated with language-specific morphological and noun-verb ratio distributional differences between the languages and data the children were exposed to. These authors explain these differences in terms of drawing the attention of the respective groups to different aspects of a common conceptual representation, and consequent facilitation of reasoning about those concepts that the language made salient. Boroditsky (2001) shows language-specific interference effects from spatial tasks on latency of temporal judgements in English dependent on differential availability of spatial metaphors in subjects' first language (English or Mandarin) and age of second language acquisition in the latter. However, these effects seem relatively transient stages on slightly different paths from a common prelinguistic conceptual representation converging on a shared adult understanding of what is overwhelmingly the same world from all linguistic points of view (Gopnik, 2001:58-62).

On this view, the languages of the world share a semantics that is partly founded on a Kantian pre-linguistic (and probably to a considerable extent pre-human) internal representation of external (and internal) reality, evolved over hundreds of millions of years of vertebrate evolution, of a kind that is available to some extent to other animals, and partly on a socially constructed, more distinctively human reality, evolved over a few million years of hominid development.

It is consistent with this view that the major contribution of descriptive linguistics since its inception has been to identify an essential similarity across human languages that reflects this shared semantics, despite the fact that Individual languages can vary widely in the ways in which their lexicons carve this common conceptual representation at the linguistic joints.

For example, the English past tensed verb "ate", as in "Five boys ate a fish", is either underspecified or ambiguous as to whether it denotes a distributive event, in which each boy ate a different fish, or a collective one, in which a single fish was eaten. It is also ambiguous as to whether the event of either kind was repeated or an isolated occurrence. (Both ambiguities can of course be resolved by the addition of further modifiers, as in "Five boys ate a fish each for the next three nights".) Navajo, on the other hand, makes all of these distinctions explicit via the highly productive agglutinative verbal morphology noted earlier (Faltz, 2000).

Unfortunately, despite such insights, linguists and psychologists have been able to discover comparatively little about this primordial conceptual representation that must underlie English and Navajo and all other languages. Our only access to it is indirect, via the forms of adult language, which commits each speaker to one specific high-level partition of the original semantic information (or perhaps to a few such partitions, if they are multi-lingual). Such partitions are so close to the syntax of our adult language that it seems in practice to be almost impossible to see through them to the universal underlying conceptual representation to which we had access when we first learned them as children.

Fortunately, linguists do know a great many useful cross-linguistic generalizations concerning the possible forms of natural languages. Crucially, all of these different kinds of language, with their different partitions of the "hidden" underlying conceptual information, seem to share a type-system over the parts-of-speech (Hale and Keyser, 2002), which presumably reflects the type-system of the original Language of Mind. For example, all languages have transitive and intransitive verbs with the types of functions from one or two entities into propositions. English and Mandarin then distribute aspectual distinctions like the perfect and progressive over various adjunct categories, while Navajo packs them all into agglutinative verbal morphology. Similarly, most if not all languages have raising and control verbs, like "seem" and "persuade" which require subjectless infinitival verbphrases like "to go" as complements. But no language seems to have "super-raising" verbs that would allow a version of English in which strings like **John seems that it is certain to leave* meant the same as *It seems that John is certain to leave*.[2]

1.2 The Problem of Child Language Acquisition

From this point of view, the problem that faces a child acquiring their first language(s) is simply that of learning which elements of the sentences they hear correspond to which typable components of the meaning that the context affords. The meaning representation itself must be an articulated symbolic structural representation of the situation that we can think of as an expression of a language in its own right, like a language of mind. In its most basic form, learning consists in examining all possible pairings of all possible decompositions of the sentence into phrases and words with all possible decompositions of the meaning representation into substructures. While most of these pairings will be incorrect, the correct pairings will be more frequently supported by the context, so that a statistical model of all the possibilities considered will soon approximate to the adult grammar. (This procedure will also work when the context makes distracting irrelevant meanings available as well as the intended one. The details are briefly considered in appendix B.)

In order to solve that problem, it is generally acknowledged that the child needs acccess to more information than the mere sound of the utterance and its possible meaning representation(s). Most basically, it would help to know

2. The asterisk "*" is a useful linguists' annotation meaning "ungrammatical with the relevant meaning".

what are the most likely boundary points in the sound-string between possible words and phrases. It is possible that this part of the task can be accomplished on purely distributional grounds, that is, by observation of statistical regularities concerning sequences that frequently recur (Saffran, Aslin, and Newport, 1996; Goldwater, Griffiths, and Johnson, 2009). However, it is also likely that speech sounds are specialized for segmentation by the auditory system, which must originally have evolved for hearing non-speech sounds (Barlow, 1961), and that some sound-structure is therefore evident to children from the start.

It is also necessary for the child to have information about the legal types into which meaning representations can be split, to avoid wasting time on semantic types like verbs that control non-subjects, which it seems that no language actually entertains.[3]

Finally, the child needs to know what are the legal ways of combining words and phrases of various types into larger units to yield sentence meanings, in order to divide potential sentence meaning representations into components representing possible phrases and words.

These last two components, constituting a *type-system* consisting of the lexical types or parts-of-speech and the syntactic rules for combining them, are often referred to as “Universal Grammar”. The name is somewhat misleading, since it suggests that these elements are specific to language. In fact, nobody knows whether they are that specific or not. It is perfectly reasonable to think that concepts of the type that underlie transitive verbs like “grasp” and “climb”, and even the operations for combining such concepts with arguments are a part of a general apparatus of cognition that we share with pre-linguistic children, other primates, and even some other animals. We will not prejudge this issue, to which we will eventually return in appendix, A.

1.2.1 The Simplest Languages: Applying Concepts

What would we expect a natural grammar to look like on the basis of the preceding remarks? First, it must be the case that all languages are *semantically transparent*, in the sense that structural units such as noun-phrases and intransitive verbs must correspond to structural units of that primordial language of mind. (Otherwise, children wouldn’t be able to learn them from exposure to sentences paired with situations in the world that they denote.)

The artificial languages that humans have constructed for reasonig about the world, such as arithmetic, geometry, and propositional logic, are semantically transparent in this sense. Such languages have a simple syntax, usually de-

3. This claim requires a careful definition of the term “subject” that we will get to in chapter 2.

fined by an unambiguous grammar, which is completely defined by the semantics (of addition, conjunction, negation, etc.), apart from some purely syntactic conventions to make it easily readable and learnable by humans (such as that all binary connectives like $*$, denoting a procedure *times*, and $\Rightarrow$, denoting a truth-table, are realised the same way, say as infix operators). Then a sentence of standard Propositional Logic, such as the following, can be viewed as an instruction to decide its truth by applying the relevant truth table to the propositions "Tuesday" and "Belgium":

(1) $Tuesday \Rightarrow Belgium$

One might think that something similar is qoing on in understanding the related English sentence "If it's Tuesday, this must be Belgium", give or take a few complications about what exactly "it", "this", and the modal predicate "must" denote,

Such simple languages can be captured with very simple grammars defined with rules like the following for arithmetic, where S is short for "SUM", O is short for "OPERATOR" and 'N is short for "NUMBER":

(2)
$$\begin{array}{lll} S & \rightarrow & (S \quad O \quad S) \\ S & \rightarrow & N \\ O & \rightarrow & \{+,-,*,\div\} \\ N & \rightarrow & \{1,2,3,4,\ldots\} \end{array}$$

Such rules are referred to as "context-free" (CF) rules, because there is only one symbol X to the left of $\rightarrow$, so that the realization of a think of type X is independent of the things on either side of the X. (Rules with more than one symbol on the left are called "context-sensitive".) Collections of context-free rules like (2) are refered to as "context-free grammars" (CFG), and the set of expressions conforming to a CFG is a "context-free language" (CFL). Sentences of a CFL are guaranteed to be parseable in time at most polynomial in their length n, actually n^3, which means there are efficient "divide and conquer" algorithms for doing so.

They can also be made to build semantically interpretable logical forms in parallel with parsing. For example, the following version of (2), in which each type is paired with an interpretation with the separator ":", computes the value of expressions like $(1+2)*(3+4)$ as sums with values like 21

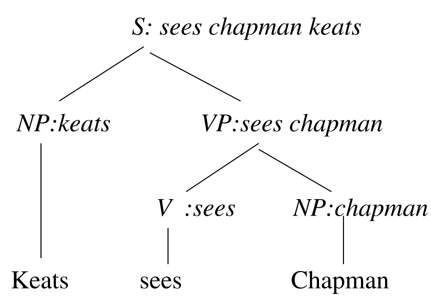
(3)
$$
\begin{array}{lll}
S : o\,s_1\,s_2 & \rightarrow & (S : s_1 \quad O : op \quad S : s_2) \\
S : n & \rightarrow & N : n \\
O : o & \rightarrow & \{+ : add, - : sub, * : mul, \div : div\} \\
N : n & \rightarrow & \{1 : 001, 2 : 010, 3 : 011, 4 : 100, \ldots\}
\end{array}
$$

Such grammars can be applied to capture some basic facts about natural languages, such as the syntax and semantics of transitive and intransitive clauses:

(4)
$$
\begin{array}{lll}
S : v\,n & \rightarrow & NP : n \quad VP : v \\
VP : v & \rightarrow & V_1 : v \\
VP : v\,n & \rightarrow & V_2 : v \quad NP : n \\
\ldots & & \\
NP : n & \rightarrow & \{\text{Keats} : keats, \text{Chapman} : chapman \ldots\} \\
V_1 : v & \rightarrow & \{\text{walks} : walks, \text{talks} : talks, \ldots\} \\
V_2 : v & \rightarrow & \{\text{sees} : sees, \text{forgets} : forgets, \ldots\} \\
\ldots & &
\end{array}
$$

The symbols *S, NP, VP*, and *V* in these rules are mnemonic for “Sentence”, “Noun-Phrase”, “Verb-Phrase”, and (tensed) “Verb”. Lower case symbols *n* and *v* are the meanings of *NP*s and *V*s. Juxtaposition of lower-case symbols, as in $v\,n$ denotes application of the fomer to the latter. The symbol $\rightarrow$ means that the thing on the left of it “can be made up of” the things on its right in that order. (So the first rule above means that “A sentence meaning *v* applied to *n* can be made up of an *NP* meaning *n* and a *VP* meaning *v*, in that order.

It is natural to think of the analysis or “derivation” of a sentence according to these rules as a “phrase-structure tree”:[4]



(5)

Even if natural language is not context-free, it would be very desirable for it to have all of these properties. We would like it to be as near context-free as possible.

4. Because they think of rules as generative, linguists always draw trees the wrong way up, with the root at the top and the leaves below, like a family tree.

Nevertheless, it might seem surprising to suggest that natural syntax is semantically transparent to the child's semantic language of mind in the same sense, since we have already noted that natural languages, unlike the different varieties of arithmetic and those of propositional logic, differ wildly in even such basic grammatical matters as what counts as a noun or a verb. However, semantic transparency does not require that the structural units of syntax and semantic representation correspond one-to-one. (It is unlikely that *Tuesday* and *Belgium* are native concepts of the child's language of mind.) We need only assume that the relevant concepts *can be expressed* in the language of mind, possibly in terms of other non-primitive concepts that have been acquired previously.[5]

1.2.2 The Simplest Languages: Forming New Concepts

The fact that language learning depends in this way on the ability to define new concepts in terms of existing concepts means that the language of mind that the child uses to first understand the world must include the capability of *abstraction*, as well as that of *application* of an concept to an instance. That is to say that the Language of Mind and the language- and culture-specific semantics that is built upon it constitute what is called an *applicative system*.

Applicative systems are calculi that not only define the notion of *application* of a function such as multiplication to its arguments, as in basic arithmetic, but also define the notion of *abstraction*. Abstraction allows us to define new functions such as the *square* of a number in terms of existing functions like multiplication, here written with the * operator:

(6) $square = \lambda x.x * x$

The abstraction operator λ "binds" a variable x which is used to pass a value to all occurrences of x in the "body" or definition of the new function, which says it is to multiply x by itself. By making this function the value of the identifier *square*, we make it part of the language of arithmetic, so that we can apply it to an argument

(7) $square\, 2 = 4$

In linguistic terms, λ-abstraction allows us to not only apply a concept or function such as *father* to individuals to yield their fathers as a result, as in (a,b), but also to define a new concept/function *grandfather* in terms of *father*,

5. Conceptually-advanced sentences like "if it's Tuesday, this must be Belgium" are is unlikely to be found in early child-directed utterance, and if they are, will just be ignored.

as in (c), which can be applied to pairs of individuals to yield their grandfathers, as in (d):[6]

(8) a. $\mathit{father}\,\mathit{Esau} = \mathit{Isaac}$
b. $\mathit{father}\,\mathit{Isaac} = \mathit{Abraham}$
c. $\mathit{grandfather} = \lambda x.\mathit{father}\,(\mathit{father}\,x)$
d. $\mathit{grandfather}\,\mathit{Esau} = \mathit{Abraham}$

Here, and throughout the book, linear concatenation of a function f and an argument a, as in $f\,a$, indicates application of the former to the latter, and $=$ (read "yields") indicates the result. Application "associates to the left", so that $f\,a\,b$ is equivalent to $(f\,a)\,b$.[7]

In (b), the abstraction operator λ declares a variable x as a parameter of the function, to which a value such as *Esau* can be bound by application and used in the body of the function definition to compute the value or result of what we name via the "=" operator the *grandfather* function, which when applied to *Esau* yields *Abraham*, as in (c).[8]

In what follows, it will be important to understand that there is another way of formalizing applicative systems, without the use of variables and λ-binding, using *combinators*. Combinators are operators that apply directly to functions like *father* in (8) to yield new functions. For example, the most basic combinator is *function composition*, Composition is sometimes written as an infix operator $\circ$ and sometimes as a prefix operator **B**. The functor *grandfather* from (8b) can be defined in either combinatory notation as follows:[9]

(9) a. $\mathit{grandfather} = \mathit{father} \circ \mathit{father}$
b. $\mathit{grandfather} = \mathbf{B}\mathit{father}\,\mathit{father}$

Both are equivalent to the right-hand side of (8b), $\lambda x.\mathit{father}\,(\mathit{father}\,x)$. However, it should be noticed that neither involves the explicit variable or variable binder λ of the latter.

Quite simple ensembles of combinatory rules of this kind can be used to define calculi equivalent to the λ calculus (Schönfinkel, 1924; Curry and Feys, 1958). The implication is that the use of variables, and hence of traces or

6. Concatenation as in *father Esau* represents *application* of a function to an argument.
7. This convention takes a little getting used to but it saves space, a fact that will become increasingly important as we progress.
8. Since *father* is also a function, we could have written it as $\lambda x.\mathit{father}\,x$. However, this elaboration would be redundant.
9. Functions and operators with more than one argument are binarized, or "Curried", and, as usual, function application associates to the left, so that $\mathbf{B}\mathit{father}\,\mathit{father}$ is equivalent to $(\mathbf{B}\mathit{father})\mathit{father}$.

copies, is not a theoretical necessity in the definition of an applicative system.

A child equipped with such an applicative system over a language of mind can acquiring their first language(s) in contexts they can understand by building the semantics of the language in question from the precursor language of mind, using abstraction or the equivalent combinatory operators to define word meaning and combine them to yield phrase meanings, gradually acquiring a semantics that the language in question is transparent to. The relation of that semantics to the original language of mind is one that a computer programmer would recognize as "compilation" of a high-level computer language into a low-level assembly code. Just like a compiler for a programming language like Python for a program developer, the language-specific semantics allows the child to develop useful thoughts much faster and more efficiently than would be possible using the primitive language of mind. Some of them will be culturally acquired concepts that that might otherwise never be accessible to an isolated agent without access to an established language. (As a result of this process, adults seem to entirely lose access to the language of mind that supported the earliest stages of language acquisition.)

However, a number of complications face the child or computer program that has to learn language on this basis.

Exercise : Define the concept "great grandfather" in both the λ-calculus notation in (8b) and the combinatory calculus notation in (9b).

1.3 The Problem of Ambiguity

One problem is the incredible degree of syntactic ambiguity characterizing natural grammars.

There is a (probably apocryphal) legend that in a early public demonstration to funders of machine translation in the 1950s, the presenters were dismayed to find that the machine's translation of the sentence "Time flies like an arrow" into Russian, when translated back into English, had been analysed as analogous to the sentence "Fruit-flies enjoy a banana"—that is, with *time flies* translated as a noun and *like* as a verb.

More generally, perfectly understandable sentences of moderate length that we encounter routinely in newspapers and in conversation have hundreds, frequently thousands, and in certain pathological cases millions of syntactically well-formed analyses, almost all of which human language users are blissfully unaware.

The fact that human languages entertain such huge ambiguity clearly calls for explanation. Such proliferation is something that we never allow in the artificial languages of mathematics, logic, or computer programming. (Nobody wants their computer to compute unintended *double entendres*.) Its profusion implies that humans have access to some very powerful mechanism for eliminating syntactic irrelevancies.

The example suggests that this mechanism is either semantic (*time flying* makes sense, unlike the spurious noun *time-flies*), or statistical (the former is more frequent in corpora of English than the latter). In computational natural language processing (NLP), statistical parsing models are currently the only practical way of limiting the huge search problem engendered by ambiguity-natural language. They have the advantage that, with neural computational methods, they can be trained on unlabeled raw text, However, such models are very large, and require many orders of magnitude more text to train them than a human encountrs in a lifetime, let along a five-year old child. It is likely that human language understanding uses a mixture of statistical modeling and knowledge-based inference for this purpose.

Ambiguity and the way it is resolved is not the linguist's problem, and further discussion of the problem is defered until the appendices. By the same token, linguists should be wary of criticizing any linguistic theory merely on the grounds of increasing ambiguity, particularly when the increase comes in only one component of the grammar, with the possibility of savings elsewhere. Some of the early criticisms directed against the application of context-free grammars to linguistic analysis of Harman and Gazdar were of this kind, and we shall need to avoid the temptation below. The fact that no language in the world shows the slightest sign of moving in the direction of reducing ambiguity (despite the prevalence of drift in other aspects of grammar) should tell us that ambiguity really isn't a problem for human language users, and *there must be a way*.

1.4 The Problem of Discontinuous Constituents

The phenomenon of *discontinuity* or long-range dependency, where two elements that are semantically dependent upon each other as predicate and argument are not structurally contiguous, is ubiquitous in natural language, and constitutes the fundamental problem of theoretical linguistics. For example, in (10a) , *a unicorn* is the semantic agent of discontiguous *approaching*, rather than of contiguous *seems*; in (b) *Chapman* is semantically the agent of discontiguous *write*; and in (c) *who* is nonadjacent to *saw*. (d) involves a discontiguous dependency of *a book* as object of *bought*, as well as that of adjacent *sold*.

(10) a. A unicorn$_i$ seems to be approaching$_i$.
b. Chapman$_i$ wants to try to begin to write$_i$ a play.
c. Who$_i$ did you say Keats thinks Chapman saw$_i$?
d. I bought$_i$ and you said you sold$_i$ a book$_i$

The long-range dependency in (10a) is referred to as *bounded*, because it occurs within the domain of a single tensed verb *seems*. That indicated in (10b) is also regarded as bounded, because semantically it is mediated by a cascade of intervening similarly bounded dependencies of the subjects of infinitival verbs on their parent's subject. (10c) is referred to as *unbounded*, because the dependency between *who* and *saw* can span any number of intervening tensed domains without being in any obvious sense an argument of the intervening verbs, here *say* and *think* (Bresnan, 1977). Right node raising (d) shows that rightward long-range dependency is also allowed, as between *bought* and *a book*.

The bounded dependencies that relate co-arguments of a single head can be handled in the lexical semantics of that head. "Seems" in (10a), can accordingly be assigned the following λ-term as its logical form, in the body of which the *lambda*-bound subject variable *y* appears as the subject or agent argument of *p*, which corresponds to *approaching* in (10a), rather than "seems":[10]

(11) $\lambda p \lambda y.seem\,(p\,y)$

As a result, the meaning of (10a) is something like the following:[11]

(12) $seem\,(approaching\,(a\,unicorn))$

Technically, the logical form in (11) is a *second-order* function/predicate, because the argument *p* is itself a predicate. In (10b), all of the verbs are semantically second-order predicates whose subject *y* appears once as the as the subject of their complement *p*, and once of the tensed verb itself, such as the following:

(13) $\lambda p \lambda y.wants\,(p\,y)\,y$

10. As usual, application associates to the left.
11. The meaning representation is simplified: we defer disussion of the ambiguity of "a unicorn" between *de dicto* and *de re* readings.

As a result, the meaning of (10b) is something like the following:[12]

(14) $\mathit{wants}\,(\mathit{try}\,(\mathit{begin}\,(\mathit{write}\,(\mathit{a\,play})\,\mathit{chapman})\,\mathit{chapman})\,\mathit{chapman})\,\mathit{chapman}$

The fragment *did you say Keats thinks Chapman saw* involved in the unbounded dependency in (10c), repeated here as (15a), seems somewhat analogous semantically to the abstraction (15b):[13]

(15) a. Who$_i$ did you say you think you saw$_i$?
b. $(\lambda x.\mathit{say}(\mathit{think}(\mathit{saw}\,x\,\mathit{you})\,\mathit{you})\,\mathit{you})\,\mathit{who}$

"Who" can then be thought of as providing a variable *who* as an argument to that function to create an open proposition corresponding to a question $\mathit{say}(\mathit{think}(\mathit{saw}\,\mathit{who}\,\mathit{you})\,\mathit{you})\,\mathit{you})$.

Similarly, the fragments "I bought" and "you said you sold" in (10d) can be thought of as λ-terms $\lambda x.\mathit{bought}\,x\,i$ and $\lambda x.\mathit{said}\,(\mathit{sold}\,x\,\mathit{you})\,\mathit{keats}$, and as being combined by conjunction to yield the λ-term $\lambda x.\mathit{bought}\,x\,i \wedge \mathit{said}\,(\mathit{sold}\,x\,\mathit{you}))\,\mathit{keats}$, which when applied to "a book" yields the following:[14]

(16) $\mathit{bought}\,(\mathit{a\,book})\,\mathit{me} \wedge \mathit{said}\,(\mathit{sold}\,(\mathit{a\,book})\,\mathit{you})\,\mathit{keats}$

MOVE Both (15a,b) exhibit a relation of structural *command* of the bound element by the λ-binder/*wh* element, such that the bindee falls in the scope of the binder. Such command relations are ubiquitous in linguistic theory, and characterise all linguistic dependencies (Epstein, 1999; Hornstein, 2009). The fact that such relations are also characteristic of all applicative systems seems likely to be an important clue to understanding the linguistic system.

The linguistic literature since Chomsky (1957) can be read as tacitly or explicitly assuming that in semantic terms, language is an applicative system. However, it is very surprising from this point of view that natural language sentences like (15a) include no phonological realization of such crucial elements of meaning as the variable x or its binder λx that appear in the abstraction (15b) (nor any equivalent combinatory operators of the kind seen in (9)). It is not clear in their absence how the two elements of the long-range dependency are identified syntactically and united semantically. The linguistic literature can be read as offering two kinds of solution to this problem.

The most common solution is to include rules of displacement in the syntax

12. We defer discussion of the mechanics of constructing such interpretations until later.
13. Such logical forms differ from those in (8) only in that functions like *met* and *think* are functions into propositions rather than individuals. We defer discussion of the mechanics of constructing them
14. We defer discussion of the ambiguity as to whether the books are the same or different.

itself, in the form of "transformational" rules which explicitly permute or associate non-structurally adjacent elements. Since the early 70's, this has been described in terms of "movement" of the displaced item from the non-displaced semantically interprtable position such as the object of "saw" in (15a) to its surface position as for "who". Sometimes this process is thought of as leaving a "trace" at the original position (Chomsky, 1975), which may be co-indexed with the displaced version, equivalent to a bound variable. ((15b) is in fact exactly the kind of logical form that Heim and Kratzer (1998):97 and Fox (2002):67 propose to derive from the output of movement via trace-conversion rules. However, movement is still doing the real work of displacement in these theories, since, somewhat surprisingly, the surface string includes no phonological realization of anything corresponding to either a variable or a binder.)

More recently, movement has been talked of in terms of "copies", leaving a complete version of the moved element in situ, rather than a trace (Chomsky, 1995b). On occasion, these "copies" are thought of as *identical*, in the sense that they are somehow simultaneously instantiated in both source and target positions of the movement, and thereby distinguished from independent repetitions (Chomsky, 2007: 10). This also is somewhat reminiscent of the instantiation of a bound variable by a λ-binder. However it raises the question of why only one "copy" is actually pronounced, if both are present. We will return to this questiion later, but it implies the possibility of rules of *deletion* of material from the surface form of the sentence. On occasion, some actually discontinuous constructions like the following, which is hard to describe in terms of movement (since neither *I gave* nor *Adlai a record* are usually thought of as constituents), are also talked of by linguists as involving deletion, here indicated by overstriking:[15]

(17) I gave Ike a bike and ~~I gave~~ Adlai a train.

The resemblance of movement to abstraction under any of its linguistic interpretations makes it seem a very general operation. The λ-calculus can essentially represent any computable fuction, so there is a question whether there is any conceivable linguistic phenomenon that could *not* be captured in terms of unconstrained movement. If not, then it is not clear that movement counts as an *explanation* of the phenomenon of displacement, rather than constituting a general notation for describing the phenomenology of discontinuous con-

15. In some recent work, both copy-movement and deletion are thought of in terms of *multidominance* (Citko, 2011). For example, in (17), the structures dominating "I" and "gave would be dominated by internal nodes of a conict of type TP on the right, as well as by those of the left conjunct, making the structure a graph, rather than a tree.

structions across languages. (Of course, the latter is an important first step in providing an explanation—in fact, the terms "movement" and "deletion" are so descriptively vivid that we will use them freely in what follows to describe constructions, without any commitment to their theoretical reality.)

The other solution to the problem of displacement on offer from linguistics since the introduction of Generalized Phrase Structure Grammar (GPSG, Gazdar, 1981) is to pass a feature through the derivation linking the two ends of the long-range dependency, and marking the result as having undergone abstraction (the path-based work of Kayne, 1983 and Pesetsky, 1982 is related.) In the Gazdar's version, the grammar was context-free, and the lf interpretation was done in parallel with derivation. Lexical Functional Grammar (LFG, Bresnan, 1982) and Head-driven Phrase Structure Grammar (HPSG, Pollard and Sag, 1994) can be seen as attempts to generalize feature passing beyond the context-free case (although both have independent origins).[16]

1.5 Some Complex Discontinuities

There are other examples of long range dependency that seem both more complex than those permitted by context-free grammar, and more limited than would be expected on the assumption of free movement and equivalence to the full λ-calculus. The remainder of this section briefly reviews the characteristics of some of the major types of long-range dependency that present problems for constructing constrained theories of grammar, and that will be analysed in more detail in the rest of the book..

1.5.1 Argument/Adjunct Cluster Coordination and the Order of Constituents

Coordinate constructions pose the greatest challenge to any theory of natural grammar, including the movement theory. In particular, (17), repeated here with some variants, is particularly important.

(18) a. I gave Ike a bike and Adlai a train.
b. I saw Ike on Monday and Adlai on Wednesday.
c. I told Ike that it was raining and Adlai that it was snowing.

Because strings like *Adlai a train*, *Adlai on Wednesday*, and *Adlai that I would leave* do not look like traditional constituents, it is common to refer to the phenomenon as "non-constituent coordination". However, the name suggests that we might be prepared to regard coordination as exempt from the

16. All of these theories are among those surveyed in Steedman, 2019.

constituent condition on rules (Chomsky, 1955/1975), which says that all rules of grammar must apply over constituents, and yield a constituent as their result. The Constituent Condition is merely a corollary of the fact that rules of grammar have to have a compositional semantics, and that constituents are things that have interpretations. Rules of coordination must therefore take constituents as their inputs, so the only room for disagreement concerns the *type* of constituents like *Adlai a train*, whether as sentences like ~~*I gave*~~ *Adlai a train*, where *I gave* is unpronounced or "deleted under coordination", or as constituting constituents in their own right, as proposed below. We will therefore eschew the term "non-constituent coordination", in favor of "argument/adjunct cluster coordination".

The phenomenon of argument/adjunct cluster coordination is very widespread, and possibly universal, in the languages of the world. So in Japanese, an SOV language, we have cases like the following:

(19) Boku-ga [Anna-ni hon-o] , [Manny-ni hana-o] yatta.
I-NOM Anna-DAT book-ACC Manny-DAT flowers-acc gave-PAST
'I gave Anna a book, and Manny flowers'

In Welsh, a VSO language, we have (Borsley, Tallerman, and Willis, 2007:52):

(20) Rhoddodd yr un dyn [lyfr i Mair] a [darlun i Megan].
Give.PAST.3S the one man book to Mair and picture to Megan
'The same man gave a book to Mair and a picture to Megan'

As Ross (1970) pointed out, there is a striking generalization concerning such coordinations cross -inguistically which the above examples illustrate, which can be stated as follows:

(21) *Ross's (1970) Generalization*:
If the material that is deleted under coordination would normally find its arguments to the right (left), then the site of deletion is in the right (left) conjunct.

For example,, in SOV languages/constructions, the verb is is missing from the *left* conjunct, as in (19). In VSO languages/constructions, the verb is is missing from the *right* conjunct, as in (20).

Ross also pointed out that SVO languages like English pattern with VSO: the deletion is in the right conjunct. The English cluster coordination examples (18) illustrate the point, as does the English medial gapping construction:

(22) Anna married Manny, and Tom Sue.

There is a little more to say about free word-order languages, and mixed word-order languages like Dutch and Zapotec, which not surprisingly show mixed gap directionality (see *SP* for some discussion). But when all is said and done, Ross's generalization is one of the strongest syntactic universals that has yet been identified. It seems to be telling us that all details of any language's syntactic projection onto coordinate constructions is determined in its lexicon. That is, if the lexicon specifies an argument as being to the right (left) of a verb, then in the absence of an explicitly direction-changing category such as a relative pronoun, its projection under operations like coordination will be, too.

1.5.2 Extraction "Across-the-Board" under coordination

Ross (1967) noticed that extraction from conjunctions is in general impossible, as in (24a,b), an observation which he enshrined in the Coordinate Structure Constraint (CSC) on extraction.

(23) a. *a man $that_i$ [I $like_i$ and you hate him]
b. *a man $that_i$ [I like him and you $hate_i$]
c. a man $that_i$ [I $like_i$ and you $hate_i$]
d. *a man $that_i$ [I $like_i$ and $hates_i$ you]
e. ?a man $that_i$ [$hates_i$ you and I $like_i$]

However, Ross also noticed that extraction out of coordinate structures *is* allowed when *all* conjuncts undergoe extraction, as in (24c), a fact that he described as the "Across the Board Exception" to the CSC (ATB). Williams (1978) further noticed that ATB extraction failed if one of the extractions was of a subject and the other of a non-subject asin (d), a fact enshrined in the "Same Case Condition" on the ATB exception to the CSC. (A number of people have noticed that the latter condition seems a little weaker in the case of (e)).[17]

The same generalization holds with even greater strength for rightward extraction, known as "Right Node-Raising", and in this case the same case condition violation on (e) is clear.

17. The prefix "?" is a useful linguists' notation marking a sentence whose grammatical status is uncertain.

(24) a. *[I like$_i$ and you hate him] the man in the Brooks Brothers shirt.
b. *[I like him and you hate$_i$] the man in the Brooks Brothers shirt.
c. a man that$_i$ [I like$_i$ and you hate$_i$] the man in the Brooks Brothers shirt.
d. *[I like$_i$ and hates$_i$ you] the man in the Brooks Brothers shirt.
e. *[hates$_i$ you and I like$_i$] the man in the Brooks Brothers shirt.

There is something rather absurd about a Condition on an Exception to a Constraint. However, it is important to note, first, that this is a strong phenomenon: (24a,b) are very bad. And, second, although also involving multiple dependencies on a single relative pronoun, ATB extraction seems to be a different phenomenon from parasitic gapping: *neither* of the two ATB extractions in (24c) is permitted on its own, as in (24a,b), in contrast to parasitic extraction, considered next.

1.5.3 Parasitic gaps

A particularly awkward phenomenon for analysis in terms of the movement metaphor arises from the multiple dependencies on a single *wh*-element that are referred to as Parasitic Gaps (Ross, 1967; Engdahl, 1983; Gazdar, Klein, Pullum, and Sag, 1984; Chomsky, 1986a; Steedman, 1987; Cinque, 1990; Nunes, 2004), exemplified by (25a), in which the relativized item *that* depends upon *two* verbs, *filed* and *reading*, and in which the non-adjunct extraction is allowed on its own, as in (25b), in constrast to ATB extraction.[18] :

(25) a. Articles that$_i$ I filed$_i$ without reading$_i$
b. Articles that$_i$ I filed$_i$ without reading your instructions.
c. *Articles that$_i$ I filed your report without reading$_i$

The movement metaphor in all its forms becomes less attractive if we need to think of one element's possibility of movement from a position within an adjunct that (in contrast to across-the-board extraction in the last section) is normally inaccessible to relativization (see (25c)) as being contingent on the movement of another element to the same place

Many different analyses of the phenomenon have been proposed, from the movement-based account of Chomsky, to anaphora-based accounts, in which the parasitic gap is realized as a proform of some kind, either a null resumptive (Cinque, 1990), a null epithet (Lasnik and Stowell, 1991), *pro* (Browning,

18. We will return to the question of the difference between ATB and parasitic extraction in chapter 11, where we will also consider some supposed exceptions to the ATB condition itself noted by Ross and Goldsmith (1985).

1987), or *PRO* (Weinberg, 1988), and the non-transformational proposals of Gazdar et al. and the present approach.

More recently, Nunes (Nunes, 2001, 2004) has proposed an analysis in terms of "sideward movement", to which we will return in chapter 9.

1.5.4 Multiple Dependencies: Nesting and Crossing

The first of these problems arises from the fact that natural languages allow *multiple* long-range dependencies. In many cases, like the following multiple *wh*-questions in English, the dependencies must nest and may not cross.

(26) a. Which violin$_i$ is which sonata$_j$ easiest to play$_j$ upon$_i$?
b. *Which sonata$_i$ is which violin$_j$ easiest to play$_i$ upon$_j$?

However, many Germanic languages and dialects including Dutch, West Flemish, and Zurich German allow unboundedly many crossed dependencies in certain contructions (examples for the latter from Shieber, 1985):

(27)

... das	mer	em Hans	es huus	haelfed	aastriiche
... that	we.NOM	Hans.DAT	the house.ACC	helped	paint

'... that we helped Hans paint the house.'

(28)

... das	mer	d'chind	em Hans	es huus	loend	haelfe	aastriiche
... that	we.NOM	the children.ACC	Hans.DAT	the house.ACC	let	help	paint

'... that we let the children help Hans paint the house.'

Such examples present an important challenge to the formal research program of defining the class of possible human languages via a theory that is more expressive than context-free grammmar, yet is more constrained than the Universal Turing machine that can capture any computable relation between strings and meanings. They provide part of the motivation for seeking some more constrained expression of long-range dependency than the original very general notion of movement proposed in the transformational tradition.

1.6 Pronominal Anaphora and Coreference

There is a strong reconstraint on whether a pronoun and and a full noun-phrase can corefer or be "bound" as indicated by the indices:

(29) a. Lola $_i$likes the person she$_i$ works for.
b. *She$_i$ likes the person that Lola$_i$ works for.

The constraint can for present purposes be stated as that a pronoun cannot be interpreted as coreferential with a full NP that it precedes and commands, where a node A commands a node B if the node that immediately dominates A dominates B, and neither of A and B dominates the other (Reinhart, 1981).

However, if it is correct to believe that the following sentences are just displaced versions of (29b), in which "the person she works for" is a displaced dependent of "like", so that "she" commands "Lola" why aren't they equally ungrammatical?

(30) The person that Lola works for $_j$, she likes$_j$.
Which person that Lola works for $_j$ does she like$_j$?

It cannot simply be that referents like "Lola" must precede as well as command coreferring pronouns, because the following seem to have the same meaning as (**??**):

(31) The person that she works for $_j$, Lola likes$_j$.
Which person that she works for $_j$ does Lola like$_j$?

1.7 Some Distracting Anaphoric Coordinate Constructions

There are a number of coordinate constructions that arguably involve anaphoric relations between elements, rather than purely syntactically mediated ones, which (like pronominal anaphora) should probably be treated as falling outside the theory of sentential grammar. Hankamer and Sag (1976) and Sag and Hankamer (1984) offer a number of criteria for distinguishing these constructions, of which the simplest and most important is their potential to be used intersententially (and even across speakers), as well as intrasententially, suggesting that in both cases they are mediated by discourse anaphora/reference, and that they should therefore be excluded from purely syntactic treatment. Among them are the following.

1.7.1 *do so* anaphora

The following examples involve an explicit anaphor *so*, and the anaphoric relation can be split across two utterances:

(32) a. I caught a fish, and you did so, too.
b. I caught a fish, and so did you.
c. Me: I caught a fish.
You: So did I

1.7.2 VP anaphora

Do so anaphora is closely related to VP anaphora, in which the anaphoric element is not explicit

(33) a. I caught a fish, and you did, too.
b. Me: I caught a fish.
You: I did, too.

1.7.3 Respectively

While we have seen that cross-serial syntactic dependencies do exist, Pullum and Gazdar (1982) argue that the very different kind of cross-serial dependency in the following construction should not be treated syntactically. Instead, "respectively" should be regarded as an anaphoric referential element meaning something like "in the order of mention".

(34) Bob and Ted married Carol and Alice, respectively.

If so, its binding should not be considered part of the problem of grammar, any more than the discourse reference of terms like "the former" and "the latter".

1.7.4 Extraposition

Extraposed modifiers constitute a further construction that appears not to obey the constraints that characterize true syntactic dependency. For example, extraposed NP modifiers do not show an asymmetry with respect to subjects, unlike the extractions in (36):

(35) a. A man came in that I didn't recognize.
b. I saw a picture in the paper of the scene of the crime.
c. The fact surprised us all that Albert had fled the country.

(36) a. *Which man do you think that came in.
b. *Of which scene did you see a picture in the paper of the crime.
c. *That Albert had fled the country the fact surprised us all.

Moreover, the antecedent that they appear to modify may not even exist grammatically as a constituent, even at the level of logical form:[19]

(37) a. A man came in and a woman went out that I didn't recognize.
b. A man came in and a woman went out that seemed to like each other.

We will follow Wittenburg (1987) and Culicover and Rochemont (1990) in concluding that at least the above varieties of extraposition are mediated by an anaphoric element under S-adjunction.

1.7.5 Sluicing

Perhaps Hankamer and Sag's most conclusive example of a construction that is anaphoric rather than syntactic is "sluicing" (Ross, 1967):

(38) a. Somebody caught a fish, but I don't know who.
b. Me: Somebody caught a fish.
You: I wonder who.

Ross's Ross (1969) syntactic account of sluicing, according to which (38a) arises from underlying relativization followed by deletion, as in *Somebody caught a fish, but I don't know who ~~caught a fish~~*, has recently been influentially revived by Merchant (2001, 2006).

Merchant bases his argument, following Ross, on the fact that in a wide range of languages with relative pronouns that agree in case with that assigned by the verb they are extracted from, including German, a sluiced relative shows that agreement as well:

(39)

Er	will	jemandem	schmeicheln,	aber	sie	wissen	nicht,	wem
He	wants	someone.DAT	flatter,	but	they	know	not	who.DAT

"He wants to flatter someone, but they don't know who(m)."

In English, dialects that preserve the "who/whom" distinction in relative pronominal caseexhibit the same case agreement, and in all dialects it shows up for the genetive:

(40) I borrowed somebody's comb, but I can't remember *who/whose.

Merchant also argues for his syntactic analysis on the basis of preposition-stranding. In languages like English which allow it, the sluiced relative can either be the relativized PP, or the relative pronoun alone, as if the sluiced

19. Compare the latter with the unacceptability of the following right-node-raising example:
(i) *Frankie seemed and Albert claimed to like each other.

fragment could include a stranded preposition:

(41) I know I spoke to someone, but I don't know who(m) ~~I spoke to~~

However, in languages like German which disallow preposition-stranding, the sluiced relative must include the preposition:[20]

(42) Ich habe mit jemandem gesprochen, aber ich weißnicht *wem/mit wem.

However, the same effect of case shows up in German two-party dialogs analogous to (38b). And since prepositions are a manifestation of case in German, it is hardly surprising that the same effect shows up for relativized PPs. (We saw in chapter 9 that English stranding propositions resemble particles, rather than case-markers.)

The strongest evidence against the Ross/Merchant claim is that sluicing remains strikingly immune to the island constraints that normally apply to relativization, as shown in the following oppositions:

(43) a. Anna caught a fish that bit her finger, but I don't remember which finger ~~Anna caught a fish that bit~~.
b. #Which finger did Anna catch a fish that bit?

(44) a. They said that a fish bit Anna, but I don't remember what kind of fish ~~they said that bit Anna,~~.
b. *What kind of fish did they say that bit Anna?

(45) a. I saw Anna and someone, but I don't remember who.
b. *Who did you see Anna and?

This fact led Ross to stipulate cross-derivational constraints determining coinditions under which island violations could be "amnestied"

The possibility of intersentential anaphoric reference of the *wh*-element in (38) leaves open the possibility of *intra*sentential anaphoric reference as well. Chung, Ladusaw, and McCloskey (1995) propose a referential account in a discourse representation-theoretic framework. See Chung (2013) for an even-handed discussion of both sides of this still-open debate.

In the rest of the book, we will ignore these potentially discourse-anaphoric constructions as external to the syntax.

20. Merchant's Preposition Stranding Generalization (PSG) is contested by Diogo and Yoshida (2007), who show that Brazilian Portuguese, which does not strand prepositions under relativization, allows the equivalent of (41), and suggest that there are two sources for preposition stranding.

1.7.6 Spoken Intonation

When first exposed to the traditional account of grammar, many students resist the traditional division of a simple transitive clause into a subject–*Manny*, say—and a predicate or verb-phrase including the object, such as *married a millionaire*. They often argue that a partion into the subject and verb *Manny married*, and the object *a millionaire*, seems just as reasonable. When asked to justify their intuition, they invariably point out that you can use intonation to partition the sentence in either way, depending on the context.

For example, in the context of the following discourse, one can answer the question *Who did Manny Marry?* as shown:

(46) Me: Manny used to date a dentist.
You: Who did he MARRY?
Me (Manny MARRied) (A MILLIONAIRE.)

Here, small caps indicate intonational accent or emphasis, with the accent on the first syllable of "married" being *late* with respect to the initial syllable onset in comparison with that on *a millionaire*. Parentheses indicate separate intonational phrases with the medial boundary marked by lengthening and/or rising pitch on the second syllable of *married*, and the final boundary marked by low pitch and length. (We will come to a more formal notation later).

This intonation seems to structure the semantic information in the sentence into a "topic", *(who) Manny married* (as opposed to dated), and a "comment", *(that it was) a millionaire* (as opposed to a dentist). The students clearly think that sentence structure ought to be the same as intonation structure.

The traditional syntactician's claim for the special syntactic status of the predicate lies in the fact that there are lexical items—intransitive verbs like *walks*—that can be substituted for *loves Mary*, but no such lexical items that that can be substituted for *John loves*. On this basis, the traditional view is that, whatever intonation structure is doing, it isn't the same as syntactic structure.

Nevertheless, it seems odd that there should exist an alternative level of structure related to meaning but orthogonal to syntax. Since the only point of syntactic structure is to support semantics, this seems to amount to a claim that natural language has *two* syntaxes. We would expect syntax and prosody to be homomorphic, as under the MATCH hypothesis of Selkirk (2011).

Interestingly, intonation structures of this kind, orthogonal to the traditional subject-predicate division, are very frequent in child-directed speech from the earliest years (Fisher and Tokura, 1996):

(47) a. That looks like a DOGGY.
b. (You LIKE) (the doggy)!

The fact that grammar exists only to map sound onto meaning, and the fact that children can learn constituent structure and intonation structure at the same time, suggests that the students (and the mothers) are right to believe that these aspects of grammar must be more directly related than traditional accounts would have us believe.

1.8 Explaining Discontinuity

In the end, all theories of grammar can be thought of as consisting of a context-free core defining the level of meaning or logical form, plus some extra machinery to handle long range dependency in surface forms. While theories may differ in the details of the context-free core—for example, on the degree to which it is lexicalized, or the specific form of the language of logical form—all linguistic theories can be considered as essentially equivalent in respect of the context-free core It is in the extra machinery they apply to derive long-range dependencies in the surface form of the language, and in particular the unbounded variety, that the theories differ in interesting ways.

Almost all of the linguistic theories mentioned above and reviewed in Steedman, 2019 take the set of syntactic constituent types comncerned in the derivation of long-range dependency to be the same as the one defined by the context-free core (Wells, 1947). This assumption is understandable, because that is what we were told in our first syntax class, and the intuition that the traditional NP, S, VP, AP, PP and the like are psychologically real is inescapable. However, it is very far from clear that this reality is *syntactic*, rather than semantic, since those constituents are also constituents of linguistic meaning. It is also striking that the traditional tests for syntactic constituency are, as noted earlier, inconsistent and unconvincing (Pesetsky, 1995; Phillips, 2003; Jacobson, 2006). Nowhere is this consensus more questionable than in the case of the VP, where the coordination and intonation tests suggest that “Keats found” is as much a constituent as the traditional VP “found the answer”.

(48) a. Keats found. and Chapman published, the answer.
b. Keats found the answer and published a proof.

The assumption that nothing *else* is a derivational constituent immediately implies that the residues of relativization and right-node raising in examples like (10c,d) cannot be consituents in their own right, but must be traditional constituents of type *S*. To prevent them from behaving like constituents and combining in their own right, they must be marked with a special feature to indicate the presence of a trace or copy, as in the case of the movement theory, where the syntax is responsible for establishing the connection between source and target. Even in G/HPSG and LFG and the related versions of Construction Grammar (CxG, e.g. Boas and Sag, 2012), where hypercyclic feature-passing does the same work within a traditional constituent structure, and in Tree-adjoining Grammar (TAG, Joshi 1985; Joshi and Schabes 1997), where there is a lexical initial tree that includes a moved element and an indexed trace for every extraction from the domain of a verb, distinct from the one with the arguments in canonical position, into which auxiliary trees may be adjoined to "stretch" the dependency unboundedly, in a manner reminiscent of "generalized" or "double-base" transformations (Chomsky, 1955/1975, 1957; Frank, 2006:18).[21]

The details need not detain us at this point, except to note that, in the face of the problem of discontinuity in constructions, assuming the traditional definition of derivational constituency constituency forces the inclusion of displacement in some form *in the rules of syntax themselves*, either as movement, or as G/HPSG hypercyclic trace or gap-feature passing, or LFG "functional uncertainty" and/or control features, or the adjunction mechanism of TAG. This adherence to traditional constituent structure holds even for theories of low expressive power such as GPSG and TAG.

In exploring other possbilities, it is important to keep the theory as low in expressive power as possible, consistent with capturing the degrees of freedom in the discontinuities such as those exemplified above that are actually observed. If our theory is capable of capturing phenomena that we are reasonably sure we will never encounter among real human languages, then we cannot claim to have explained the degrees of freedom in the data that actually are attested.

In this connection, Joshi (1985) and Joshi, Vijay-Shanker, and Weir (1991) proposed a number of properties that should characterize all languages permitted by a theory of grammar if it is to be taken seriously as an explanatory theory of natural languages, a class which he called "Mildly Context-Sensitive", without identifying this class with any automata-theoretic level known at the time. They were the following:

21. An analysis related to TAG seems to be what Chomsky has in mind as the interpretation of "copies" (Chomsky, 2007:6).

(49) *Mild Context-Sensitivity (MCS)*

1. The Context-Free Languages (CFL) are properly contained by the Mildly Context Sensitive Languages (MCSL);

2. All languages in MCSL are parsable in polynomial time;

3. MCSL do not include arbitrary permutation-complete languages;

4. MCSL have the property of "constant growth", such that if their sentences are ordered in terms of length, then two consecutive lengths cannot differ by an arbitrarily large amount.

(The last of these criteria excludes languages like a^{2^n} and $a^{n!}$.)

It is important to understand that these properties do not in themselves identify any specific level of the language hierarchy, intermediate between context-free and context sensitive. There may be many such intermediate levels (Weir, 1988).

Moreover, many of these theories are still very expressive. What we are interested in is the *least expressive* mildly context-sensitive class that will adequately capture the kind of discontinuities discussed in this chapter, which we might distinguish from the larger mildly context-sensitive (MCS) set as 'near-context-free".

In particular, in order to allow for the manifest possibility of language acquisition in children consistent under the semantic bootstrapping assuption laid out above, we shall need to assume a homomorphic relation between syntax and semantics down to the level of the morpho-lexicon, so that the derivation of syntactic types and logical and phonological forms can proceed synchronously and in parallel lock-step. Our watchword will be "no syntax without semantics."

The hypothesis to be explored below is that such a theory can be defined in which, without exception, *all rules of syntax apply to strictly contiguous non-empty constituents*. There are no discontiguous operators, along the lines of movement or TAG adjunction. In such a theory, such derivational residues of relativization, coordination, and intonational phrasing as "I think she found" and "Adlai a train" are first-class citizens of the grammar, with the standing of constituents complete with an interpretation or logical form, free to combine in their own right with other constituents, just as long as they are contiguous to them in the sentence.

In order to do this, we will need to make the following key assumptions:

- **Categorial Grammar**: All constituents are syntactically typed as either functions or arguments.

- **Case**: Counterintuitively, it is the entity-denoting terms such as subjects and objects that are the functions, while the property and relation-denoting terms such as verbs are their arguments.

- **Composition Rules**: Categorial merger is generalized from simple application of functions to contiguous arguments to a small number of "Combinatory" operators, of which *composition* of contiguous functions is the most significant, with the consequence of radically generalizing the classical notion of constituency.

The result will be to reduce the combination of all "displaced" elements with their residues to exactly the same rules of adjacent merger as that of the corresponding "in situ" complements with their heads.

Part I

Categories, Combinators, and Case

Chapter 2

Categorial Grammar

These correspondences [between formal and semantic features] should be studied in some more general theory of languages that will include a theory of linguistic form and a theory of the use of language as subparts.
—*Syntactic Structures* Noam Chomsky, 1957:102

We will assume in what follws a particularly strong form of Chomsky's 1995b; 2001; 2001/2004 "Inclusiveness Condition" on grammar, which says that rules of syntactic derivation cannot add any information such as "indices, traces, syntactic categories or bar levels, and so on" that has not been specified ab initio in the lexicon for the language concerned. This principle entails that all relations between "displaced" elements and their origin must be specified in the lexicon, and be projected unchanged onto the sentences of the language by language-independent universal rules of derivations

2.1 The Categorial Lexicon

In the rest of the book, the categorial notation for lexical entries exemplified for the English transitive verb in (1a) will be used:

(1) a. sees $:= (S\backslash NP_{3s})/NP : \lambda x \lambda y.pres\,(see\,x\,y)$

b. $$\overbrace{\text{sees}}^{\text{phonological form}} := \overbrace{\overbrace{(S\backslash NP\ \underbrace{_{3s}}_{\text{feature}}\ )/NP}^{\text{syntactic type}} : \overbrace{\underbrace{\lambda x \lambda y.}_{\lambda\text{-binders}}\ \ \underbrace{pres\,(see\,x\,y)}_{\text{predicate-argument structure}}}^{\text{logical form}}}^{\text{category}}$$

The category (1a) is anatomized as in (1b). Syntactic types are written in uppercase italic. A syntactic type of the form X/Y (or $X\backslash Y$) denotes something that combines with something of type Y to its right (left) to form an X.[1]

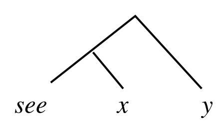
1. We use the "result leftmost" notation of Ajdukiewicz (1935) for syntactic categories because it gives a simpler mapping from syntactic types to logical forms. There is another widely-used

Subscripted feature-values like *3s* specify atomic values or ranges of values for attributes such as tense and agreement which do no real theoretical work in CCG apart from further specifying subcategorization, and are frequently omitted. (In particular, features cannot take unbounded structures as values.) Nevertheless, they are there in the grammar, limiting overgeneration and ambiguity.

The syntactic type of the transitive verb *sees*, $(S\backslash NP_{3s})/NP$, therefore identifies it as something that combines to its right with an NP to yield something with the category of an intransitive verb $S\backslash NP_{3s}$—that is, something that in turn combines to its left with a NP compatible with third person singular agreement to yield a sentence.

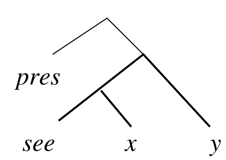
The logical form in (1) is written as a lambda-term, in the body or predicate argument structure of which, as usual, left-to-right juxtaposition denotes the application of a function to an argument under a convention of left associativity. That is, $see\,x\,y$ is equivalent to $(see\,(x))\,(y)$, defining the following structure:

(2)



Such a structure defines a notion of "command" or structural dominance at the level of logical form. Specifically, the predicate argument structure defines the order of application of the logical predicate *see* to its arguments *x* and *y*, such that the second argument *y* "commands" the first argument *x*, in the sense that *y* is attached higher in the argument structure (2) than *x*. The present tense element *pres* then applies to the proposition $see\,x\,y$ to yield the following structure:

(3)



Predicate-argument structures are order-free, in the sense that they represent only dominance relations, not the alignment of their elements with the ordered

"result on top" convention due to Bar-Hillel (1953) and Lambek (1958)—cf. Morrill (2011). The result-leftmost notation is more transparent to cross-linguistic comparison. (For example, the transitive verb is always of the form $(S \mid NP) \mid NP$, in which "|" is a slash whose value is either / or \, regardless of word-order.

strings of the language. It is the syntactic component of the lexical category $(S\backslash NP)/NP$ that defines it as a function applying to its arguments in a fixed order, first to the object *NP* to the right, and then to the subject *NP* to the left.

The binders $\lambda x \lambda y$ to the left of the predicate-argument structure in the logical form then merely express the mapping from the two syntactically aligned arguments to the corresponding two arguments of the predicate *see* in the predicate argument structure.

An important detail to be clear about is that the variables x, y, etc., bound by a λ operator in a logical form are "local" to that logical form. That is, they are distinct from any *other* variable x, y bound by some other λ in some other logical form. The locality of variable binding means that, for example, we can use the same identifiers x and y for the arguments of *every* transitive verb.

2.2 Combining Categories I: Application

Functors combine with arguments via the forward and backward rules of function application (4):

(4) MERGE I: THE APPLICATION RULES
   a. Forward Application:
   $$X/_{\star}Y : f \quad Y : a \quad \Rightarrow \quad X : f\,a \qquad (>)$$
   b. Backward Application:
   $$Y : a \quad X\backslash_{\star}Y : f \quad \Rightarrow \quad X : f\,a \qquad (<)$$

The $\star$ annotation on the slashes in rules (4) is one of a number of slash-types or "modalities" which can be used via the lexicon to limit the rules by which categories may combine. (For example, these will turn out to be the *only* rules by which the conjunction category for "and", $(\mathrm{X}\backslash_{\star}\mathrm{X})/_{\star}\mathrm{X}$, discussed in chapter 11 can apply.) We defer further discussion of slash-typing until chapter 9, since all categories in the examples in the present chapter and the next are unconstrained, and can combine by any rule, including the above.

Like all rules of syntactic derivation in CCG, the Application rules (4) are subject to the following Condition

(5) *The Combinatory Projection Principle (CPP)*
Syntactic combinatory rules are binary linearly ordered type-dependent rules, applying to string-adjacent categories, consistent with their directional types and linear order, and must project unchanged onto the result category the type and directionality of any argument of the input categories that also appears in the result.

This principle is defined more formally in *SP* in terms of three more fundamental principles of Adjacency, Directional Inheritance, and Directional Consistency, which collectively forbid rules like (a), (b), and (c), as indicated by the non-reduction symbol "$\not\Rightarrow$"

(6) a. $Y: a \quad X/Y: f \quad \not\Rightarrow \quad X: fa$

b. $(X/Y)/Z: f \quad Y: a \quad \not\Rightarrow \quad X/Z: fa$

c. $(X/Y)/Z: f \quad Z: a \quad \not\Rightarrow \quad X\backslash Y: fa$

The Combinatory Projection Principle (5) rules out (6a) because it has a rightward function combining to its left, and rules out (6b) because it has the second argument of a function combining before its first argument, an operation of the general class that has been proposed under other categorial approaches under the name of WRAP (Bach, 1976; Dowty, 1979a), but is disallowed under the present interpretation of adjacency. Rule (6c) is disallowed because it switches the directionality of the *Y* argument. We shall see in later sections that this principle limits all rules of syntax, and is the source of the low "near context-free" expressive power of the present theory.[2]

The sub-principles of Directional Consistency and Inheritance are simply corollaries of the Inclusiveness Condition, which says that derivational rules cannot override and must project the relations of Linear Precedence specified in the lexicon. The sub-principle of Adjacency extends this stricture to the relations of Immediate Dominance specified there.

At this point we have a choice: when (say) a verb and its complement(s) combine by one of these rules, we could either choose the verb to be the functor and the complements as its arguments, or we could define the arguments as functors and verbs as their arguments

Since we have defined verbs as functors via categories like exrefex:sees, there is a natural temptation to make the former assumption, giving rise to

2. The above is a stronger interpretation of the Combinatory Projection Principle than is assumed in some earlier publications.

derivations like the following:

(7)
$$\begin{array}{ccc}
\text{I} & \text{saw} & \text{Esau.}\\
\overline{NP_{1s}} & \overline{(S\backslash NP_{agr})/NP} & \overline{NP}\\
: me & : \lambda x \lambda y.past\,(see\,x\,y) & : esau\\
 & \multicolumn{2}{c}{\overline{S\backslash NP_{agr} \quad : \lambda y.past\,(see\,esau\,y)}{>}}\\
\multicolumn{3}{c}{\overline{S : past\,(see\,esau\,me)}{<}}
\end{array}$$

(By convention, CCG derivations are shown in the accepting direction, with the lexical leaves or terminals on top. Combination is indicated by underlining decorated with the relevant combinatory rule for easier exposition, e.g. (>) for the last derivation line in (7), in which the material covered by the rule is *(I)* and *(saw Harry)*, as shown by the ranges of the underlines.)

However, assuming that entities like subjects and objects are arguments does not work well semantically for NP complements in general. In particular it fails to assign the right semantic scope to quantified subjects and objects:

(8)
$$\begin{array}{ccc}
*\text{I} & \text{saw} & \text{every boy.}\\
\overline{NP_{1s}} & \overline{(S\backslash NP_{agr})/NP} & \overline{NP}\\
: me & : \lambda x \lambda y.past\,(see\,x\,y) & : \forall x.[boy\,x]\\
 & \multicolumn{2}{c}{\overline{S\backslash NP_{agr} \quad : \lambda y.past\,(see\,(\forall x[boy\,x])\,y}{>}}\\
\multicolumn{3}{c}{\overline{S : past\,(see\,(\forall x[boy\,x])\,me)}{<}}
\end{array}$$

If $past\,(see\,(\forall x[boy\,x])\,me$ means anything, it means that *I saw that everything was a boy*.

What we want as a meaning for (9) is $\forall x[boy\,x \Rightarrow past\,(see\,x\,me)]$ (that is, *everything is such that, if its a boy, I saw it* or perhaps *for everything of type boy, I saw it*. That reading can be obtained directly if we adopt the other assumption, and make all NPs be (second-order) functors over verbs, capturing the implicative relation between boys and seeing in derivations like the following:

(9)
$$\begin{array}{ccc}
\text{I} & \text{saw} & \text{every boy.}\\
\overline{S/(S\backslash NP_{1s})} & \overline{(S\backslash NP_{agr})/NP} & \overline{(S\backslash NP)\backslash((S\backslash NP)/NP)}\\
: \lambda p.p\,me & : \lambda x \lambda y.past\,(see\,x\,y) & : \lambda p.\forall x.[boy\,x \Rightarrow p\,x]\\
 & \multicolumn{2}{c}{\overline{S\backslash NP_{agr} \quad : \lambda y.\forall x.[boy\,x \Rightarrow past\,(see\,x\,y)]}{<}}\\
\multicolumn{3}{c}{\overline{S : \forall x.[boy\,x \Rightarrow past\,(see\,x\,me)]}{>}}
\end{array}$$

The assumption in the above derivation is that *all* NPs bear order-preserving "type-raised" categories of functions over verbs. Apart from the syntactic type and logical forms of the subject and object, the only diffence between the derivations (8) and (9) is that the direction of the two applications has been reversed—in other words the raised categories of the subject and object are *order-preserving*. Since the raised categories define the NP as a particular argument of the verb, such as subject or object, we identify each such category with a grammatical *case*, such as nominative or accusative, as if English were a cased language like Latin or Japanese. Despite the lack of morphology in English, we assume that case/type-raising is an essentially morpho-lexical process, as it is in those languages, rather than a rule of syntactic derivation.

NPs inherit such categories from the lexical entries for their heads/specifiers—that is, determiners like "every". For example, the object "every boy" in (9) is derived as follows:

(10)
$$\begin{array}{cc}
\text{every} & \text{boy} \\
\hline
((S\backslash NP)\backslash((S\backslash NP)/NP))/N & N \\
:\lambda n\lambda p.\forall x.[n\,x \Rightarrow p\,x] & :\lambda x.boy\,x \\
\hline
\multicolumn{2}{c}{(S\backslash NP)\backslash((S\backslash NP)/NP)} \\
\multicolumn{2}{c}{:\lambda p.\forall x.[boy\,x \Rightarrow p\,x]}
\end{array}{>}$$

Because English is in this respect lexically ambiguous as to case, and because the logical form of raied NPs is the same across cases, it will often be convenient to abbreviate raised syntactic types as $NP^{\uparrow}$, denoting "whichever raised type the derivation requires", writing the above rather formidable derivation more readably as follows:[3]

(11)
$$\begin{array}{cc}
\text{every} & \text{boy} \\
\hline
NP^{\uparrow}_{3s}/N_{3s} & N_{3s} \\
:\lambda n\lambda p.\forall x.[n\,x \Rightarrow p\,x] & :\lambda x.boy\,x \\
\hline
\multicolumn{2}{c}{NP^{\uparrow}_{3s}} \\
\multicolumn{2}{c}{:\lambda p.\forall x.[boy\,x \Rightarrow p\,x]}
\end{array}{>}$$

Verbs can of course take categories other than NP as complements, including S, inducing recursion into the syntax and semantics:

3. In fact, such underspecification is routinely built into parsers for CCG.

(12)

$$\begin{array}{cccccc}
\text{I} & \text{believe} & \text{I} & \text{saw} & \text{every} & \text{boy.} \\
\hline
NP^{\uparrow}_{1s} & (S\backslash NP_{agr})/S & NP^{\uparrow}_{1s} & (S\backslash NP_{agr})/NP & NP^{\uparrow}/N & N \\
:\lambda p.p\,me & :\lambda s\lambda y.believe\,s\,y & :\lambda p.p\,me & :\lambda x\lambda y.past\,(see\,x\,y) & :\lambda n\lambda p.\forall x[n\,x \Rightarrow p\,x] & :\lambda x.boy\,x
\end{array}$$

$$\overline{}^{>} \quad NP^{\uparrow} : \lambda p.\forall x[boy\,x \Rightarrow p\,x]$$

$$\overline{}^{<} \quad S\backslash NP_{agr} : \lambda y.\forall x[boy\,x \Rightarrow past\,(see\,x\,y)]$$

$$\overline{}^{>} \quad S : \forall x[boy\,x \Rightarrow past\,(see\,x\,me)]$$

$$\overline{}^{>} \quad S\backslash NP_{agr} : \lambda y.believe\,(\forall x[boy\,x \Rightarrow past\,(see\,x\,me)])\,y$$

$$\overline{}^{>} \quad S : believe\,(\forall x[boy\,x \Rightarrow past\,(see\,x\,me)])\,me$$

Clearly, the theory presented so far is equivalent to context-free grammar (CFG), with the λ-calculus merely acting as a "glue-language" putting together distinct but equally context-free simple predicate-argument structural logical forms synchronously with syntactic derivation. In comparison with a traditional context-free phrase-structure grammar like (4) of chapter 1, all that we have done is trade an increase in the number and specificity of lexical types for a decrease in the number and specificity of syntactic rules, for example replacing V by language-specific categories like $S\backslash NP$, $(S\backslash NP)/NP$, $(S\backslash NP)/S$, etc., and replacing language-specific production rules like $S \rightarrow NP\ VP$ by universal rules of functional application like $X/Y\ Y \Rightarrow X$.

The transitive verb such as (1) is representative of a number of verbal function types or *subcategorization* frames, which under standard linguistic definitions of argument may have a valency of up to three.[4]

Such syntactic function-types are always binarized or "Curried": they take one argument, and yield a binarized function over any remaining arguments.[5]

For example, the following is the category for an "object control" verb for a sentence such as *He persuaded her to leave*:[6]

4. More pragmatic traditions like those used in annotating the Penn Treebank may allow somewhat higher valencies, where what are here regarded as adjucts such as ethic datives.

5. Schönfinkel (1924) showed that Curried functions support exactly the same class of computations as *n*-ary ones.

6. The logical form is simplified as usual.

(13) a. persuaded $:= ((S\backslash NP_{3s})/VP_{to})/NP : \lambda x\lambda p\lambda y.past\,(persuade\,(p\,x)\,x\,y)$

b. $$\overbrace{\text{persuaded}}^{\text{phonological form}} := \overbrace{\overbrace{((S\underbrace{_{fin}}_{\text{feature}}\backslash NP\underbrace{_{3s}}_{\text{feature}})/VP\underbrace{_{to}}_{\text{feature}})/NP}^{\text{syntactic type}} : \overbrace{\underbrace{\lambda x\lambda p\lambda y.}_{\lambda\text{-binders}}\underbrace{past\,(persuade\,(p\,x)\,x\,y)}_{\text{predicate-argument structure}}}^{\text{logical form}}}^{\text{category}}$$

Verbs like (13a) also combine with their arguments by the application rules (4), as in the following derivation:

(14)
$$\begin{array}{ccccc}
\text{Keats} & \text{persuaded} & \text{Chapman} & \text{to} & \text{go} \\
\hline
NP^{\uparrow} & ((S\backslash NP_{3s})/VP_{to})/NP & NP^{\uparrow} & VP_{to}/VP & VP \\
: keats & : \lambda x\lambda p\lambda y.past\,(persuade\,(p\,x)\,x\,y) & : chapman & : \lambda p.p & : \lambda y.go\,y \\
\hline
 & \multicolumn{2}{c}{(S\backslash NP_{3s})/VP_{to} : \lambda p\lambda y.past\,(persuade\,(p\,chapman)\,chapman\,y)} <& \multicolumn{2}{c}{VP_{to} : \lambda y.go\,y} > \\
\hline
 & \multicolumn{4}{c}{S\backslash NP : \lambda y.past\,(persuade\,(go\,chapman)\,chapman\,y)} > \\
\hline
\multicolumn{5}{c}{S : past\,(persuade\,(go\,chapman)\,chapman\,keats)} >
\end{array}$$

The derivation computes the meaning $past\,(persuade\,(go\,chapman)\,chapman\,keats$, restoring continuity between “go” and the object of “persuaded”. Nevertheless the derivation consists entirely of combinations of adjacent functions and arguments. The apparent discontinuity is baked into the lexical logical form (13a) via the variable p, which we noted earlier is second-order, taking a function as its value, which is applied to the value x of the object of *persuaded*. at the level of argument structure, which is independent of linear order. Linear order is defined by the syntactic category, and linged to argument structure by the λ-binding.[7]

Unlike control verbs, in which the nominal argument seems to have two distinct lf roles, as both the object of the main verb (such as “persuade” in (14)), and as the subject of an infinitival complement (such as “to go”), reflected in two occurrences of the bound variable x, the subject of a raising verb like “seems” seems to have a single role as the subject of the infinitival complement. For that reason, it is invariably talked of as involving movement of the complement subject to the the subject position of “seems”.

However, the Inclusiveness Condition requires that such a displacement be

7. The present predicate-argument structure resembles the ARG-ST terms of HPSG and the “grammatical function tier” of SimSyn in not committing to a fixed repertoire of thematic role labels (cf. Dowty, 1991b). Indeed, Landau (2001, 2015) shows that ther are a number of semantically distinct families of control verbs, with each of which the present underspecified lgical forms are compatible.

defined in the lexicon, rather than established dynamically, as a side-effect of the derivation. We can lexicalize the observation as follows:

(15) a. seems $:= (S\backslash NP_{3s})/VP_{to} : \lambda p\lambda y.pres\,(seem\,(p\,y))$

b. $$\overbrace{\text{seems}}^{\text{phonological form}} \;:=\; \overbrace{\overbrace{(S\backslash NP_{\underbrace{3s}_{\text{feature}}})/VP_{to}}^{\text{syntactic type}} : \overbrace{\underbrace{\lambda p\,\lambda y.}_{\lambda\text{-binders}}\;\underbrace{pres\,(seem\,(p\,y))}_{\text{predicate-argument structure}}}^{\text{logical form}}}^{\text{category}}$$

(16)

a	unicorn	seems	to	be	approaching
$NP^{\uparrow}/N$ $: \lambda n\lambda p.p\,(a\,n)$	N $: unicorn$	$(S\backslash NP_{3s})/VP_{to}$ $: \lambda p\lambda y.pres\,(seem\,(p\,y))$	VP_{to}/VP $: \lambda p.p$	VP/XP_{pred} $\lambda p\lambda y.p\,y$	VP_{ing} $: \lambda y.approach\,y$
$NP^{\uparrow} : \lambda p.p\,(a\,unicorn)$ (>)				$VP : \lambda y.approach\,y$ (>)	
			$VP_{to} : \lambda y.approach\,y$ (>)		
		$S\backslash NP_{3s} : \lambda y.pres\,(seem\,(approach\,y))$ (>)			
$S : pres\,(seem\,(approach\,(a\,unicorn)))$ (>)					

Rather than giving rise to a cascade of lf roles, raising leaves the complement subject in situ at LF, via the variable y, which the raised nominative subject gives a value via the binder λy, and which its syntactic type aligns to the left of the tensed verb. This gives the appearence of discontinuity, but the derivation is via entirely contiguous application merger, exactly parallel to that in the previous derivations (9) and (14).

We should note in passing concerning derivation exrefex:seemstobe that, while "a unicorn" has the syntactic category $NP^{\uparrow}$ of a generalized quantifier, its logical form does not include a classical existential quantifier $\exists$ but is rather an underspecified Skolem term *a unicorn* which may either be an unbound Skolem constant unicorn (the "de re" reading, which commits the speaker to the existence of the unicorn), or may become bound by an intensional operator associated with *seem* (the "de dicto" reading, which does not.) It is only the true universal quantifier determiners like "every" that introduce classical quantifiers.[8]

The above lexicalization of phenomena of control and binding crucially depends on the verbal heads of those constructions selecting syntactically and se-

8. We defer discussion of the mechanism by which Skolem terms get bound to operators until chapter **??**.

mantically for properties, or functions of type $e \rightarrow t$ such as *VP*represented in lexical logical form by bound variables like *p*—that is to say, on the availability of second-order functions in the theory. We must assume that the involvement of second-order variables like *p* denoting properties such as VPs reflects their presence as primitives of the universal language of mind that was claimed earlier to underpin the child's ability to learn lexical categories like (13a). But there is no evidence for the involvement of *third-* or higher-order variables—that is, variables whose value is a second-order function like *persuade*.

As a consequence, categories like (13) automatically obey a minimality condition on relations between verbs and their clausal arguments that used under the movement theory to be called "subjacency" (Chomsky, 1981). In present terms, this condition expresses the observation that in the attested natural languages, a matrix verb like *persuade*, may bind an argument of its infinitival complement, here the variable *p*, to one of its own arguments, here the variable *y*. But we never see a matrix verb binding an argument of any more embedded verb—say, an argument of the complement of *p*. Such a contingency woulr require a third-order variable. Subjacency is an empirically-observed condition on possible lexical categories, excluding monsters like the following "super-control" verb, in which λq is the binder for the controlled VP argument of a controlled VP:

(17) *foo := $(S\backslash NP)/(VP/VP) : \lambda p \lambda q \lambda y.foo'(p(q\,y)\,y)\,y)$

2.2.1 "Abstract" Case

Cases other than the nominative require multiple raised categories in both morphologically and structurally-cased languages. For example, the phenomenon of subject "pro-drop"—which in English is confined the first and second person subjects, but in other languages like Hindi is perfectly general–mean that *non*-subject NPs need more than one category—for example, we have:[9]

(18)
$$\dfrac{\dfrac{\text{Missed}}{S/NP : \lambda x.missed\,x\,me} \quad \dfrac{\text{the Saturday Dance.}}{S\backslash(S/NP) : \lambda p.p\,saturdaydance}}{S : missed\,saturdaydance\,me}<$$

(We will assume that such pro-drop verb categories, including a pronominal subject in their lf predicate-argument structure, but with no corresponding syn-

9. The Inclusiveness Condition as realized in the Combinatory Projection Principle (5) forbids any acclount of pro-drop in terms of introducing an inaudible pronoun to act as the subject in the derivation, as opposed to its introduction in lexical logical form. For the sake of simplicity, we assume here that "the Saturday Dance" translates as the proper name *saturdaydance*$'$

tactic argument or λ-binder, are derived from the basic form by a lexical rule for all verbs for those proforms which support drop in all languages.)

We shall see later that the type-polymorphism for accusatives assumed in (18) also does crucial work elsewhere in the grammar of English, in particular in allowing right-node-raising and argument-adjuct cluster coordination in chapter 11 We shall also see in chapter **??** that it is also responsible for the phenomenon of "scrambling" in freer argument-order languages like Japanese and German.

In particular, even though such languages typically carry relatively unambiguous morphological case-markers, those cases apply to verbs of various syntactic valencies. They must therefore bear multiple categories, type-raised over those verb-categories, so that even in cased languages like Latin and Japanese, morphological case-markers are typically ambiguous (or equivalently underspecified).

It follows that in identifying type-raising with nominal case, we are embracing the idea of "abstract" case, (Legate, 2008; Bobaljik and Wurmbrand, 2009), divorced from any fixed relation to semantic or thematic role, which remains the responsibility of the verb itself. According to this theory of case, it is a coincidence that the subject of the intransitive and of the transitive are marked in Latin by the same case. That is, we could define a language just like Latin, apart from marking the intransitive subject with the same morphological case as the transitive object.

Such languages exist, and are known as "ergative" languages, in which the case of the intransitive subject and the transitive object is the "absolutive" case, and are contrasted with "accusative" languages like Latin, Japanese, and English. Yup'ik Bok-Bennema (1991) is such a language:

(19) a. Arnaq yurar-tuq
woman.ABS dance.IND.3S
"The/a woman dances."

b. Angutem tangrr-aa arnaq
man.ERG see.IND.3S.3S woman.ABS
"The/a man sees the/a woman."

These examples suggest the following morpho-lexical categories, possibly among others, for absolutive "arnaq" in Yup'ik:

(20) arnaq := $S/(S\backslash NP_{abs,3s}) : \lambda p.p\,woman$
$S\backslash NP_{erg}/(S\backslash NP_{erg})/NP_{abs,3s}) : \lambda p.p\,woman$

Of course, it does not follow that Yup'ik speakers think about transitive and intransitive events any differntly from English speakers. Subject like the door in English "unaccusative" sentences like "The door opened" are as their name suggests semantically patients, rather than agents, despite bearing the same structural case as agents of transitives.

Languages are clearly (somewhat) free as to whether they assign the subject of the intransitive the case of the more agent-like or patient-like of the arguments of the transitive, althought there is a clear bias towards the former, presumably because of its more salient commanding level at the level of lf predicate-argument or "thematic" structure (cf. exrefex:lexitem). Seen in this light, there is nothing more remarkable in the fact that egative languages assign the case of the transitive patient to the subject of "unergative" intransitives like "yurar-" ("dance") than the fact that the accusative language English assigns the case of the transitive agent to the subject of "unaccusative" intransitives like "open".

One might suspect on this basis that there is a pressure on languages to use the same case pattern on all verbs of a given valency, such as intransitives and transitives, etc. However, there exist "split" ergative languages like Dyirbal, in which certain classes of nominal such as pronouns carry accusative pattern case-marking (Dixon, 1972; Nordlinger, 1998:75). Similarly, languages like Icelandic are free to specify "quirky" morphological case, so that accusatives and datives have to bear the category of the subject, in addition to those of the object etc., in order to combine with certain verbs that specify those cases on the syntactic subject, reflecting semantics, the history of the language, or both (Butt, 2006; Baker, 2015). In particular, many of the verbs that in English are referred to as "unaccusative" intransitives such as "bjróta" ("break") take "quirky" accusative subjects (Zaenen and Maling, 1990). Similarly, certain Icelandic transitive verbs take morphologically accusative, dative, or genetive subjects (Thráinsson, 2007:181). For example:

(21) Þeim likar maturinn
Them.DAT likes.SG food-the.NOM
"They like the food"

Of cousre, it does not follow that speakers of Icelandic, any more than those of Yup'ik, think about people liking food any differently from English speakers

The analysis of case in terms of morpho-lexical type-raising has some resemblance to the LFG analysis of "constructive case" (Butt and King, 1991; Nordlinger, 1997). It is worth noticing at this point that the effect of case when interpreted in this way is to turn arguments such as NP subjects and objects into something very like *adjuncts* to whatever category specifies them as an argument. Jelinek, 1984:44, *passim* claims that the status of NPs as (optional) adjuncts is characteristic of non-configurational languages. CCG embodies the claim that NP arguments are quasi-adjuncts in *all* languages, and that their optionality in nonconfigurational languages is linked to paratactic properties of those languages, such as pro-drop, rather than adjuct-hood itself. NOORDLINGER EXAMPLE HERE?

If this wide degree of variation and ambiguity seems confusing, it is worth recalling again that Type-raising is in CCG (as opposed to Type Logical Grammar (Moortgat, 1988; Morrill, 1994, 2011) and some other generalizations of categorial grammar using the related notion of *continuation* (Barker and Shan, 2014)) a strictly morpholexical operation, rather than an operation of projective syntax and semantics. It is therefore subject to lexical processes like "bleaching" of thematic role and freezing of archaic forms as morpho-lexical "irregularities". It can also only apply where A is an elementary argument type such as NP or VP. While T can itself be a raised type, as we shall see in the case of pied-piping relatives and "roll-up" extraction), A cannot be a raised type. This is not contradicted by the existence of "case-stacking" languages, in which multiple case markers act to disambiguate scope, case, or agreement, rather than as distinct cases on the same argument (Plank, 1995; Schweiger, 2000; Nordlinger, 1997).

Such morpholexical operations are strictly subject to the combinatory projection principle, (5), and do not override the lexical directionality of the verbal categories they are raised over. While we shall see categories such as topics and *wh*-elements with similar second-order types representing displaced elements, they will have to change the syntactic type of their result, marking it as S_{top}, S_{whq}, and the like.

2.3 CCG and the Minimalist Program

There is a close relation between CG categories and derivations of the kind seen in the present chapter and the minimalist notions of "Bare Phrase Structure" (which eliminates phrase structure rules in favor of head-projection Chomsky, 1995a), and to a lesser extent "Phase" (which defines a domain of locality for movement with similar effects to the transformational cycle Chomsky, 2001). In particular, despite the fact that the generative approach makes a different division of responsibility between lexical types and rules of syntactic derivation, the core CCG lexical logical forms labeled "predicate-argument structure" in (1) and (13) seems to correspond quite directly to the minimalist phasal vP, including rather obviously conforming to the "predicate-internal subject hypothesis" (PISH) of Fukui (1986, 1995), differing only in being unordered (that is, unlinearized), and having the effects of operations like "A-movement" and "Head-movement" compiled into logical form via λ-bound variables. In particular, we saw in derivation (**??**), *every dog barks*, that logical operators such as modals, negation, and quantifiers take their scope at the *edge* of the predicate argument structure, as they do in the minimalist proposal of Johnson (2000). However, levels corresponding to IP/TP are added by tense morphology, as in section 3, while others like modality, negation, and CP are in English added by independent lexical elements—in the case of the latter, by complementisers and relativisers, limited in their application by minor features such as S_{fin}.

In the case of Welsh tense morphology, we saw in (5) that it has the effect of minimalist Head-movement to the extent that it specifies VS linear order for the finite sentence, although it does not in any sense involve movement, since the nonfinite verb stem is not specified for subject linearization.[10]

It follows, according to the present proposal, that *every* lexical governor such as a verb defines a domain of locality for such operators to scope over. The Phase Impenetrability Condition (PIC, Chomsky, 2001) follows from the fact that syntactic category and syntactic combinatory rules are strictly type-dependent and entirely blind to predicate argument structure. No PIC need separately be stipulated.

The syntactic type in categories like (1) and (13), together with the λ-binders, then defines the mapping of predicate argument structure onto linearized surface derivations such as (7) and (**??**), via the morphology of examples (2) and (5), and by combinatory rules such as (4), which correspond directly in minimalist terms to merger. No finer distinction between phasal and non-phasal nodes is needed. In fact, the role of CP is greatly reduced in CCG, which will be seen in chapter (9) to avoid any idea of "cycle" in its analysis of the phenomena that fall under the heading of $\overline{\text{A}}$-movement in Minimalism.

The close relation of CCG to the "Bare Phrase Structure/Derivation by Phase" instance of the Minimalist Program as extended in Chomsky, 1995a, 2000, 2001 (and to the related Pregroup Grammars of Lambek, 2001) should

10. This may or may not be consistent with Chomsky's 2001 suggestion that Head movement should not be included in "Narrow Syntax", which seems to mean it occurs after "Spellout" or lexical insertion.

be clear. Lexical categories like that of "works", $S\backslash NP_{3s}$, and "saw", $(S\backslash NP_{agr})/NP$, are comparable to lexical categories in "Bare Phrase Structural" Minimalism (Chomsky, 1995a, 2001), such as the following, in which "*u*N" (for "uninterpretable N-feature") takes the place of both "$/NP$" and "$\backslash NP$" (Adger, 2003: 86):

(22) work [V, *u*N] ("yields V; selects N")

(23) see [V, *u*N, *u*N] ("yields V; selects two N")

"Uninterpretable features" such as *u*N must be "checked" against or "canceled" by matching "interpretable features" such as N, carried by their their arguments, a process which corresponds to matching of $/NP$ and $\backslash NP$ under function application in the earlier derivations (**??**) and (**??**).

In particular, this category allows a derivation isomorphic to the rejected CCG derivation (7) for *I saw Esau*.

All CCG derivations also necessarily conform to a particularly strong form of the Projection Principle or "Inclusiveness Condition" (Chomsky, 1995b: 228, 2001: 2, 2001/2004: 109), in that derivations add no information such as "indices, traces, syntactic categories or bar-levels and so on" that has not already been specified in the lexicon. Minimalism can therefore be seen as Categorial Grammar with the addition of discontinuity in rules such as movement/internal merge (Berwick and Epstein, 1995a,b; Adger, 2003, 2013; Smith and Cormack, 2015). The most important difference is that CG specifies the equivalent of bar-level of arguments as NP, N', N, etc. *in the lexicon*, avoiding the use of a structure-dependent "labeling algorithm" (Chomsky, 2008, 2013), and includes *linearization* information in language-specific lexical categories via the slash notation, specifying that, in English, subjects are found to the left, and objects to the right, avoiding "head movement". It should be noted that this is a quite different interpretation of slashes to the one used in GPSG, and defines *selection* rather than extraction *per se* (Gazdar, 1981: 159).[11]

We shall see later that the usual argument from the existence of free argument order languages such as Japanese for leaving linearization unspecified in the lexicon and attempting to derive it from universal principles such as Kayne's 1994 Linear Coherence Axiom is obviated by the involvement of case

11. The categorially-influenced Minimalist Grammars of Harkema (2001) and Torr (2019); Torr, Stanojević, Steedman, and Cohen (2019) also lexicalize linearity. Linearization and linearization of categories and rules is a source of strength in the theory presented below, for example in predicting the dependency of island effects and deletion under coordination on basic word-order (Ross, 1967, 1970).

in the form of type-raising.

Rather than the minimalist notion of phase being identified with that of *morpholexically defined domain*, it is in MG defined structurally, in terms of Phase-bounding nodes v and C, along with the Phase Impenetrability Condition (PIC).

Minimalist grammars can therefore be seen as adding movement and its attendant constraints to a form of categorial grammar restricted to first order functions over atomic types, while CCG is full second-order CG, with the addition of a few strictly adjacent combinatory rules.

The rest of the book explores the consequences of these fundamental differences for various kinds of construction.

Exercise : Turn the context free phrase structure grammar (4) of chapter 1 into an equivalent categorial lexicon. (Hint: you can make tensed verbs like "met" lexical items. You don't have to spell out the morphology unless you want to). Test your grammar by doing a derivation. Then add logical forms to the lexicon. Test again. Then extend your grammar to cover the passive. (Hint: you will need another lexical entry for "met" as passive participle.) Finally, add a logical form semantics for the passive.