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

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 *asdzą́ą́*, "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 *dilyį́hí*, "lead" is literally "that which melts", derived from the verb *dilyį́h*, "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) $\textit{Tuesday} \Rightarrow \textit{Belgium}$

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

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 $(\mathit{1}+\mathit{2})*(\mathit{3}+\mathit{4})$ as sums with values like 21

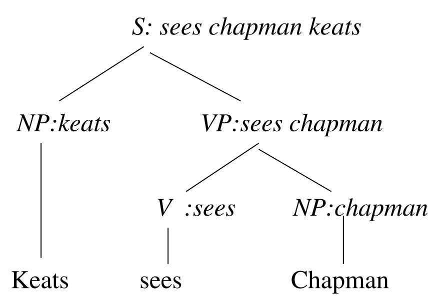
(3)
$$\begin{array}{lll} S : o\,s_1\,s_2 & \rightarrow & (S : s_1 \quad O : op \quad S : s_2) \\ S : n & \rightarrow & N : n \\ O : o & \rightarrow & \{+ : add, - : sub, * : mul, \div : div\} \\ N : n & \rightarrow & \{1 : 001, 2 : 010, 3 : 011, 4 : 100, \ldots\} \end{array}$$

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)
$$\begin{array}{lll} S : v\,n & \rightarrow & NP : n \quad VP : v \\ VP : v & \rightarrow & V_{\mathit{1}} : v \\ VP : v\,n & \rightarrow & V_2 : v \quad NP : n \\ \ldots & & \\ NP : n & \rightarrow & \{\text{Keats} : keats, \text{Chapman} : chapman \ldots\} \\ V_{\mathit{1}} : v & \rightarrow & \{\text{walks} : walks, \text{talks} : talks, \ldots\} \\ V_2 : v & \rightarrow & \{\text{sees} : sees, \text{forgets} : forgets, \ldots\} \\ \ldots & & \end{array}$$

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 Vs. 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. $father\,Esau = Isaac$
b. $father\,Isaac = Abraham$
c. $grandfather = \lambda x.father\,(father\,x)$
d. $grandfather\,Esau = Abraham$

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 $\mathbf{B}$. The functor *grandfather* from (8b) can be defined in either combinatory notation as follows:[9]

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

Both are equivalent to the right-hand side of (8b), $\lambda x.father\,(father\,x)$. 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 $\lambda x.father\,x$. 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 $\mathbf{B}\,father\,father$ is equivalent to $(\mathbf{B}\,father)\,father$.

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) $wants\,(try\,(begin\,(write\,(a\,play)\,chapman)\,chapman)\,chapman)\,chapman$

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. $(\lambda x.say(think(saw\,x\,you)\,you)\,you)\,who$

"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 $say(think(saw\,who\,you)\,you)\,you)$.

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

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

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

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]

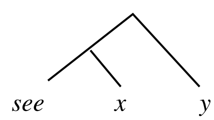
1. We use the "result leftmost" notation of Ajdukiewicz (1935) for syntactic categories because it gives a simpler mapping from syntactic types to logical forms. There is another widely-used

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.

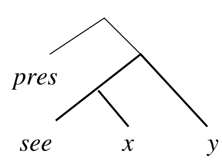
The logical form in (1) is written as a lambda-term, in the body or predicate argument structure of which, as usual, left-to-right juxtaposition denotes the application of a function to an argument under a convention of left associativity. That is, $see\,x\,y$ is equivalent to $(see\,(x))\,(y)$, defining 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 $\backslash$, 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)
$$\begin{array}{cccl}
\text{I} & \text{saw} & \text{Esau.} & \\
\cline{1-1}\cline{2-2}\cline{3-3}
NP_{\mathit{1s}} & (S\backslash NP_{agr})/NP & NP & \\
: me & : \lambda x\lambda y.past\,(see\,x\,y) & : esau & \\
\cline{2-3}
 & \multicolumn{2}{c}{S\backslash NP_{agr}} & > \\
 & \multicolumn{2}{c}{: \lambda y.past\,(see\,esau\,y)} & \\
\cline{1-3}
\multicolumn{3}{c}{S : past\,(see\,esau\,me)} & <
\end{array}$$

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

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

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)
$$\dfrac{\dfrac{\text{every}}{((S\backslash NP)\backslash((S\backslash NP)/NP))/N \; : \lambda n \lambda p.\forall x.[n\,x \Rightarrow p\,x]} \quad \dfrac{\text{boy}}{N \; : \lambda x.boy\,x}}{(S\backslash NP)\backslash((S\backslash NP)/NP) \; : \lambda p.\forall x.[boy\,x \Rightarrow p\,x]}>$$

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)
$$\dfrac{\dfrac{\text{every}}{NP^{\uparrow}_{3s}/N_{3s} \; : \lambda n \lambda p.\forall x.[n\,x \Rightarrow p\,x]} \quad \dfrac{\text{boy}}{N_{3s} \; : \lambda x.boy\,x}}{NP^{\uparrow}_{3s} \; : \lambda p.\forall x.[boy\,x \Rightarrow p\,x]}>$$

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)
$$\begin{array}{cccccc|c}
\text{I} & \text{believe} & \text{I} & \text{saw} & \text{every} & \text{boy.} & \\
\hline
NP^{\uparrow}_{1s} & (S\backslash NP_{agr})/S & NP^{\uparrow}_{1s} & (S\backslash NP_{agr})/NP & NP^{\uparrow}/N & N & \\
:\lambda p.p\,me & :\lambda s\lambda y.believe\,s\,y & :\lambda p.p\,me & :\lambda x\lambda y.past\,(see\,x\,y) & :\lambda n\lambda p.\forall x[n\,x\Rightarrow p\,x] & :\lambda x.boy\,x & \\
\cline{5-6}
 & & & & \multicolumn{2}{c}{NP^{\uparrow}} & > \\
 & & & & \multicolumn{2}{c}{:\lambda p.\forall x[boy\,x\Rightarrow p\,x]} & \\
\cline{4-6}
 & & & \multicolumn{3}{c}{S\backslash NP_{agr}} & < \\
 & & & \multicolumn{3}{c}{:\lambda y.\forall x[boy\,x\Rightarrow past\,(see\,x\,y)]} & \\
\cline{3-6}
 & & \multicolumn{4}{c}{S} & > \\
 & & \multicolumn{4}{c}{:\forall x[boy\,x\Rightarrow past\,(see\,x\,me)]} & \\
\cline{2-6}
 & \multicolumn{5}{c}{S\backslash NP_{agr}} & > \\
 & \multicolumn{5}{c}{:\lambda y.believe\,(\forall x[boy\,x\Rightarrow past\,(see\,x\,me)])\,y} & \\
\cline{1-6}
\multicolumn{6}{c}{S : believe\,(\forall x[boy\,x\Rightarrow past\,(see\,x\,me)])\,me} & >
\end{array}$$

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

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

(16)
$$\begin{array}{cccccc}
\text{a} & \text{unicorn} & \text{seems} & \text{to} & \text{be} & \text{approaching}\\
\hline
NP^{\uparrow}/N & N & (S\backslash NP_{3s})/VP_{to} & VP_{to}/VP & VP/XP_{pred} & VP_{ing}\\
:\lambda n\lambda p.p\,(a\,n) & :unicorn & :\lambda p\lambda y.pres\,(seem\,(p\,y)) & :\lambda p.p & \lambda p\lambda y.p\,y & :\lambda y.approach\,y\\
\cline{1-2}\cline{5-6}
\multicolumn{2}{c}{NP^{\uparrow}} & & & \multicolumn{2}{c}{VP}\\
\multicolumn{2}{c}{:\lambda p.p\,(a\,unicorn)} & & & \multicolumn{2}{c}{:\lambda y.approach\,y}\\
\cline{4-6}
 & & & \multicolumn{3}{c}{VP_{to} : \lambda y.approach\,y}\\
\cline{3-5}
 & & \multicolumn{3}{c}{S\backslash NP_{3s} : \lambda y.pres\,(seem\,(approach\,y))} & \\
\cline{1-5}
\multicolumn{5}{c}{S : pres\,(seem\,(approach\,(a\,unicorn)))} &
\end{array}$$

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 $e \to t$ 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.

Chapter 3

Case

I presented an analysis of case-marking in which the case morphology itelf directly constructs the larger syntactic context in which it appears.
—*Constructive Case* Rachel Nordlinger 1998:131

As in any lexicalized theory, lexical categories are derived by morpho-lexical processes, either by explicit morphological inflection or by lexical rules, which we assume to be distinct from syntactic derivation.

3.1 Tense Morphology and Agreement

For example, the English tensed verb *sees* is derived morph-lexically as follows:[1]

(1) $sees = \lambda x \lambda y.pres\,(see\,x\,y)$

In English, this logical form is assembled by present tense morphology, thus:

(2)
$$\dfrac{\dfrac{\text{see+}}{\begin{array}{c}VP/NP\\ :\lambda x\lambda y.see\,x\,y\end{array}} \quad \dfrac{\text{+s}}{\begin{array}{c}(S\backslash NP_{3s})/\mathrm{X}\backslash\!\!\backslash VP/\mathrm{X}\\ :\lambda p\lambda x\lambda y.pres\,(p\,x\,y)\end{array}}}{\begin{array}{c}(S\backslash NP_{3s})/NP\\ :\lambda x\lambda y.pres\,(see\,x\,y)\end{array}}{<}LEX$$

X is a variable over argument categories like *NP*. The double slash $\backslash\!\!\backslash$ and the combination marked <*LEX* indicate that the tense morpheme can only apply in the morphology, as a suffix during the off-line process of learning and updating the lexicon, rather than during on-line syntactic derivation.[2]

1. This logical form is itself a schema covering a number of distinct alternatives, including a habitual stative and an achievement eventive, in a finer-grain semantics that is outlined in appendix D, distinguishing the event time *e* of *seeing* from the reference time *r* referred to and the time of utterance *u*, all of which are coincide in the case of the achievement reading but can on occasion be different (Reichenbach, 1947)
2. In the case of irregular (and frequent) forms like "saw", the tensed verb is learned directly, although "#seed" may surface during acquisition or in adults under stress.

The same tense morpheme does the right thing for the raising verb "seems"

(3)
$$\dfrac{\dfrac{\text{seem+}}{VP/VP_{to} : \lambda p \lambda y.seem\,(p\,y)} \quad \dfrac{\text{+s}}{(S\backslash NP_{3s})/\mathrm{X}\backslash\!\!\backslash VP/\mathrm{X} : \lambda p \lambda x \lambda y.pres\,(p\,x\,y)}}{(S\backslash NP_{3s})/VP_{to} : \lambda x \lambda y.pres\,(seem\,(x\,y))}{<}LEX$$

For the ternary verb "persuade" we need a second category for the tense morpheme:[3]

(4)
$$\dfrac{\dfrac{\text{persuade+}}{(VP/VP_{to})/NP : \lambda x \lambda p \lambda y.persuade\,(p\,x)\,x\,y} \quad \dfrac{\text{+ed}}{((S\backslash NP_{3s})/\mathrm{W})/\mathrm{X}\backslash\!\!\backslash(VP/\mathrm{W})/\mathrm{X} : \lambda p \lambda x \lambda w \lambda y.pres\,(p\,x\,w\,y)}}{((S\backslash NP_{3s})/VP_{to})/NP : \lambda x \lambda p \lambda y.pres\,(persuade\,(p\,x)\,x\,y))}{<}LEX$$

Welsh is similar, apart from the VSOX category of the main verb. (Note that we are forced under CCG assumptions to assign the infinitival stem the same logical form as the English equivalent, in which the infinitival subject is the last argument.) For example, we have the following morphological derivation for the Welsh present-tensed transitive, differing only syntactically and in binding order from English (2):[4]

(5)
$$\dfrac{\dfrac{\text{gwel+}}{VP/NP : \lambda x \lambda y.see\,x\,y} \quad \dfrac{\text{+odd}}{(S/\mathrm{X})/NP_{3s}\backslash\!\!\backslash VP/\mathrm{X} : \lambda p \lambda y \lambda x.past\,(p\,x\,y)}}{S/NP/NP_{3s} : \lambda y \lambda x.past\,(see\,x\,y)}{<}LEX$$

It is the tense morpheme that imposes VSO order on the tensed verb. The morpholexicon can therefore be seen as doing the work of the operation of "Head movement", which Roberts (2005:9) invokes to derive Welsh VSO order from underlying SVO order, except that the CCG infinitival does not specify any linearization on the infinitival subject, which is never realized in phonology. We return to this question in the discussion at the end of the chapter,

SOV Japanese is like English, modulo verb-finality:

3. The two types for the tense morpheme could be schematized as one.
4. We shall see later that the agreement system in Welsh differs from English in a way that will not be obvious from the discussion to this point.

(6)
$$\dfrac{\dfrac{\text{mi+}}{VP\backslash NP_{acc} : \lambda x\lambda y.see\,x\,y} \quad \dfrac{\text{+ru}}{(S\backslash NP_{nom}\backslash \mathrm{X}\backslash\!\backslash VP\backslash \mathrm{X} : \lambda p\lambda x\lambda y.pres\,(p\,x\,y)}}{(S\backslash NP_{3nom}\backslash NP_{acc} : \lambda x\lambda y.pres\,(see\,x\,y)}{<}LEX$$

Predicate-argument structures like $past\,(see\,x\,y))$ in the above examples can be seen as corresponding in some general sense to the cartographic structures of functional projections proposed by Cinque and Rizzi (2008), among others. In English, the further elaboration of cartographic structure is accomplished by unmarked lexical rules, auxiliary verbs, and adjuncts of various kinds, but in more highly inflected languages like Latin, morphology does the work. For example,*videram*, "had seen" acquires the following category via Latin morphology:[5]

(7) videram := $S\backslash NP_{nom,1sg}\backslash NP_{acc} : \lambda x\lambda y.past(perfect(see'\,x\,y))$

3.2 Case and Morpho-Lexical Type-raising

Case, on Latin nominative NPs like the proper name "Balbus" means that the NP in question must combine with a verb as its subject, as in:

(8) a. Balbus ambulat
B.NOM.3SM walks.3s
"Balbus walks"

b. *Balbus amo
B.NOM.3SM love.1S
"I love Balbus"

As in the case of English, we must assume that Latin subjects and objects are functors raised over verbs as arguments. In Latin, case and hence type-raising is specified by nominal morphology.

For example, the Latin nominative *Balbus* is morphologically marked by the suffix *-us* as a masculine singular nominative as follows, making the noun a function over the predicate:

5. This is a case where a fuller semantics would distinguish the Reichenbachian utterence time *u* from a past reference time *r* and an event time *e* of a Vendlerian achievement of seeing preceding *r*.

(9)

$$\dfrac{\dfrac{\overset{\text{Balb}}{\text{B.}}}{N_{3sm} : balbus} \quad \dfrac{\overset{\text{+us}}{\text{.NOM.3sm}}}{(S/(S\backslash NP_{nom,3sm}))\backslash\!\backslash N_{3sm} : \lambda n \lambda p.p\,n}}{S/(S\backslash NP_{nom,3sm}) : \lambda p.p\,balbus}{<}_{LEX}$$

We then have the following derivation *Balbus ambulat* ("Balbus walks"):

(10)

$$\dfrac{\dfrac{\overset{\text{Balbus}}{\text{B.NOM}}}{S/(S\backslash NP_{nom,3sm}) : \lambda p.p\,balbus} \quad \dfrac{\overset{\text{ambulat}}{\text{walks}}}{S\backslash NP_{nom,3s} : \lambda y.pres\,(walk\,y)}}{S : pres\,(walk\,balbus)}{>}$$

The same nominative category, together with the accusative "murum" ("a/the/wall") yields the transitive derivation in figure 3.1.

In English, determiners like "the", "a(n)", and "every" do much the same work of type-raising nominal stems as does Latin case-morphology, as in (11). However, such type-raising is generally ambiguous as to the category raised over. Like all ambiguity, its resolution is a matter for the processor, rather than the grammar. In English, it is tolerable because it can be resolved by underspecifying the type as $NP^{\uparrow}$ until the point in the derivation where it can be resolved by the comparitively rigid word order word-order of English. $NP^{\uparrow}$ then schematizes over the following order-preserving directional schemata:

(11) a. Forward Morpholexical Case:
$\mathrm{T}/(\mathrm{T}\backslash\mathrm{A}) : \lambda \mathrm{p}.\mathbf{f}\,\mathrm{p}$ $\quad >_{LEX}\mathbf{T}$
b. Backward Morpholexical Case:
$\mathrm{T}\backslash(\mathrm{T}/\mathrm{A}) : \lambda \mathrm{p}.\mathbf{f}\,\mathrm{p}$ $\quad <_{LEX}\mathbf{T}$

where $T\backslash A$ and T/A are existing lexical syntactic types (such as verb-types) over arguments of type A into results of type T and logical form a function $\mathbf{f}$ of p, where $\mathbf{f}$ is the same for every T, and typically originates in an $A^{\uparrow}$-determiner, as in (11).

In Welsh, non-subject complements of finite verbs differ from subjects in appearing (where applicable) in the "soft-mutated" form, in which the initial consonent is changed from citation form, as in the following minimal pair of derivations:

a. Balb+ +us mur+ +um aedificat

B. .NOM.3SM *wall* .ACC *build*.PRES.3S

N $(S/(S\backslash NP_{nom,3s}))\backslash\backslash N$ N $((S\backslash NP_{nom,agr})/((S\backslash NP_{nom,agr})\backslash NP_{acc}))\backslash\backslash N$ $(S_{pres}\backslash NP_{nom,3s})\backslash NP_{acc}$

$S/(S\backslash NP_{nom,3sm})$ $<LEX$ $(S\backslash NP_{nom,agr})/((S\backslash NP_{nom,agr})\backslash NP_{acc})$ $<LEX$

$S\backslash NP_{nom,3s}$ $<$

S $>$

"Balbus is building a wall."

Figure 3.1:

(12)

$$
\begin{array}{lll}
\text{Gwelodd} & \text{y ddynes} & \textbf{gath} \\
\hline
{}^{\ulcorner}(S/NP)/NP^{\urcorner -M} & NP\uparrow & {}^{\ulcorner SM}NP\uparrow^{\urcorner} \\
: \lambda y \lambda x.saw'xy & : the'woman' & : a'\,cat' \\
\hline
\multicolumn{2}{c}{S/NP \quad >} & \\
\multicolumn{2}{c}{: \lambda x.saw'x(the'woman')} & \\
\hline
\multicolumn{3}{c}{S \quad >} \\
\end{array}
\quad : saw'(a'\,cat')(the'woman')
$$

("The woman saw a cat")

(13)

$$
\begin{array}{lll}
\text{Gwelodd} & \textbf{cath} & \text{y ddynes} \\
\hline
{}^{\ulcorner}(S/NP)/NP^{\urcorner -M} & {}^{\ulcorner -M}NP\uparrow^{\urcorner} & NP\uparrow \\
: \lambda y \lambda x.saw'xy & : a'\,cat' & : the'woman' \\
\hline
\multicolumn{2}{c}{S/NP \quad >} & \\
\multicolumn{2}{c}{: \lambda x.saw'x\,(a'\,cat')} & \\
\hline
\multicolumn{3}{c}{S \quad >} \\
\multicolumn{3}{c}{: saw'(the'woman')(a'\,cat')} \\
\end{array}
$$

("A cat saw the woman")

This observation represents something of a reversion to an account like that of Zwicky (1984), revived in the P&P framework by Roberts (2005), as well as traditional grammars of Welsh, who identify soft mutation with non-nominative case-marking, a claim that seemed at the time to be contradicted by the fact that the objects of non-finite verbs are non-mutated in Welsh. However, the present association of abstract case with multiple raised categories in the sense of chapter 3.2, according to which morphological case-markers can be associated with multiple second-order raised categories in arbitrary or "quirky" ways that are independent of their structural role (as with Icelandic subjects) or their thematic role as Agent or Patient (as in Ergative languages). In particular, nothing prevents unmutated NPs like "cath" from carrying the category $VP\backslash(VP/NP)$ of the object of an infinitival verb, as in "yn gweld cath" ("seeing the cat") in (14a), or of the object of the locative/existential construction in (14b), as well as the category $(S/NP)\backslash((S/NP)/NP_{agr})$ of the subject of a finite verb. In fact, the unmutated Welsh NP carries the categories of the first argument of the verb, like the absolutive of an ergative language.[6]

6. Of course we do not imply that Case-marking is the *only* determinant of soft mutation. There are many other instances, such as marking feminine agreement.

(14) a. Roedd y ddynes yn gweld cath/*gath "the woman is see-
be.IMPF.3S the woman PROG see.INF cat
ing a cat."

b. Mae yn yr ardd gath/*cath "There is a cat in the garden"
be.PRES.3S in the garden cat

In short, according to the present theory, and as argued by Legate (2008), the logical form of a cased or type-raised category says nothing about the semantic or "thematic (θ)" role it will play in the sentence, say as agent or patient. That is the responsibility of the predicate that subcategorizes for that argument, such as the verb, and provides a basis for the accusative/ergative contrast, and for the possibility of "split" ergativity and Icelandic quirky case noted in the preceding chapter.

However, as we saw in the case of the generalized quantifier determiner "every", a cased/raised category may in its own right contribute other kinds of θ-role-independent syntactic and semantic information.

3.3 Modal Case

Apart from coordination, considered in the next chapter, perhaps the most compelling evidence for interpreting case formally as type-raising arises from what is sometimes called *differential subject marking* (Butt, 2009), where case not only saturates an argument and the corresponding thematic role, but also contributes to properties of the predication itself, such as modality or aspect of the verb. Butt gives the following Minimal pair from Urdu as an example:

(15) a. nadya=ne zu ja-na hɛ
Nadya.F.S=ERG ZOO.M.S.OBL go-INF.M.S be.PRES.3.S
'Nadya wants to go to the zoo'

b. nadya=ko zu ja-na hɛ
Nadya.F.S=DAT ZOO.M.S.OBL go-INF.M.S be.PRES.3.S
'Nadya has to go to the zoo'

Rather than mutliplying lexical entries for the copula and many other verbs, it is consistent with the linguistic idea of differential subject alternation to define the verb as underspecified for case, with the cased ergative and dative categories as actively contributing to the semantics of the predicate itself via the type-raising morphology, as in the following:

(16)
$$\begin{array}{cc}
\text{nadya} & =\text{ne} \\
\text{N.3}\textsc{fs} & .\textsc{erg} \\
\hline
N_{3sf} : nadya & (S/(S\backslash NP_{erg.3s}))\backslash\!\backslash N_{3s} : \lambda n \lambda p.deliberately\,(p\,n) \\
\hline
\multicolumn{2}{c}{S/(S\backslash NP_{erg.3s}) : \lambda p.deliberately\,(p\,nadya))} \quad {<}LEX
\end{array}$$

(17)
$$\begin{array}{cc}
\text{nadya} & =\text{ko} \\
\text{N.3}\textsc{fs} & .\textsc{dat} \\
\hline
N_{sf} : nadya & (S/(S\backslash NP_{dat.3s}))\backslash\!\backslash N_{3s} : \lambda n \lambda p.necessarily\,(p\,n) \\
\hline
\multicolumn{2}{c}{S/(S\backslash NP_{dat,3s}) : \lambda p.necessarily(p\,nadya)} \quad {<}LEX
\end{array}$$

We therefore have minimal pairs of derivations like the following (Butt, 2009:22):

(18) a.
$$\begin{array}{cc}
\text{yassin} & k^{\text{h}}\tilde{a}s-a \\
\text{N.3}\textsc{ms}.(\textsc{abs}) & \text{coughed}-\textsc{perf}.3\textsc{ms} \\
\hline
S/(S\backslash NP_{abs.3s}) : \lambda p.p\,yassin & S\backslash NP_{3ms} \\
\hline
\multicolumn{2}{c}{S : coughed\,yassin} \quad {>}
\end{array}$$

b.
$$\begin{array}{cc}
\text{yassin}=\text{ne} & k^{\text{h}}\tilde{a}s-a \\
\text{N.}\textsc{erg}.3\textsc{ms} & \text{coughed-}\textsc{perf}.3\textsc{s} \\
\hline
S/(S\backslash NP_{erg.3ms}) : \lambda p.deliberately(p\,yassin) & S\backslash NP_{3s} \\
\hline
\multicolumn{2}{c}{S : deliberately(coughed\,yassin)} \quad {>}
\end{array}$$

The first sentence (18a), with unmarked case involves bare unergative type-raising, However, the ergative case in the second has an effect like a raising verb of adding a modal predication.

The copula and main verb in (15) and (18) are semantically compatible with either ergative, dative, or (unmarked) absolutive subjects. Other verbs are more restricted, as in the following pair (Butt, 2009:24c-d):

(19) a. nadya=ne kohani yad k-i
Nadya.3FS.ERG story.3FS.ABS memory do.PERF.ERG.3FS
'Nadya called to mind the story'

b. nadya=ko kohani yad a-yi
Nadya.3FS.DAT story.3FS.ABS memory come.PERF.DAT.3FS
'Nadya remembered the story'

We will probably want to make such verbs subcategorize for NP_{erg} and NP_{dat} subjects respectively. This move leaves open the option to treat the idiosyncratic "quirky" subject case requirements of verbs like *la-yi*, ("brought"), which take the absolutive rather than the ergative standardly found with agentive transitive verbs (Butt, 2009:(25))

3.4 Case and Case Agreement

It seems in the nature of case that it should correspond to a single argument of the governing verb. Where multiple arguments of the same verb appear to carry the same case-marker, we will assume that they are related by agreement rather than type-raising, and that one of them is a modifier of the other, denoting either apposition or an associative relation such as inalienable possession. For example, in Korean, same-case NPs denote a part-whole relation, such as inalienable possession of the second by the first (Maling and Kim, 1992):

(20) Chelsoo-ka Suni-lul meli-lul piskiko-iss-ta
Chelsoo.NOM Suni.ACC hair.ACC combing.BE.IND
'Chelsoo is combing Suni's hair'

In keeping with our restriction of type-raising to the morph-lexicon, rather than establishing such relations at the level of noun-modification, prior to the application of case-raising, we assume that one of the same-case NPs is a cased NP modifier $NP^{\uparrow}\backslash NP$, to which a case-raised NP can apply as it can to any function over NP, to yield a standard case-raised NP, as in the following analysis:

(21)
$$
\begin{array}{cccc}
\mathrm{Chelsoo-ka} & \mathrm{Suni-lul} & \mathrm{meli-lul} & \mathrm{piskiko-iss-ta} \\
\mathrm{Chelsoo.NOM} & \mathrm{Suni.ACC} & \mathrm{hair.ACC} & \mathrm{combing.BE.IND} \\
\hline
NP^{\uparrow}_{nom} & NP^{\uparrow}_{acc} & NP^{\uparrow}_{acc}\backslash NP_{acc} & (S\backslash NP_{nom})\backslash NP_{acc} \\
\lambda p.p\,chelsoo & \lambda p.p\,suni & \lambda y\lambda p.p\,sk_{\lambda x.hair\,x\wedge partof\,y\,x} & \lambda x\lambda y.comb\,x\,y \\
\cline{2-3}
 & \multicolumn{2}{c}{NP^{\uparrow}_{acc} : \lambda p.p\,sk_{\lambda x.hair\,x\wedge partof\,suni\,x}} & \\
\cline{2-4}
 & \multicolumn{3}{c}{S\backslash NP_{nom} : \lambda y.comb\,sk_{\lambda x.hair\,x\wedge partof\,suni\,x}\,y} \\
\cline{1-3}
\multicolumn{3}{c}{S : comb\,sk_{\lambda x.hair\,x\wedge partof\,suni\,x}\,chelsoo} &
\end{array}
$$

It is predicted that this construction can iterate over embedded multiple part whole relations, as in examples discussed by by Maling and Kim, 1992: 61 ,n4, (i):[7]

7. The example involves a Korean "quirky" dative subject, which the analysis follows Maling and Kim in assuming (nonessentially) to be lexicalized via a verb category $(S\backslash NP_{dat})\backslash NP_{nom}$, rather than derived via the composition of the nominative.

(22)

Hangkwuk – ey	Sewul – ey	namtaymun.sicang – ey	pul – i	na – (a)ss – ta
Korea.DAT	Seoul.DAT	market.DAT	fire.NOM	break.out.PST.IND
$NP^{\uparrow}_{dat}$	$NP^{\uparrow}_{dat}\backslash NP_{dat}$	$NP^{\uparrow}_{dat}\backslash NP_{dat}$	$NP^{\uparrow}_{nom}$	$(S\backslash NP_{dat})\backslash NP_{nom}$
$\lambda p.p\,korea$	$\lambda y\lambda p.p\,sk_{\lambda x.seoul\,x \wedge partof\,y\,x}$	$\lambda y\lambda p.p\,sk_{\lambda x.market\,x \wedge partof\,y\,x}$	$\lambda p.p\,fire$	$\lambda x\lambda y.breakout\,x\,y$

$$NP^{\uparrow}_{dat} : \lambda p.p\,sk_{\lambda x.seoul\,x \wedge partof\,korea\,x} \quad (>)$$

$$S\backslash NP_{dat} : \lambda x.breakout\,x\,fire \quad (>)$$

$$NP^{\uparrow}_{dat} : \lambda p.p\,sk_{\lambda x.market\,x \wedge partof\,sk_{\lambda x.seoul\,x \wedge partof\,korea\,x}\,x} \quad (>)$$

$$S : breakout\,sk_{market\,x \wedge partof\,sk_{\lambda x.seoul\,x \wedge partof\,korea\,x}\,x}\,fire \quad (>)$$

3.5 Stacked Case

The phenomenon of "stacked" multiple case agreement found in some Australian languages, where multiple case markers apply to the same stem, with only the innermost case marker corresponding to its actual semantic role, while others relate to the markers of other arguments, is a rather diffeerent case that we only touch on briefly here. In general it appears to involve anaphoric relations of the kind assumed below to apply to extraposed adjuncts and relative modifiers in English. For example, Simpson (1991) cites the following minimal pair for Warlpiri:

(23) a.

Karnta-ngyu	kar-la	kurdu-ku	miyi	yi-nyi	parraja-rla-ku
woman.ERG	PRES.3DAT	baby.DAT	food.(ABS)	give.NPST	coolamon.LOC.DAT

"The woman is giving food to the baby (who is) in the coolamon."

b.

Karnta-ngyu	kar-la	kurdu-ku	miyi	yi-nyi	parraja-rla-ku
woman.ERG	PRES.3DAT	baby.DAT	food.(ABS)	give.NPST	coolamon.LOC(.ABS)

"The woman is giving the baby food (which is) in the coolamon

Butt and Nordlinger (1997, 2000), citing Dench, 1987/1995:60, discuss the following example from Martuthunira: a language of the Pilbara region of Western Australia:[8]

(24)

Ngayu	nhawu-lha	ngurnu	tharnta-a	mirtily-marta-a	thara-ngka-marta-a
I	saw.PAST	that.ACC	euro.ACC	joey.PROP.ACC	pouch.LOC.PROP.ACC

"I saw that euro with its joey in its pouch"

The three nouns carry progressively more and more stacked morphological markers, each appearing to agree with its predecessor, so that the last has three suffixes. Nordlinger notes that "the locative suffix relates 'pouch' with

8. A euro is a kind of kangaroo, and its joey is its offspring.

'joey'; the proprietive suffix relates the embedded phrase 'joey in pouch' to the head nominal 'euro' (note that it is marked on both elements of the embedded phrase) and the accusative case suffix appears on all elements of the higher noun phrase to indicate that [they function] as the object of the clause."

We will concentrate on the noun-group meaning "euro with its joey in (its) pouch,". In English, which lacks almost all case agreement, the same meaning would be derived as follows, in which *in its pouch* modifies *joey* and *with its joey in its pouch* modifyies *euro*, and where the relations of the joey, the pouch, and the euro are established paratactically, via anaphora, rather than grammatically:

(25)

that	euro	with	its	joey	in	its
$NP^{\uparrow}/N$ $: \lambda n \lambda p.p\,sk_n$	N $: euro$	$(N\backslash N)/NP$ $: \lambda y \lambda n \lambda x.n\,x \wedge with\,y\,x$	NP/N $: \lambda n.sk_{\lambda x.n\,x \wedge own\,x\,pro_{it}}$	N $: joey$	$(N\backslash N)/NP$ $: \lambda y \lambda n \lambda x.n\,x \wedge in\,y\,x$	NP/N $: \lambda n.sk_{\lambda x.n\,x \wedge own\,x\,p}$

$$NP\ : sk_{\lambda x.pouch\,x \wedge o}$$

$$N\backslash N\ : \lambda n \lambda x.n\,x \wedge in\,sk_{\lambda x.pouch\,x \wedge own\,x\,p}$$

$$N : \lambda y.joey\,y \wedge in\,sk_{\lambda x.pouch\,x \wedge own\,x\,pro_{it}}\,y$$

$$NP : sk_{\lambda y.joey\,y \wedge in\,sk_{\lambda x.pouch\,x \wedge own\,x\,pro_{it}}\,y \wedge own\,y\,pro_{it}}$$

$$\overline{N\backslash N : \lambda n \lambda z.n\,z \wedge with\,sk_{\lambda y.joey\,y \wedge in\,sk_{\lambda x.pouch\,x \wedge own\,x\,pro_{it}}\,y \wedge own\,y\,pro_{it}}\,z}^{>}$$

$$\overline{N : \lambda z.euro\,z \wedge with\,sk_{\lambda y.joey\,y \wedge in\,sk_{\lambda x.pouch\,x \wedge own\,x\,pro_{it}}\,y \wedge own\,y\,pro_{it}}\,z}^{>}$$

$$\overline{NP^{\uparrow} : \lambda z.euro\,z \wedge with\,sk_{\lambda y.joey\,y \wedge in\,sk_{\lambda x.pouch\,x \wedge own\,x\,pro_{it}}\,y \wedge own\,y\,pro_{it}}\,z}^{>}$$

We will similarly assume that the Martuthinira determiner *ngurnu* is an accusative type-raised determiner $NP^{\uparrow}_{acc}/N_{acc}$, and that the binding of pronominial arguments is non-syntactic:

(26)

ngurnu that.ACC	tharnt – a euro.ACC	mirtily – marta – a joey.PROP.ACC	thara – ngka – marta – a pouch.LOC.PROP.ACC
$NP^{\uparrow}_{acc}/N_{acc}$ $: \lambda n \lambda p.p\,sk_n$	N_{acc} $: euro$	$N_{acc}\backslash N_{acc}$ $: \lambda p \lambda x.p\,x \wedge partof\,pro_{acc}\,sk_{joey}$	$N_{acc}\backslash N_{acc}$ $: \lambda p \lambda x.p\,x \wedge in\,sk_{pouch}\,pro_{prop}$

$$\overline{N_{acc} : \lambda x.p\,x \wedge partof\,pro_{acc}\,sk_{joey}}^{<}$$

$$\overline{N : \lambda x.euro\,x \wedge partof\,pro_{acc}\,sk_{joey} \wedge in\,sk_{pouch}\,pro_{prop}}^{<}$$

$$\overline{NP^{\uparrow}_{acc} : \lambda p.p\,sk_{\lambda x.euro\,x \wedge partof\,pro_{acc}\,sk_{joey} \wedge in\,sk_{pouch}\,pro_{prop}}}^{>}$$

While the binding of the possessor of the joey to the euro could reasonably be accomplished in the semantics, it has been specified in (26) via an anaphor pro_{acc} for uniformity with the occupant of the pouch pro_{prop}, which could not. Following Jelinek (1984), it is assumed that the stacked locatives and in fact all the arguments in the Warlpiri examples (23) are similarly anaphorically-bound adjuncts to a lexically pronominally saturated finite verb-phrase.[9] NOTE ON IMPLICATIONS OF CASE-RAISING FOR ACQUISITION..

9. This is not the analysis of Nordlinger, 1998:140-145, who has stacked morphology construct a skeletal LFG f-structure for the whole NP which is then filled in by "inside-out" unification, which Dalrymple (1993) applied to anaphor binding via "functional uncertainty", the mechanism used in LFG for unbounded wh-movement.

Chapter 4

Composition

The possibility of conjunction offers one of the best criteria for the initial determination of phrase structure.
—*Syntactic Structure* Noam Chomsky, 1957:36

Conjunctions like "and" and "or" appear happy to combine any pair of categories just so long as they are of the same type, to yield a category of that same type:

(1) a. Gabbitas [[walks]$_{S\backslash NP}$ and [talks]$_{S\backslash NP}$]$_{S\backslash NP}$.
b. Gabbitas [[walks]$_{S\backslash NP}$ and [chews gum]$_{S\backslash NP}$]$_{S\backslash NP}$.
c. Gabbitas [[bought]$_{(S\backslash NP)/NP}$ and [sold]$_{(S\backslash NP/)NP}$]$_{(S\backslash NP)/NP}$ a car..
d. Gabbitas [[bought]$_{(S\backslash NP)/NP}$ and [sold my brother]$_{(S\backslash NP)/NP}$]$_{(S\backslash NP)/NP}$ various books.
e. Gabbitas [[gave]$_{((S\backslash NP)/NP)/NP}$ or [sold]$_{((S\backslash NP)/NP)/NP}$]$_{((S\backslash NP)/NP)/NP}$ my brother various books.
(etc.)

We must therefore write the category for conjunctions as follows:

(2) and := $(\mathrm{T}\backslash_{\star}\mathrm{T})/_{\star}\mathrm{T} : \lambda p \lambda q.p \sqcap q$

—where T is S or any function into S, and $\sqcap$ is the pointwise recursive extension of logical conjuntion $\wedge$ to functions of any valency into S (Partee and Rooth, 1983).

As in any theory of coordination, category (2) itself must be excluded as a value for T, to disallow the following:

(3) a. *John walks and and and talks.
b. *John walks and stalks and and talks.

The $\star$-type slashes mean that the conjunction category can *only* combine by the application rules (4), so that (1c) can be derived as follows

(4)
$$
\begin{array}{ccccc}
\text{Gabbitas} & \text{bought} & \text{and} & \text{sold} & \text{a car} \\
\hline
NP^{\uparrow} & (S\backslash NP)/NP & (\mathrm{T}\backslash_{\star}\mathrm{T})/_{\star}\mathrm{T} & (S\backslash NP)/NP & NP^{\uparrow} \\
:\lambda p.p\,gabbitas & :\lambda x\lambda y.bought\,x\,y & :\lambda p\lambda q.p\sqcap q & :\lambda x\lambda y.sold\,x\,y & :\lambda p.p\,(a\,car) \\
\cline{3-4}
 & & \multicolumn{2}{c}{((S\backslash NP)/NP)\backslash((S\backslash NP)/NP)\ {>}} & \\
 & & \multicolumn{2}{c}{\lambda q\lambda x\lambda y.sold\,x\,y\wedge q\,x\,y} & \\
\cline{2-4}
 & \multicolumn{3}{c}{(S\backslash NP)/NP\ {<}} & \\
 & \multicolumn{3}{c}{:\lambda x\lambda y.sold\,x\,y\wedge bought\,x\,y} & \\
\cline{2-5}
 & \multicolumn{4}{c}{S\backslash NP:\lambda y.sold\,(a\,car)\,y\wedge bought\,(a\,car)\,y\ {<}} \\
\hline
\multicolumn{5}{c}{S:sold\,(a\,car)\,gabbitas\wedge bought\,(a\,car)\,gabbitas\ {>}}
\end{array}
$$

However, conjunction can also apply to things that are neither words nor traditional constituents, like "might sell" in the following example:

(5) a. I bought and might sell a car.
 b. I gave or will send those boys these books.

The reasoning behind the conjunction category (2) forces us to believe that "might sell" must have the same syntactic type as "bought", namely that of a transitive verb, $(S\backslash NP)/NP$. However, we cannot combine "might", $(S\backslash NP)/VP$, and "sell", VP/NP using either of the application rules (4). We need another rule from the family of rules of function composition:

4.1 Combining Categories II: Composition

The following rules of functional composition will be needed:

(6) MERGE IIA: THE COMPOSITION RULES
 a. Forward Composition:
 $$X/_{\diamond}Y:f \quad Y/Z:g \quad \Rightarrow \quad X/Z:\lambda z.f\,(g\,z) \qquad (>\mathbf{B}\ )$$
 b. Backward Composition:
 $$Y\backslash Z:g \quad X\backslash_{\diamond}Y:f \quad \Rightarrow \quad X\backslash Z:\lambda z.f\,(g\,z) \qquad (<\mathbf{B}\ )$$
 c. Forward Crossing Composition:
 $$X/_{\times}Y:f \quad Y\backslash Z:g \quad \Rightarrow \quad X\backslash Z:\lambda z.f\,(g\,z) \qquad (>\mathbf{B}_{\times})$$
 d. Backward Crossing Composition:
 $$Y/Z:g \quad X\backslash_{\times}Y:f \quad \Rightarrow \quad X/Z:\lambda z.f\,(g\,z) \qquad (<\mathbf{B}_{\times})$$

These rules conform to the Combinatory Projection Principle (5) in applying to strictly adjacent categories, consistent with the directionality of the governing category $X|Y$ and projecting the type and directionality of the argument(s) $|Z$

onto the result.

The $\diamond$- and $\times$–type slashes on the governing functor X/Y or $X\backslash Y$ mean that only categories whose slash is compatible with that type can combine by them. In particular, the conjunction category (2) cannot do so. The unrestricted slashes on the secondary functor Y/Z or $Y\backslash Z$ mean that any slash-type is compatible. (However, the CPP says that the type is inherited by the result X/Z or $X\backslash Z$.)

The unrestricted $/$ and $\backslash$ slashes on the lexical categories of the modals and infinitivals mean that they can act as the governing category in rules of composition, allowing the following derivation for (5):[1]

(7)

$$
\begin{array}{cccccc}
\text{I} & \text{bought} & \text{and} & \text{might} & \text{sell} & \text{a car} \\
\hline
NP^{\uparrow} & (S\backslash NP)/NP & (\mathrm{T}\backslash_{\star}\mathrm{T})/_{\star}\mathrm{T} & (S\backslash NP)/VP & VP/NP & NP^{\uparrow} \\
:\lambda p.p\,me & :\lambda x\lambda y.bought\,x\,y & :\lambda p\lambda q.p\sqcap q & :\lambda p\lambda y.possible\,(p\,y) & :\lambda x\lambda y.sell\,x\,y & :\lambda p.p\,(a\,car) \\
\hline
 & & & \multicolumn{2}{c}{(S\backslash NP)/NP \quad >\mathbf{B}} & \\
 & & & \multicolumn{2}{c}{:\lambda x\lambda y.possible\,(sell\,x\,y)} & \\
\hline
 & \multicolumn{4}{c}{(S\backslash NP)/NP \quad <>} & \\
 & \multicolumn{4}{c}{:\lambda x\lambda y.possible\,(sell\,x\,y)\wedge bought\,x\,y} & \\
\hline
 & \multicolumn{5}{c}{S\backslash NP:\lambda y.possible\,(sell\,(a\,car)\,y)\wedge bought\,(a\,car)\,y \quad <} \\
\hline
\multicolumn{5}{c}{S:possible\,(sell\,(a\,car)\,me)\wedge bought\,(a\,car)\,me \quad >} & \\
\end{array}
$$

This amounts to saying that, grace of the forward composition rule (6a), “might sell” is a *constituent*, on an even footing with the transitive verb, even if traditional notions of constituency don’t recognize it as such.

To allow (5) under a similar argument, we also need the following similarly CPP-compliant “level 2” generalization of the composition rules to allow composition into bivalent dependent functors $(Y|Z)|W$:

1. There is quite a lot going on in the logical forms, the details of which we will pass over for now. From now on we also abbreviate the successive forward and backward applications of the conjunction category as a single combinationindexed $<>$.

(8) MERGE IIB: THE LEVEL 2 COMPOSITION RULES

a. Forward Level 2 Composition:

$$X/_{\diamond}Y : f \quad (Y/Z)|W : g \quad \Rightarrow \quad (X/Z)|W : \lambda w \lambda z.f\,(g\,w\,z) \qquad (>\mathbf{B}^2\ )$$

b. Backward Level 2 Composition:

$$(Y\backslash Z)|W : g \quad X\backslash_{\diamond}Y : f \quad \Rightarrow \quad (X\backslash Z)|W : \lambda w \lambda z.f\,(g\,w\,z) \qquad (<\mathbf{B}^2\ )$$

c. Forward Level 2 Crossing Composition:

$$X/_{\times}Y : f \quad (Y\backslash Z)|W : g \quad \Rightarrow \quad (X\backslash Z)|W : \lambda w \lambda z.f\,(g\,w\,z) \qquad (>\mathbf{B}^2_{\times})$$

d. Backward Level 2 Crossing Composition:

$$(Y/Z)|W : g \quad X\backslash_{\times}Y : f \quad \Rightarrow \quad (X/Z)|W : \lambda w \lambda z.f\,(g\,w\,z) \qquad (<\mathbf{B}^2_{\times})$$

The first of these rules, $>\mathbf{B}^2$, allows the following derivation for (**??**b):

(9)

$$\begin{array}{lllllll}
\text{I} & \text{gave} & \text{or} & \text{will} & \text{send} & \textit{these boys} & \text{those books} \\
NP^{\uparrow} & ((S\backslash NP)/NP)/NP & (\mathrm{T}\backslash_{\star}\mathrm{T})/_{\star}\mathrm{T} & (S\backslash NP)/VP & (VP/NP)/NP & NP^{\uparrow} & NP^{\uparrow} \\
: \lambda p.p\,me & : \lambda x \lambda w \lambda y.past\,(give\,w\,x\,y) & : \lambda p \lambda q.p \sqcup q & : \lambda p \lambda y.predicted\,(p\,y) & : \lambda x \lambda w \lambda y.send\,w\,x\,y & : \lambda p.p\,(these\,boys) & : \lambda p.p\,(those\,books)
\end{array}$$

$$\dfrac{will \quad send}{((S\backslash NP)/NP)/NP : \lambda x \lambda w \lambda y.predicted\,(send\,w\,x\,y)}\ {>}\mathbf{B}^2$$

$$\dfrac{gave \quad or \quad will\ send}{((S\backslash NP)/NP)/NP : \lambda x \lambda w \lambda y.predicted\,(send\,w\,x\,y) \vee past(give\,w\,x\,y)}\ {<>}$$

$$\dfrac{}{(S\backslash NP)/NP : \lambda w \lambda y.predicted\,(send\,w\,(those\,boys)\,y) \vee past(give\,w\,(these\ boys)\,y)}\ {<}$$

$$\dfrac{}{S\backslash NP : \lambda y.predicted\,(send\,(thosebooks)\,(these\,boys)\,y) \vee past(give\,(thosebooks)\,(these\ boys)\,y)}\ {<}$$

$$\dfrac{}{S : predicted\,(send\,(thosebooks)\,(these\,boys)\,me) \vee past(give\,(thosebooks)\,(these\ boys)\,me)}\ {>}$$

It will become important later to note that repeated application of the above rules can derive categories of unboundedly high valency, because X in the governing functor can itself be a functor, as it is unboundedly Since the inclusion of second-order composition $\mathbf{B}^2$ allows derived syntactic types of unboundedly high valency, to avoid increasing expressive power, the variable T in the above category must be restricted to syntactic categories of bounded valency.

Modals like “might” are a variety of adjunct to the VP. It is perhaps not surprising that standard adjuncts like “tomorrow” can also compose with transitive verbs (via the backward crossing composition rule (6d)) in the “Heavy NP Shift” construction (10), and that the result of the composition, having the same type $(S\backslash NP)/NP$ as the transitive verb, can coordinate with it, as in (11).

(10)

$$
\begin{array}{ccccc}
\text{I} & \text{might} & \text{sell} & \text{tomorrow} & \text{a very heavy book} \\
\hline
NP^{\uparrow} & (S\backslash NP)/VP & VP/NP & VP\backslash VP & NP^{\uparrow} \\
:\lambda p.p\,me & :\lambda p\lambda y.possible\,(p\,y) & \lambda x\lambda y.sell\,x\,y & :\lambda p\lambda y.tomorrow\,p\,y & :\lambda p.p\,(a\,book)
\end{array}
$$

$$
\begin{array}{l}
\text{sell tomorrow} \;{<}\mathbf{B}_{\times}:\; VP/NP : \lambda x\lambda y.tomorrow\,(sell\,x)\,y \\
\text{might sell tomorrow} \;{>}\mathbf{B}:\; (S\backslash NP)/NP : \lambda x\lambda y.possible\,(tomorrow\,(sell\,x)\,y) \\
\text{might sell tomorrow a very heavy book} \;{<}:\; S\backslash NP : \lambda y.possible\,(tomorrow\,(sell\,(a\,book))\,y) \\
\text{I might sell tomorrow a very heavy book} \;{>}:\; S : possible\,(tomorrow\,(sell\,(a\,book))\,me)
\end{array}
$$

This rearrangement is a case of "scrambling" canonical order, which in English is restricted to adjuncts The next chapter will show that other languages allow much more general scrambling.

Again, "sell tomorrow" has the status of a constituent in the derivation above, so we are not surprised to find that it can undergo coordination

(11)

$$
\begin{array}{ccccccc}
\text{I} & \text{bought} & \text{and} & \text{might} & \text{sell} & \text{tomorrow} & \text{a very fast car} \\
\hline
NP^{\uparrow} & (S\backslash NP)/NP & (\mathrm{T}\backslash_{\star}\mathrm{T})/_{\star}\mathrm{T} & (S\backslash NP)/VP & VP/NP & VP\backslash VP & NP^{\uparrow} \\
:\lambda p.p\,me & :\lambda x\lambda y.bought\,x\,y & :\lambda p\lambda q.p\sqcap q & :\lambda p\lambda y.possible\,(p\,y) & :\lambda x\lambda y.sold\,x\,y & :\lambda p\lambda y.tomorrow\,p\,y & :\lambda p.p\,(a\,car)
\end{array}
$$

$$
\begin{array}{l}
\text{sell tomorrow} \;{<}\mathbf{B}_{\times}:\; VP/NP : \lambda x\lambda y.tomorrow\,(sell\,x)\,y \\
\text{might sell tomorrow} \;{>}\mathbf{B}:\; (S\backslash NP)/NP : \lambda x\lambda y.possible\,(tomorrow\,(sell\,x)\,y) \\
\text{bought and might sell tomorrow} \;{<}{>}:\; (S\backslash NP)/NP : \lambda x\lambda y.possible(tomorrow\,(sell\,x)\,y) \wedge bought\,x\,y \\
\text{bought and might sell tomorrow a very fast car} \;{<}:\; S\backslash NP : \lambda y.possible(tomorrow\,(sell\,(a\,car))\,y) \wedge bought\,(a\,car)\,y \\
\text{I bought and might sell tomorrow a very fast car} \;{>}:\; S : possible(tomorrow\,(sell\,(a\,car))\,y) \wedge bought\,(a\,car)\,me
\end{array}
$$

However, in order to prevent overgeneration of examples like (14) via an analogous scrambling derivation, we must assume that the type raised NP arguments that we schematize as $NP^{\uparrow}$ must in English be of the general form $T\backslash_{\diamond\star}(T/NP)$, incompatible with crossing composition. In particular, the dative argument of ditransitives must be $VP\backslash_{\diamond\star}(VP/NP)$.[2]

(12) $*$I will

$$
\begin{array}{ccc}
\text{give} & \text{flowers} & \text{my very heavy friends.} \\
\hline
(VP/NP)/NP & VP\backslash_{\diamond\star}(VP/NP) & (VP/NP)\backslash_{\diamond\star}((VP/NP)/NP) \\
\hline
\multicolumn{3}{r}{***}
\end{array}
$$

This is a reflex of a general observation about English NP components and English fixed word-order, rather than a stipulation specific to Heavy NP shift: *all* nominal functors have to have this restriction to reflect the fact that English

2. We shall see below that cased arguments in other languages that allow "scrambled" word order, such as Japanese and German, are less restricted in this way.

nominal word order is more rigid than verbal, excluding the following:

(13) $$\begin{array}{llllll} *\text{He was} & \text{a} & \text{little} & \text{who said he worked on the railways} & \text{man} \\ \cline{2-2}\cline{3-3}\cline{4-4} & NP^{\uparrow}/N & N/_{\diamond\star}N & N\backslash_{\diamond\star}N & N \\ \cline{3-5} & & & & {***} \end{array}$$

We shall see in the next chapter that other languages such as German are less restricted in this respect.

The particle-verb constructions like *call up*, *show off* etc., that are so astonishingly abundant in English seem similarly to exclude "light" objects such as pronouns when the particle is medial:

(14) a. I called the girl.
b. I called the girl/her up.
c. I called up the girl/#her.

What about coordinate sentences like the following?

(15) a. I caught and you cooked a fish.
b. I gave and you sold old books to the library

By the logic of the argument so far, since "I caught" and "you cooked" can coordinate to yield something that combines with an object to yield a sentence, they must be syntactically typable in the same sense as "might sell" in (7) and the VP in (**??**b). In fact by this logic they must be *consituents* of type S/NP, despite the fact that there is no traditional name for that type in English. (We might be encouraged in this believe by the fact that languages like Latin that decline transitive verbs according to person and number ("amo, amas, amat," etc.) actually lexicalize such elements as "I like, you like, he/she/it likes," etc.

Similarly, "I gave" and "you sold" must be constituents of type $(S/PP)/NP$.

FROM MINIMALISM7.

Because what we were thinking of as arguments are now seen to be adjunct-like functions, they can do everything functions can do. In particular, like adjuncts, they can compose by the composition rules (6). We immediately predict several varieties of so called "non-constituent" coordination, such as (15a):[3]

3. We continue to abbreviate the forward and backward applications of the conjunction category as a single combination, indexed $<>$.

(16)

$$\begin{array}{cccccc} \text{I} & \text{caught} & \text{and} & \text{you} & \text{cooked} & \text{a fish} \\ \hline S/(S\backslash NP) & (S\backslash NP)/NP & (\mathrm{T}\backslash_{\star}\mathrm{T})/_{\star}\mathrm{T} & S/(S\backslash NP) & (S\backslash NP)/NP & S\backslash(S/NP) \\ :\lambda p.p\,me & :\lambda x\lambda y,caught\,x\,y & :\lambda p\lambda q.p\sqcap q & :\lambda p.p\,you & :\lambda x\lambda y,cooked\,x\,y & afish \end{array}$$

$$S/NP : \lambda x.caught\,x\,me \;{>}\mathbf{B} \qquad S/NP : \lambda x.cooked\,x\,you \;{>}\mathbf{B}$$

$$S/NP : \lambda x.caught\,x\,me \wedge cooked\,x\,you \;{<}{>}$$

$$S : caught\,(afish)\,me \wedge cooked\,(afish)\,you \;{<}$$

Similarly, by the level 2 rule

(17)

$$\begin{array}{ccccccc} \text{I} & \text{gave} & \text{and} & \text{you} & \text{sold} & \text{old books} & \text{to the library} \\ \hline S/(S\backslash NP) & ((S\backslash NP)/PP)/NP & (\mathrm{T}\backslash_{\star}\mathrm{T})/_{\star}\mathrm{T} & S/(S\backslash NP) & ((S\backslash NP)/PP)/NP & NP^{\uparrow} & PP^{\uparrow} \\ :\lambda p.p\,me & :\lambda w\lambda x\lambda y,gave\,w\,x\,y & :\lambda p\lambda q.p\sqcap q & :\lambda p.p\,you & :\lambda x\lambda y,cooked\,x\,y & afish & \end{array}$$

$$S/NP : \lambda x.caught\,x\,me \;{>}\mathbf{B} \qquad S/NP : \lambda x.cooked\,x\,you \;{>}\mathbf{B}$$

$$S/NP : \lambda x.caught\,x\,me \wedge cooked\,x\,you \;{<}{>}$$

$$S : caught\,(afish)\,me \wedge cooked\,(afish)\,you \;{<}$$

Remarkably, we also immediately capture the (in traditional terms non-constituent) phenomenon of argument-adjunct cluster coordination, (18) in Chapter 1, as in Figure **??** (TODO).[4]

(18)

$$\begin{array}{ccccccc} \text{I} & \text{gave} & \text{Ike} & \text{a bike} & \text{and} & \text{Adlai} & \text{a train} \\ \hline S/(S\backslash NP) & ((S\backslash NP)/NP)/NP & ((S\backslash NP)/NP)\backslash(((S\backslash NP)/NP)/NP) & (S\backslash NP)\backslash((S\backslash NP)/NP) & (\mathrm{T}\backslash_{\star}\mathrm{T})/_{\star}\mathrm{T} & ((S\backslash NP)/NP)\backslash(((S\backslash NP)/NP)/NP) & (S\backslash NP)\backslash((S\backslash N \\ :\lambda p.p\,me & :\lambda x\lambda w\lambda y.gave\,w\,x\,y & :\lambda p\lambda w\lambda y.p\,w\,ike\,y & :\lambda p\lambda y.p\,(a\,bike)\,y & :\lambda p\lambda q.p\sqcap q & :\lambda p\lambda w\lambda y.p\,w\,adlai\,y & :\lambda p\lambda y.p\,(a\,tr \end{array}$$

$$(S\backslash NP)\backslash(((S\backslash NP)/NP)/NP) : \lambda p\lambda y.p\,(a\,bike)\,ike\,y \;{<}\mathbf{B} \qquad (S\backslash NP)\backslash(((S\backslash NP)/NP)/NP) : \lambda p\lambda y.p\,(a\,train)\,adlai\,y$$

$$(S\backslash NP)\backslash(((S\backslash NP)/NP)/NP) : \lambda p\lambda y.p\,(a\,train)\,adlai\,y \wedge p\,(a\,bike)\,ike\,y$$

$$S\backslash NP : \lambda y.gave\,(a\,train)\,adlai\,y \wedge gave\,(a\,bike)\,ike\,y \;{<}$$

$$S : gave\,(a\,train)\,adlai\,me \wedge gave\,(a\,bike)\,ike\,me \;{<}$$

(19)

$$\begin{array}{ccccccc} \text{I} & \text{saw} & \text{Ike} & \text{on Monday} & \text{and} & \text{Adlai} & \text{onWednesday} \\ \hline S/(S\backslash NP) & (S\backslash NP)/NP & (S\backslash NP)\backslash((S\backslash NP)/NP) & (S\backslash NP)\backslash(S\backslash NP) & (\mathrm{T}\backslash_{\star}\mathrm{T})/_{\star}\mathrm{T} & (S\backslash NP)\backslash((S\backslash NP)/NP) & (S\backslash NP)\backslash(S\backslash NP) \\ :\lambda p.p\,me & :\lambda x\lambda y.saw\,x\,y & :\lambda p\lambda y.p\,ike\,y & :\lambda p\lambda y.(on\,monday)\,p\,y & :\lambda p\lambda q.p\sqcap q & :\lambda p\lambda y.p\,adlai\,y & :\lambda p\lambda y.(on\,wednesday)\,p\,y \end{array}$$

$$(S\backslash NP)\backslash((S\backslash NP)/NP) : \lambda p\lambda y.(on\,monday)\,p\,ike\,y \;{<}\mathbf{B} \qquad (S\backslash NP)\backslash((S\backslash NP)/NP) : \lambda p\lambda y.(onwednesday)\,p\,adlai\,y \;{<}\mathbf{B}$$

$$(S\backslash NP)\backslash((S\backslash NP)/NP) : \lambda p\lambda y.(onwednesday)\,p\,adlai\,y \wedge (on\,monday)\,p\,ike\,y \;{<}{>}$$

$$S\backslash NP : \lambda y.(onwednesday)\,saw\,adlai\,y \wedge (on\,monday)\,saw\,ike\,y \;{<}$$

$$S : (onwednesday)\,saw\,adlai\,me \wedge (on\,monday)\,saw\,ike\,y,me \;{<}$$

4. Cf. Dowty (1985/1988); Steedman (1985).

(20) I told Ike that it was raining and Adlai that it was snowing

$$\begin{array}{ccccccc}
\text{I} & \text{told} & \text{Ike} & \text{that it was raining} & \text{and} & \text{Adlai} & \text{that it was snowing} \\
\hline
S/(S\backslash NP) & ((S\backslash NP)/S')/NP & ((S\backslash NP)/S')\backslash(((S\backslash NP)/S)/NP) & (S\backslash NP)\backslash((S\backslash NP)/S') & (T\backslash_{\star}T)/_{\star}T & ((S\backslash NP)/S')\backslash(((S\backslash NP)/S')/NP) & (S\backslash NP)\backslash((S\backslash NP)/S') \\
: \lambda p.p\,me & : \lambda x\lambda w\lambda y.told\,w\,x\,y & : \lambda p\lambda w\lambda y.p\,w\,ike\,y & : \lambda p\lambda y.p\,raining\,y & : \lambda p\lambda q.p\sqcap q & : \lambda p\lambda w\lambda y.p\,w\,adlai\,y & : \lambda p\lambda y.p\,snowing\,y
\end{array}$$

$$\text{Ike} + \text{that it was raining} \;(<\mathbf{B}):\; (S\backslash NP)\backslash(((S\backslash NP)/S')/NP) : \lambda p\lambda y.p\,raining\,ike\,y$$

$$\text{Adlai} + \text{that it was snowing} \;(<\mathbf{B}):\; (S\backslash NP)\backslash(((S\backslash NP)/S')/NP) : \lambda p\lambda y.p\,snowing\,adlai\,y$$

$$(<\!\!>):\; (S\backslash NP)\backslash(((S\backslash NP)/S')/NP) : \lambda p\lambda y.p\,snowing\,adlai\,y \wedge p\,raining\,ike\,y$$

$$(<):\; S\backslash NP : \lambda y.told\,snowing\,adlai\,y \wedge told\,raining\,ike\,y$$

$$(<):\; S : told\,snowing\,adlai\,me \wedge told\,raining\,ike\,me$$

4.2 Discussion

The above discussion requires us to rethink the traditional notion of constituence. If strings like “might sell”, “I caught”, and “Adlai a train” are typable by the grammar as $(S\backslash NP)/NP$, S/NP, and $(S\backslash NP)\backslash(((S\backslash NP)/NP)/NP)$ for purposes of coordination, then they must be possible constituents of canonical sentences as well. For example, as well as the standard derivation (21) for the simple transitive “I saw Esau”, we must allow a non-standard derivation like (22):[5]

(21)

$$\begin{array}{ccc}
\text{I} & \text{saw} & \text{Esau} \\
\hline
S/(S\backslash NP) : \lambda p.p\,me & (S\backslash NP)/NP : \lambda x\lambda y,saw\,x\,y & (S\backslash NP)\backslash((S\backslash NP)/NP) : \lambda p.p\,esau \\
\hline
 & \multicolumn{2}{c}{S\backslash NP : \lambda y.saw\,esau\,y \quad <} \\
\hline
\multicolumn{3}{c}{S : saw\,esau\,me \quad >}
\end{array}$$

(22)

$$\begin{array}{ccc}
\text{I} & \text{saw} & \text{Esau} \\
\hline
S/(S\backslash NP) : \lambda p.p\,me & (S\backslash NP)/NP : \lambda x\lambda y,saw\,x\,y & S\backslash(S/NP) : \lambda p.p\,esau \\
\hline
\multicolumn{2}{c}{S/NP : \lambda x.saw\,x\,me \quad >\mathbf{B}} & \\
\hline
\multicolumn{3}{c}{S : saw\,esau\,me \quad <}
\end{array}$$

In this connection, it might be pointed out in defense of this position that the traditional tests for constiutuency, namely lexical substitutability, ability to undergo movement, ability to undergo coordination, and ability to be marked as an intonational phrase, are mutually inconsistent (Jacobson, 2006), and in the case of the last two, on the side of the present definition, rather than the traditional one.

In more positive support of our proposal, it might also be pointed out that it seems to give us an account of apparent discontinuity under coordination

5. Note that the transitive object needs two distinct syntactic types, although their semantics is the same. We retunr to this point in the next chapter.

that does not require discontinuity in rules of grammar, such as movement, deletion, multidominance, or transderivational parallel structure constraints.

Chapter 5

Word-Order

The possibility of conjunction offers one of the best criteria for the initial determination of phrase structure.
—*Syntactic Structure* Noam Chomsky, 1957:36

Languages like Japanese and German notoriously exhibit the phenomenon of "scrambling", whereby the arguments of a verbs are rather freely ordered linearly within the local domain to its left. The combination of case as morpholexical type-raising and rules of composition merger—in particular, crossed composition— predicts the existence of such languages, despite the assumption of a universal ordered lexicon.

5.1 Scrambling in Japanese

Miyagawa, 1997 :1 cites the following example:

(1) a. John-ga Mary-ni piza-o ageta.
John.NOM Mary.DAT pizza.ACC give.PAST
"John gave Mary a pizza."

b. John-ga piza-o Mary-ni ageta.
John.NOM pizza.ACC Mary.DAT give.PAST
"John gave Mary a pizza."

c. piza-o John-ga Mary-ni ageta.
pizza.ACC John.NOM Mary.DAT give.PAST
"John gave Mary a pizza."
(etc.)

Arguments, including subjects and objects, are accordingly taken to move boundedly from their θ or argument-structural position (which in these languages is consistent with the default linearization) to an adjoined position (Saito, 1992; Bošković and Takahashi, 1998). In terms of the theory proposed here, this means that arguments themselves must be second-order type-raised functors of the general form $\mathrm{T}/(\mathrm{T}\backslash NP)$ where T is a variable ranging over lexical types S, $S\backslash NP$, etc., with invariant semantics.

In the case of the basic Japanese ditransitive order (1a), we have the following purely applicative derivation with morphologically-derived cased argument categories, shown fully-specified:[1]

(2)
$$
\begin{array}{cccc}
\text{John}-\text{ga} & \text{Mary}-\text{ni} & \text{piza}-\text{o} & \text{ageta.}\\
\hline
S/(S\backslash NP_{nom}) & (S\backslash NP_{nom})/((S\backslash NP_{nom})\backslash NP_{dat}) & ((S\backslash NP_{nom})\backslash NP_{dat})/(((S\backslash NP_{nom})\backslash NP_{dat})\backslash NP_{acc}) & ((S\backslash NP_{nom})\backslash NP_{dat})\backslash NP_{acc}\\
:\lambda p.p\,john & :\lambda p.p\,mary & :\lambda p.p\,pizza & :\lambda w\lambda x\lambda y.past\,(give\,w\,x\,y)\\
\cline{3-4}
 & & \multicolumn{2}{c}{(S\backslash NP_{nom})\backslash NP_{dat} : \lambda x\lambda y.past\,(give\,pizza\,x\,y)} \; > \\
\cline{2-4}
 & \multicolumn{3}{c}{S\backslash NP_{nom} : \lambda y.past\,(give\,pizza\,mary\,y)} \; > \\
\hline
\multicolumn{4}{c}{S : past\,(give\,pizza\,mary\,john)} \; >
\end{array}
$$

However, to capture scrambling of the kind seen in (1), we must allow raised categories to combine by the earlier composition rules (**??**), and in particular by the forward crossing rule (**??**c), indexed $>\mathbf{B}_{\times}$ in the following derivation:

(3)
$$
\begin{array}{cccc}
\text{John}-\text{ga} & \text{piza}-\text{o} & \text{Mary}-\text{ni} & \text{ageta.}\\
\hline
S/(S\backslash NP_{nom}) & (S\backslash NP_{nom})/((S\backslash NP_{nom})\backslash NP_{acc}) & (S\backslash NP_{nom})/((S\backslash NP_{nom})\backslash NP_{dat}) & ((S\backslash NP_{nom})\backslash NP_{dat})\backslash NP_{acc}\\
:\lambda p.p\,john & :\lambda p.p\,pizza & :\lambda p.p\,mary & :\lambda w\lambda x\lambda y.past\,(give\,w\,x\,y)\\
\cline{3-4}
 & & \multicolumn{2}{c}{(S\backslash NP_{nom})\backslash NP_{acc} : \lambda w\lambda y.past\,(give\,w\,mary\,y)} \; >\mathbf{B}_{\times} \\
\cline{2-4}
 & \multicolumn{3}{c}{S\backslash NP_{nom} : \lambda y.past\,(give\,pizza\,mary\,y)} \; > \\
\hline
\multicolumn{4}{c}{S : past\,(give\,pizza\,mary\,john)} \; >
\end{array}
$$

In the above derivation, the variable Y in the forward crossing composition rule (**??**c) matches the result $(S\backslash NP_{nom})\backslash NP_{dat}$ of the Japanese SXOV ditransitive category $((S\backslash NP_{nom})\backslash NP_{dat})\backslash NP_{acc}$, "canceling" the Ys to yield the Japanese SOV transitive category $(S\backslash NP_{nom})\backslash NP_{acc}$. (The rest of the derivation is purely applicative.)

To capture (1c) takes two crossing compositions:

(4)
$$
\begin{array}{cccc}
\text{piza}-\text{o} & \text{John}-\text{ga} & \text{Mary}-\text{ni} & \text{ageta.}\\
\hline
S/(S\backslash NP_{acc}) & S/(S\backslash NP_{nom}) & (S\backslash NP_{nom})/((S\backslash NP_{nom})\backslash NP_{dat}) & ((S\backslash NP_{nom})\backslash NP_{dat})\backslash NP_{acc}\\
:\lambda p.p\,john & :\lambda p.p\,pizza & :\lambda p.p\,mary & :\lambda w\lambda x\lambda y.past\,(give\,w\,x\,y)\\
\cline{3-4}
 & & \multicolumn{2}{c}{(S\backslash NP_{nom})\backslash NP_{acc} : \lambda w\lambda y.past\,(give\,w\,mary\,y)} \; >\mathbf{B}_{\times} \\
\cline{2-4}
 & \multicolumn{3}{c}{S\backslash NP_{acc} : \lambda y.past\,(give\,w\,mary\,john)} \; >\mathbf{B}_{\times} \\
\hline
\multicolumn{4}{c}{S : past\,(give\,pizza\,mary\,john)} \; >
\end{array}
$$

1. We might think of Japanese case morphemes like nominative "-ga" as bearing categories like $(S/(S\backslash NP_{nom}))\backslash\!\backslash NP$, where double-slash $\backslash\!\backslash$ as usual indicates a suffixal function applying in the lexicon only, like Latin "-us" of chapter 3.

It is important to notice that even in morphologically explicit languages like Japanese, case is still ambiguous between multiple verbal types selecting for that case. In general, there will be up to n raised categories for a given cased argument, including those raised over verbs needing all ordered subsets of arguments that include that argument itself. For example:

$$
\text{(5)}\quad
\begin{array}{ll}
\text{John-ga} & := S/(S\backslash NP_{nom}) : \lambda p.p\,john \\
\text{Mary-ni} & := (S\backslash NP_{nom})/((S\backslash NP_{nom})\backslash NP_{dat}) : \lambda p.p\,mary \\
 & := S/(S\backslash NP_{dat}) : \lambda p.p\,mary \\
\text{piza-o} & := ((S\backslash NP_{nom})\backslash NP_{dat})/(((S\backslash NP_{nom})\backslash NP_{dat})\backslash NP_{acc}) \\
 & := (S\backslash NP_{nom})/((S\backslash NP_{nom})\backslash NP_{acc}) : \lambda p.p\,pizza \\
 & := S/(S\backslash NP_{acc}) : \lambda p.p\,pizza
\end{array}
$$

For example, the Japanese accusative object "piza-o" requires all three of the above categories for the earlier derivations.[2]

Japanese case therefore does much the same job in the morphology as the English determiners seen in (**??**) do syntactically. Since this ambiguity is always resolved by the type of the argument the nominal applies to, it will often be convenient to similarly schematize over such same-case groups as $NP^{\uparrow}_{acc}$ etc.

To obtain scrambled orders like the following, we need one of the second-level generalization of the composition rules (**??**c), indexed as $>\mathbf{B}^2_{\times}$ in the derivation:

$$
\text{(6)}\quad
\begin{array}{cccc}
\text{Mary}-\text{ni} & \text{piza}-\text{o} & \text{John}-\text{ga} & \text{ageta.} \\
\hline
NP^{\uparrow}_{dat} & NP^{\uparrow}_{acc} & NP^{\uparrow}_{nom} & ((S\backslash NP_{nom})\backslash NP_{dat})\backslash NP_{acc} \\
: \lambda p.p\,mary & : \lambda p.p\,pizza & : \lambda p.p\,john & : \lambda w\lambda x\lambda y.past\,(give\,w\,x\,y) \\
\cline{3-4}
 & & \multicolumn{2}{r}{>\mathbf{B}^2_{\times}} \\
 & & \multicolumn{2}{c}{(S\backslash NP_{dat})\backslash NP_{acc} : \lambda x\lambda y.past\,(give\,w\,x\,john)} \\
\cline{2-4}
 & \multicolumn{3}{r}{>} \\
 & \multicolumn{3}{c}{S\backslash NP_{nom} : \lambda y.past\,(give\,pizza\,x\,john)} \\
\cline{1-2}
\multicolumn{2}{r}{>} & & \\
\multicolumn{2}{c}{S : past\,(give\,pizza\,mary\,y)} & &
\end{array}
$$

In general, scrambling n arguments past the subject in an SOX^{n-1}V construction, needs level-n composition rules. However, 2 is here conjectured to be a universal upper limit on that level.[3]

Similarly, in Latin, crossing composition allows the following "scrambled" variant of figure 3.1 of Chapter 3:

2. Schematization is possible as always.
3. We return to this point below.

(7)
$$\begin{array}{ccccc}
\text{Mur} & \text{+um} & \text{Balb} & \text{+us} & \text{dificat} \\
\text{wall} & \textit{.ACC} & \text{Balbus} & \textit{.NOM} & \text{build.pres.3s} \\
\multicolumn{2}{c}{\overline{S/(S\backslash NP_{acc})}^{<LEX}} & \multicolumn{2}{c}{\overline{S/(S\backslash NP_{nom,3s})}^{<LEX}} & \overline{(S_{pres}\backslash NP_{nom,3s})\backslash NP_{acc}} \\
 & & \multicolumn{3}{c}{\overline{\qquad\qquad S\backslash NP_{acc} \qquad\qquad}^{>\mathbf{B}_\times}} \\
\multicolumn{5}{c}{\overline{\qquad\qquad\qquad\qquad S \qquad\qquad\qquad\qquad}^{>}}
\end{array}$$

"Balbus is building a/the wall"

In most standard British and American dialects, English does not allow scrambling of NP arguments. Unlike Japanese and German cased arguments, English (and Welsh) case-raised NPs, including those in earlier derivations like (**??**), must be prevented from combining by crossed composition using the slash-types $/_{\diamond\star}$ and $\backslash_{\diamond\star}$, thereby blocking examples like the following:[4]

(8)
$$\begin{array}{cccc}
\text{*John} & \text{gave} & \text{pizza} & \text{Mary.} \\
\overline{S/_{\diamond\star}(S\backslash NP)} & \overline{((S\backslash NP)/NP)/NP} & \overline{(S\backslash NP)\backslash_{\diamond\star}((S\backslash NP)/NP)} & \overline{(S\backslash NP)\backslash_{\diamond\star}((S\backslash NP)/NP)} \\
 & \multicolumn{2}{c}{\overline{\qquad\qquad * \qquad\qquad}^{*<\mathbf{B}_\times}} &
\end{array}$$

However, English VP adjuncts like "without telling anyone" are unrestricted and can scramble, as in the following example of "Heavy NP-shift":

(9)
$$\begin{array}{cccc}
\text{Harry} & \text{filed} & \text{without telling anyone} & \text{any report longer than three pages.} \\
\overline{S/_{\diamond\star}(S\backslash NP)} & \overline{(S\backslash NP)/NP} & \overline{(S\backslash NP)\backslash(S\backslash NP)} & \overline{(S\backslash NP)\backslash_{\diamond\star}((S\backslash NP)/NP)} \\
 & \multicolumn{2}{c}{\overline{\qquad (S\backslash NP)/NP \qquad}^{<\mathbf{B}_\times}} & \\
 & \multicolumn{3}{c}{\overline{\qquad\qquad\qquad S\backslash NP \qquad\qquad\qquad}^{<}} \\
\multicolumn{4}{c}{\overline{\qquad\qquad\qquad\qquad S \qquad\qquad\qquad\qquad}^{>}}
\end{array}$$

5.2 Germanic scrambling and crossing dependencies

Under the assumption that case-related type-raising of arguments is universal, its inclusion in verb-final Germanic constructions like (17) above, where the arguments have scrambled from their heads across intervening verbs, correctly allows them to scramble over each other further to the left, under a generalization that Wallenberg (2009, 2013) formulated as follows:

(10) *The Generalized Holmberg Condition (*GHC*):*
Scrambling and object shift cannot move elements leftward past a c-commanding head.

4. Usually this detail of the notation can be ignored.

For example, in the German example (11), the object "die Lebensmittel" ("the groceries") can scramble out of the VP past the adjunct, as in (12):

(11) Johann hat [auf dem Markt] [die Lebensmittel gekauft].
Johann has at the market the groceries bought.
"Johann bought the groceries at the market."

(12) Johann hat [die Lebensmittel] [auf dem Markt] gekauft.
Johann has the groceries at the market bought.
"Johann bought the groceries at the market."

But "die Lebensmittel" cannot scramble further to the left past the main clause auxiliary "hat", which c-commands that VP:

(13) *Johann [die Lebensmittel] hat [auf dem Markt] gekauft.
Johann the groceries has at the market bought.
"Johann bought the groceries at the market."

This restriction stands in contrast to consistently verb-final Japanese, where we saw in (4) that the object can scramble all the way to the left periphery of the clause.

(12) is possible because the adjunct PP that intervenes between the scrambled object and the verb can compose with it to yield a category $VP\backslash NP$ adjacent to the case-raised object:[5]

(14)
$$\begin{array}{ccccc}
\text{Johann} & \text{hat} & \text{die Lebensmittel} & \text{auf dem Markt} & \text{gekauft} \\
\text{Johann} & \text{has} & \text{the groceries} & \text{at the market} & \text{bought.} \\
\hline
S/(S\backslash NP) & (S_{main}\backslash NP_{nom})/VP & VP/(VP\backslash NP_{acc}) & VP/(VP\backslash PP) & (VP\backslash PP)\backslash NP_{acc} \\
\cline{4-5}
 & & & \multicolumn{2}{c}{VP\backslash NP_{acc} \quad >\mathbf{B}_\times} \\
\cline{3-5}
 & & \multicolumn{3}{c}{VP \quad >} \\
\cline{2-5}
 & \multicolumn{4}{c}{S\backslash NP_{nom} \quad >} \\
\cline{1-5}
\multicolumn{5}{c}{S \quad >}
\end{array}$$

However, if the same object is to the left of the main verb, it is no longer adjacent to a category of the required type, even if the tensed verb is allowed to compose with the residue of scrambling. (In fact, it is only because arguments are lexically type-raised, and ground categories like NP_{acc} are excluded, that examples like (13) can in present terms be blocked:

5. The same is true if the PP is an adjunct.

(15) ∗Johann die Lebensmittel hat auf dem Markt gekauft
Johann the groceries has at the market bought.

$$\begin{array}{ccccc}
S/(S\backslash NP) & VP/(VP\backslash NP_{acc}) & (S_{main}\backslash NP_{nom})/VP & VP/(VP\backslash PP) & (VP\backslash PP)\backslash NP_{acc} \\
 & & & \multicolumn{2}{c}{\overline{\quad VP\backslash NP_{acc} \quad}^{>\mathbf{B}_\times}} \\
 & & \multicolumn{3}{c}{\overline{\quad (S_{main}\backslash NP_{nom})\backslash NP_{acc} \quad}^{>\mathbf{B}_\times}} \\
\multicolumn{4}{c}{\overline{\quad ***\quad}^{*}} &
\end{array}$$

There is no German morpholexically cased accusative category that can combine with “hat auf dem Markt gekauft”, since $(S_{main}\backslash NP_{nom})\backslash NP_{acc}$ is not a legal German main verb category. (The Germanic subordinate clause transitive category, which accusatives are also raised over, is $(S_{sub}\backslash NP_{nom})\backslash NP_{acc}$, while the main clause topicalized accusative is $S_t/(S_{inv}\backslash NP_{acc})$.

Wallenberg’s generalization (10) therefore depends on the fact that these are German main clauses, in which the legal orders are V2, departing from verb-finality. In the subordinate clauses of other Germanic languages whose control verbs mix verb finality and initiality, arguments can scramble leftward “past” tense For example, the following application-only derivation for the class of Swiss German object-control verbs that take bare infinitival complements resembles the derivation for the corresponding English sentence, except that VP is V-final:

(16) das mer d′chind lönd em Hans hälfe es huus aastriiche
that we-NOM the children-ACC let Hans-DAT help the house-ACC paint

$$\begin{array}{cccccccc}
 & NP_{nom} & NP_{acc} & ((S'\backslash NP_{nom})\backslash NP_{acc})/VP & NP_{dat} & (VP\backslash NP_{dat})/VP & NP_{acc} & VP\backslash NP_{acc} \\
 & & & & & & \multicolumn{2}{c}{\overline{\quad VP \quad}^{<}} \\
 & & & & & \multicolumn{3}{c}{\overline{\quad VP\backslash NP_{dat} \quad}^{>}} \\
 & & & & \multicolumn{4}{c}{\overline{\quad VP \quad}^{<}} \\
 & & & \multicolumn{5}{c}{\overline{\quad (S'\backslash NP_{nom})\backslash NP_{acc} \quad}^{>}} \\
 & & \multicolumn{6}{c}{\overline{\quad S'\backslash NP_{nom} \quad}^{<}} \\
 & \multicolumn{7}{c}{\overline{\quad S' \quad}^{<}}
\end{array}$$

“that we let the children help Hans paint the house” (Shieber, 1985)

However, we noted earlier that the inclusion of crossing composition and the second-level composition rules (**??**) made the grammar non–context-free. In the case of Swiss German above, as well as in similar constructions in Dutch and West Flemish, it can capture unboundedly-many crossing dependencies of the kind that constituted the basis for Shieber’s 1985 proof that human languages are not even weakly context-free, as in (17), which makes crucial use of the forward crossing second-level composition rule (**??**c) (Steedman, 1985, 2000b, 2020):

(17)

das	mer	d′chind	em Hans	es huus	lönd	hälfe	aastriiche
that	we-NOM	the children-ACC	Hans-DAT	the house-ACC	let	help	paint
	$NP^{\uparrow}_{nom}$	$NP^{\uparrow}_{acc}$	$NP^{\uparrow}_{dat}$	NP_{acc}	$((S'\backslash NP_{nom})\backslash NP_{acc})/VP$	$(VP\backslash NP_{dat})/VP$	$VP\backslash NP_{acc}$

$$(((S'\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat})/VP \quad >\mathbf{B}^2_{\times}$$
$$(((S'\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat})\backslash NP_{acc} \quad >\mathbf{B}_{\times}$$
$$((S'\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat} \quad >$$
$$(S'\backslash NP_{nom})\backslash NP_{acc} \quad >$$
$$S'\backslash NP_{nom} \quad >$$
$$S' \quad >$$

"that we let the children help Hans paint the house" (Shieber, 1985)

Such scrambling or partial free-order of arguments in Germanic and other languages has led to proposals for "verb projection raising" (Haegeman and van Riemsdijk, 1986; Wallenberg, 2009), clause-union (Haider, 2003, esp. §4), or compound "roll-up" or "remnant" varieties of movement (Koopman, 1996; Koopman and Szabolcsi, 2000), which combinatory derivations like the above reduce to serial contiguous compositional merger.

Steedman, 2020:646-7 examines in greater detail the scrambling possibilities under this generalization for the Germanic constructions exemplified above, following Wurmbrand (2004), and in Hungarian clause union, following Koopman and Szabolcsi. Wallenberg's own analysis (2009:166-167) of such examples is more complicated, assuming head movement of the verbs to adjoin to their respective vPs, followed by raising of the entire "Verb Projection", an argument structure including multiple verb traces, to spec of TP (den Dikken, 1994; Den Dikken, 1995).[6]

It is important in connection with the theories of these authors to note that the derivations (16) and (17) are unaffected by the fact that the arguments are lexically type-raised by case-morphology, which as usual merely reverses the direction of function application mergers. However, the involvement of type-raising does allow some further derivations for the latter, because the case-raised argument categories can also compose—for example, the following:

6. Wallenberg also replaces the GHC by a more general principle of Conservation of C-Command (CoCC).

(18)

das	mer	d'chind	em Hans	es huus	lönd	hälfe	aastriiche
that	we-NOM	the children-ACC	Hans-DAT	the house-ACC	let	help	paint
	$NP^{\uparrow}_{nom} : \lambda p.p\,us$	$NP^{\uparrow}_{acc} : \lambda p.p\,(the\,kids)$	$NP^{\uparrow}_{dat} : \lambda p.p\,hans$	$NP^{\uparrow}_{acc} : \lambda p.p\,(the\,house)$	$((S'\backslash NP_{nom})\backslash NP_{acc})/VP : \lambda p\lambda x\lambda y.let(p\,x)\,x\,y$	$(VP\backslash NP_{dat})/VP : \lambda p\lambda x\lambda y.help(p\,x)\,x\,y$	$VP\backslash NP_{acc} : \lambda x\lambda y.paint\,x\,y$

$$\frac{NP^{\uparrow}_{nom} \quad NP^{\uparrow}_{acc}}{S'/((S'\backslash NP_{nom})\backslash NP_{acc}) : \lambda p.p\,(the\,kids)\,us} >\mathbf{B}$$

$$\frac{}{S'/(((S'\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat}) : \lambda p.p\,hans\,(the\,kids)\,us} >\mathbf{B}$$

$$\frac{}{S'/((((S'\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat})\backslash NP_{acc}) : \lambda p.p\,(the\,house)\,hans\,(the\,kids)\,us} >\mathbf{B}$$

$$\frac{lönd \quad hälfe}{(((S'\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat})/VP : \lambda p\lambda w\lambda x\lambda y.let(help\,(p\,w)\,w\,x)\,x\,y} >\mathbf{B}^2_{\times}$$

$$\frac{}{(((S'\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat})\backslash NP_{acc} : \lambda v\lambda w\lambda x\lambda y.let\,(help\,(paint\,v\,w)\,w\,x)\,x\,y} >\mathbf{B}_{\times}$$

$$\frac{}{S' : let\,(help\,(paint\,(the\,house)\,hans)\,hans\,(the\,kids))\,(the\,kids)\,us} >$$

"that we let the children help Hans paint the house" (Shieber, 1985)

Interestingly, the second-order argument-cluster category $S'/((((S'\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat})\backslash NP_{acc})$ that is built by successive composition mergers for "mer d'chind em Hans es huus", with the logical form $\lambda p.p\,house\,hans\,children\,us$, can be viewed as the syntactic *type* of Haegeman and van Riemsdijk's and Wallenberg's raising "verb projection". with the eliminable second-order bound variable p doing the work of head-movement at the level of lf, However, this mechanism for clause union is type-dependent, rather than structure- dependent, and is entirely blind to the derivation of the composite verb *lönd hälfe aastriiche* of category $(((S'\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat})\backslash NP_{acc}$ required to instantiate the bound variable p, which reduces with it by application merge, rather than movement, and without the involvement of verb traces (cf. Epstein, Groat, Kawashima, and Kitahara, 1998; Epstein and Seely, 2006:178-179.

Haegeman and van Riemsdijk, 1986:432 discuss alternative orders for the following subordinate clause from Zürich German for a clause meaning "(that) he wants to let his children study medicine", for which the first (standard German-like) order (a) and the last order (g) are deprecated:[7]

(19) a. *

(das)	er	sini chind	mediziin	studiere	laa	wil
(that)	he	his children	medicine	study	let	wants
(S'/S')	$NP^{\uparrow}_{nom}$	$NP^{\uparrow}_{acc}$	$NP^{\uparrow}_{acc}$	$VP\backslash NP_{acc}$	$(VP\backslash NP_{acc})/_{\times\star}VP$	$(S'\backslash NP_{nom})/_{\times\star}VP$

7. These derivations crucially involve the generalization of the composition rules to second-order rules discussed in note **??**, more specifically, the forward crossing rule shown there as (i), here indicated as $>\mathbf{B}^2_{\times}$.

b.

(das)	er	sini chind	mediziin	wil	laa	studiere
(that)	he	his children	medicine	wants	let	study
	$NP^{\uparrow}_{nom}$	$NP^{\uparrow}_{acc}$	$NP^{\uparrow}_{acc}$	$(S\backslash NP_{nom})/_{\times\star}VP$	$(VP\backslash NP_{acc})/_{\times\star}VP$	$VP\backslash NP_{acc}$

$$\text{laa studiere: } ((VP\backslash NP_{acc})\backslash NP_{acc}) \quad >\mathbf{B}_{\times}$$
$$\text{wil laa studiere: } ((S\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{acc} \quad >\mathbf{B}^2_{\times}$$
$$\text{mediziin wil laa studiere: } (S\backslash NP_{nom})\backslash NP_{acc} \quad >$$
$$\text{sini chind} \ldots\text{: } S\backslash NP_{nom} \quad >$$
$$\text{er} \ldots\text{: } S \quad >$$

c.

(das)	er	sini chind	wil	mediziin	laa	studiere
(that)	he	his children	wants	medicine	let	study
	$NP^{\uparrow}_{nom}$	$NP^{\uparrow}_{acc}$	$(S\backslash NP_{nom})/_{\times\star}VP$	$NP^{\uparrow}_{acc}$	$(VP\backslash NP_{acc})/_{\times\star}VP$	$VP\backslash NP_{acc}$

$$\text{laa studiere: } ((VP\backslash NP_{acc})\backslash NP_{acc}) \quad >\mathbf{B}_{\times}$$
$$\text{mediziin laa studiere: } VP\backslash NP_{acc} \quad >$$
$$\text{wil mediziin laa studiere: } (S\backslash NP_{nom})\backslash NP_{acc} \quad >\mathbf{B}_{\times}$$
$$\text{sini chind} \ldots\text{: } S\backslash NP_{nom} \quad >$$
$$\text{er} \ldots\text{: } S \quad >$$

d.

(das)	er	sini chind	wil	laa	mediziin	studiere
(that)	he	his children	wants	let	medicine	study
	$NP^{\uparrow}_{nom}$	$NP^{\uparrow}_{acc}$	$(S\backslash NP_{nom})/_{\times\star}VP$	$(VP\backslash NP_{acc})/_{\times\star}VP$	$NP^{\uparrow}_{acc}$	$VP\backslash NP_{acc}$

$$\text{mediziin studiere: } VP \quad >$$
$$\text{laa mediziin studiere: } VP\backslash NP_{acc} \quad >$$
$$\text{wil laa mediziin studiere: } (S\backslash NP_{nom})\backslash NP_{acc} \quad >\mathbf{B}_{\times}$$
$$\text{sini chind} \ldots\text{: } S\backslash NP_{nom} \quad >$$
$$\text{er} \ldots\text{: } S \quad >$$

e.

(das)	er	wil	sini chind	mediziin	laa	studiere
(that)	he	wants	his children	medicine	let	study
	$NP^{\uparrow}_{nom}$	$(S\backslash NP_{nom})/_{\times\star}VP$	$NP^{\uparrow}_{acc}$	$NP^{\uparrow}_{acc}$	$(VP\backslash NP_{acc})/_{\times\star}VP$	$VP\backslash NP_{acc}$

$$\text{laa studiere: } ((VP\backslash NP_{acc})\backslash NP_{acc}) \quad >\mathbf{B}_{\times}$$
$$\text{mediziin laa studiere: } VP\backslash NP_{acc} \quad >$$
$$\text{sini chind mediziin laa studiere: } VP \quad >$$
$$\text{wil} \ldots\text{: } S\backslash NP_{nom} \quad >$$
$$\text{er} \ldots\text{: } S \quad >$$

f.

(das)	er	wil	sini chind	laa	mediziin	studiere
(that)	he	wants	his children	let	medicine	study
	$NP^{\uparrow}_{nom}$	$(S\backslash NP_{nom})/_{\times\star}VP$	$NP^{\uparrow}_{acc}$	$(VP\backslash NP_{acc})/_{\times\star}VP$	$NP^{\uparrow}_{acc}$	$VP\backslash NP_{acc}$

$$\text{mediziin studiere: } VP \quad >$$
$$\text{laa mediziin studiere: } VP\backslash NP_{acc} \quad >$$
$$\text{sini chind laa mediziin studiere: } VP \quad >$$
$$\text{wil} \ldots\text{: } S\backslash NP_{nom} \quad >$$
$$\text{er} \ldots\text{: } S \quad >$$

g. *	(das)	er	wil	laa	sini chind	mediziin	studiere
	(that)	he	wants	let	his children	medicine	study
		$NP^{\uparrow}_{nom}$	$(S\backslash NP_{nom})/_{\times\star}VP$	$(VP\backslash NP_{acc})/_{\times\star}VP$	$NP^{\uparrow}_{acc}$	$NP^{\uparrow}_{acc}$	$VP\backslash NP_{acc}$
						VP (>)	
					$\star$		
				$\star$			
			$\star$				

In the latter case (g), type-raised “sini chind” cannot be instantiated as $VP\backslash(VP\backslash NP_{acc})$ and compose by $<\mathbf{B}_{\times}$, because the necessary category-type is non-order-preserving, but also type-preserving over the result, and is therefore not available in CCG. (See discussion of the topicalized object category in chapter 9 CHECK).

Many of the above alternates differ in the possibilities for positioning prosodic boundaries and information-structurally relevant properties such as the definiteness of NPs.

We noted earlier that to restrict a CCG grammar to a single language, the only location for language specific information is the lexicon. It is striking that the variety of word-order found in Zürich German raising subordinate clauses, a construction that has provided the classic proofs of strong trans-context-free grammar (Huybregts, 1984; Shieber, 1985), can be captured in such a simple lexicon, with one directional category per verb, and that the complex process of “reanalysis” invoked by Haegeman and van Riemsdijk can be replaced by the independently-motivated rules of composition—crucially, crossing composition—reducing Reanalysis to Movement, and Movement in turn to contiguous adjacent merger.

The Zürich German alternation exemplified above is closely mirrored in West Flemish (Haegeman, 1992), and in German and Dutch by the *zu/te*-infinitival complement verbs such as *proberen/probeeren* (“try”).[8]

However, the small set of German/Dutch bare-infinitival verbs like *sien/zien*, (“see”) are more restricted, allowing only orders in which all NPs precede all verbs as in (19a,b) (the order of the verbs may vary Bech, 1955; Evers, 1975; Bresnan, Kaplan, Peters, and Zaenen, 1982; Seuren, 1985), and disallowing alternations like (19c,d,e,f, and g).

The idiosyncrasy of these verbs can be captured if we assume that the small class of Dutch and German verbs that require bare infinitival complements, like *zag voeren* and *zag helpen voeren* in examples like the following are actually lexicalized in those languages, with categories like $((S\backslash NP)\backslash NP)\backslash NP$ and $(((S\backslash NP)\backslash NP)\backslash NP)$, with the small closed class of verbs like *zien* and *helpen*

8. The analysis of Haegeman, 1992:193 in terms of “head adjunction” is in fact very similar to the present account in terms of serial verb composition.

and their German equivalents acting as morphemes:

(20) a. dat ik Henk de paarden zag voeren "that I saw Harry feed the hippos"
that I Harry the horses saw feed

b. dat ik Henk Cecilia de paarden zag helpen voeren "that I saw Harry help Cecilia feed the horses"
that I Harry Cecilia the horses saw help feed

Some support for the proposal may be found in the fact that for those Dutch infinitival verbs like *probeeren* that allow the alternate orders, when the *te*-infinitival complement itself consists of serial infinitivals, these *cannot* carry the *te*-complementizer, consistent with the idea that the bare serial infinitivals are morphemic (cf. Steedman, 2000b:144-146):

(21) a. dat hij probeerde Jan *(te) leren het lied *(te) zingen.
b. dat hij probeerde Jan het lied *(te) leren (*te) zingen.
"that he tried to teach Jan to sing the song"

A similar suggestion concerning the Dutch bare infinitival construction was made by Moortgat (1988). Williams (2003):231-234 reaches a similar conclusion concerning certain cases of forbidden inversion, under his combinatory categorial analysis of a related class of Hungarian serial verbs, originally analyzed in terms of "remnant movement" by Koopman and Szabolcsi (2000).

Williams (2003):231-234 points out that this restriction, together with a related restriction on inversion in Hungarian verb-complexes originally analyzed in terms of "remnant movement" by Koopman and Szabolcsi (2000), cannot be captured in his categorial calculus CAT, and proposes a lexicalized solution equivalent to the present one.

5.3 Discussion

It should not be assumed that word-order in truly non-configurational languages such as Warlpiri (Hale, 1983; Jelinek, 1984) can be accounted for solely in terms of type-raising and composition, as has been argued above for Japanese and Germanic word-order variation. Jelinek claims that a number of properties characteristic of non-configurational languages, including freedom to drop nominal arguments entirely, and for nominals to behave like freely ordered adjuncts, stem from the fact that nominal dependency is essentially anaphoric to lexically cliticized pronominals associated with tense, and hence not subject to syntactic linearization constraints of the kind medled in CCG. There is considerable disagreement on how nonconfigurationality should be handled (Marácz and Muysken, 1989). We return to the question of anaphora and coreferentiality in chapter 14 below.

Chapter 6

Intonation Structure

Notice that I am sweeping under the rug questions . . . involving topic-focus and theme-rheme structures, figure-ground properties, effects of adjacency and linearity, and many others
—*The Minimalist Program* Noam Chomsky, 1995:220

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 intuirion, they usually 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 follows:

(1) 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 rise 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 must be the same as information 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.

The evolution of the view in linguistic theory of the relation between syntactic and prosodic structure has gone from the early isomorphism position of Chomsky, 1955/1975; Chomsky, Halle, and Lukoff, 1956; Chomsky and Halle, 1968; Bresnan, 1971; Selkirk, 1972, and Kaisse, 1985, through the total autonomy of Bolinger, 1972a; Jackendoff, 1972, and Selkirk, 1984 (where prosody was subject only to non-syntactic "sense-unit" conditions), to the partial autonomy of "edge-based" theories of Selkirk, 1986, 1990 and Truckenbrodt, 1999, and the return to the contemporary syntactic isomorphism of Match Theory (Wagner, 2005; Selkirk, 2011), which assumes that every language has a single syntax-derived default prosody, and that semantically unmotivated deviations from the default prosody are found only to the extent that they are an optimal solution to various purely phonological or focus-related ranked optimality-theoretic (OT) "markedness" constraints (although Selkirk, 2011:471 makes it clear that such considerations can unduce phonological-domains that do not correspond to syntactic phrases).

Throughout this development, the element that has remained essentially unchanged has been the assumptions concerning the nature of constituency and surface syntactic structure. Most recently, a foundational principle of the Minimalist Program of Chomsky (1995b) has been that the properties of syntax are largely if not entirely determined by the need to create objects that are both phonologically and semantically well-formed. Following Chomsky (2001) Cinque, 1993; Legate, 2003; Adger, 2007; Richards, 2010, 2016, identifiy the notion of "phase", the domain of "spell out", with the domain of the phonological phrase.

This principle seems entirely reasonable. What else is syntax for but to specify the computation of meanings from the sounds of a language, and vice versa, and to do so in the most efficient way possible? The notions of Phase and Spell-Out, are related to the earlier notion of the transformational cycle, which are all related to the categorial notion of lexical domain.

The point of this chapter is to argue that the isomorphism assumption is correct. However, these authors are trying to impose isomorphism to the wrong syntactic structures. As previous chapters have shown, syntactic derivations are more diverse than traditional surface syntax allows. Traditional notions of syntactic structure as embodied in such touchstones as lexical substitutability, actually arise from some rather unreliable intuitions about the underlying semantics of logical form, a level at which there really is something special about the subject.

It follows that we should be able to hang intonation structure more directly into CCG derivations of the kind we have seen in the earlier chapters. It also follows that the semantics of topic and comment or information structure can be combined with the rest of logical form in a unified compositional semantics. To do this, we need a notation for intonation structure.

6.1 The Autosegmental-Metrical account of English Intonation Structure

Selkirk (1984), Hirschberg and Pierrehumbert (1986), Pierrehumbert and Hirschberg (1990), and the present author, have offered different but related accounts of intonation structure in English and some other languages using the the system of abstract "tones" identified by Liberman (1975) and Pierrehumbert (1980), as modified by Pierrehumbert and Beckman (1988) and in particular Pierrehumbert and Hirschberg (1990), to the last of which which the reader is directed for details of some typical pitch contours corrsponding to the prosodic notation.

These accounts share the assumption that intonation has as transparent and type-driven a semantics as do words and phrases. While the semantics of intonation in English concerns information structure and propositional attitude, rather than the predicate-argument relations and operator-scope relations that are familiar from standard semantics and the preceding chapters, this information-structural semantics is fully compositional. The present chapter will show that it can be regarded as a component of the same compositional semantic system.

The chapter follows Steedman (1991, 2000a, and 2014, hereafter *CSEI*) in developing a semantics for intonation structure that is fully integrated into the rest of CCG. This grammar treats intonation structure as identical to surface derivational structure, and the associated information structure as a component of the standard Montague-style compositional semantics, even when the intonation structure departs from the restrictions of traditional surface structure. Many of the diverse discourse meanings that have been attributed to intonational tunes are then shown to arise via conversational implicature from more primitive literal meanings distinguished along three dimensions of information structure, speaker/hearer agency, and presence or absence of the corresponding information unit from the common ground.

6.1.1 Accents

It is standard to assume, following Bolinger (1958, 1961) and Halliday (1963, 1967a,b), that accents are properties of the *words* that they fall on, and that they mark the interpretation of those words as distinguishing the speaker's actual utterance from other things that they might be expected to have said in the context to hand, as in the "Alternative Semantics" of Karttunen (1976), Karttunen and Peters (1979), Rooth (1985, 1992), and Büring (1997a,b).[1] In this sense, all accents are *contrastive*. For example, in response to the question "Which finger did he bite?", the word that contributes to distinguishing the following answer from other possible answers via reference is the deictic "this", so the following intonation is appropriate.

(2) He bit THIS one .
          H*   LL%

We use the tonal notation of Pierrehumbert and colleagues to identify the contrastive accent H*, which in many speakers is marked by a relatively high maximum pitch, as well as the low level final boundary LL%. However it is important to be clear that we do not assume that the accents we indicate in this way are invariably marked by variation in F_0 pitch. Other dimensions such as lengthening and alignment are equally important (Hart, Collier, and Cohen, 1990; Calhoun, 2012), and some speakers use little or no pitch variation (Calhoun, 2010). It is also known that accent can be conveyed by whispered speech (Meyer-Eppler, 1957; Thomas, 1969; Higashikawa and Minifie, 1999), which entirely lacks F_0.

Nevertheless, trained speakers such as actors, lawyers, and radio newsreaders do tend to use F_0 pitch, so we use tones as an abstract notation denoting a more general range of accent markers.

It is also important to be clear from the start that the set of alternative utterances from which the actual utterance is distinguished by prosody is in no sense the presumably unboundedly large set of all possible utterances appropriate to this context. Rather, the set of alternative utterances presupposed by the speaker is either established by the context such as a question, or *accommodated* by the hearer in the sense of Lewis (1979b) and Thomason (1990).

1. The term "accent" is here restricted to what Ladd (2008) calls "primary" accents, sometimes called "nuclear" accents (although there may be more than one in a sentence). Ladd follows Bolinger and many others in distinguishing primary accents from certain other accents that arise from the interaction of lexical stress with the metrical grid. While there is still no objective measure to distinguish the two varieties, it is the primary accents that are perceived as emphatic or contrastive.

This does not imply that such alternative sets are confined to things that have been mentioned, or that they are mentally enumerated by the participants—or even that their extensions are finite.

In terms of Halliday's given/new distinction, accents are markers of "new" information, although the words that receive accent may have been recently mentioned, and it might be better to call them markers of "not given" information. The latter locution seems a little cumbersome, so the present work follows Vallduví and Vilkuna 1998 (although not their spelling) in using the term "contrast" for the property of English words bearing accents and "background" or "non-contrast" for words that do not carry accent.

There are two further independent binary-valued semantic dimensions along which the literal meanings of the various accent-types are distinguished. The first of these dimensions has been identified in the literature under various names, and distinguishes between what from now on will be referred to as "theme" and "rheme" components of the utterance, using these terms in the sense of Bolinger (1958, 1961) rather than Halliday.

The theme can be thought of informally as the part of the sentence corresponding to a question or topic that has been established by the context, and more formally as the speakers claim that either the speaker or the hearer supposes that information unit to be (or not to be) common ground. The rheme can be thought of informally as the part of the utterance that constitutes the speaker's novel contribution on that question or topic, or more formally as the speakers claim that either the speaker or the hearer makes (or fails to make) that information unit common ground.[2]

A great deal of the literature on information structure can be summarized as distinguishing two dimensions corresponding to the background/contrast and theme/rheme distinctions, although the consensus has tended to be obscured by the very different nomenclatures that have been applied. (See discussion by Kruijff-Korbayová and Steedman (2003) for a summary of this ramifying terminology and its lines of descent, along with some contiguous semantic influences.)

There is a further dimension of discourse meaning along which the accent types are distinguished which is less widely identified in this literature. It concerns whether or not the particular theme or rheme to hand is claimed to already be or to come to be mutually believed or in the Common Ground in the terms of Lewis (1969), Stalnaker (1978, 2002) Cohen (1978), Clark and

2. The theme in this sense differs from the notion of topic as defined by, for example, Gundel (1974); Gundel and Fretheim (2001) in being *speaker-defined rather than text-based*.

Marshall (1981) and Clark (1996).[3]

Both of these accent-related components of meaning are projected by the process of grammatical derivation defined in pervious chapters from the words that carry the accent in question to the prosodic phrase corresponding to these information units, following *CSEI*.

6.1.2 Boundaries

The intonational boundaries which delimit the prosodic phrase, such as those sometimes referred to as "continuation rises," mark a different property of the information units corresponding to intonational phrases. Although boundary tones are necessrily co-articulated with words, they themselves are analysed as autonomous string elements analogous to punctuation, rather than properties of words like accent.

According to the present theory, the boundary tones fall into two classes respectively distinguishing the speaker or the hearer as responsible for the supposition or update associated with the corresponding information unit.[4]

The choice of these terms reflects a view of the discourse context established by and refered to by a speaker as a database or set of propositions in some convenient logical language, with the λ-calculus acting as usual as a "glue-language", in which the semantics of themes and rhemes can be defined as functions. This set is divided into two subdomains. The first is a set S of information units that the speaker claims *they themselves* either suppose to be common ground (themes), or make common ground (rhemes). The second is a set H of themes and rhemes which the speaker claims *the hearer* either supposes to be or makes common ground.

Themes and rhemes are further distinguished on a dimension $\top/\bot$ according to whether the speaker claims a theme to be or not to be (or in the case of a rheme, to come to be or not to come to be) in the common ground.

It is important to be clear that the claims that the speaker makes about the discourse context and what is going on in the common ground are entirely distinct from the speaker and hearers actual beliefs including actual mutual belief. While the former may in the most neutral cases coincide with the latter, we shall see that speakers frequently make claims about the state of the context which are blatently false.

3. Stalnaker (2002) attributes the term to Grice (1967/1989). Hobbs (1990) proposes a different revision of Pierrehumbert and Hirschberg (1990) to the present one, and also gives a central role to Mutual Belief.

4. Earlier papers refer to this relation of the speaker and hearer to the information units as "ownership" and/or "commitment"—cf. Gussenhoven (1983, p.201) and Gunlogson (2001, 2002)).

These classifications of the accents and boundaries can be set out diagrammatically as in the tables 6.1 and 6.2 (adapted from CSEI), in which θ signifies theme, ρ signifies rheme, and $[S]$ and $[H]$ respectively denote speaker and hearer agency. If a theme or a rheme is marked $\top$ then it is claimed by the

	$\top$	$\perp$
θ	L+H*	L*+H
ρ	H*, (H*+L)	L*, (H+L*)

Table 6.1: The Literal Meaning of the Accents

S	L, LL%, HL%
H	H, HH%, LH%

Table 6.2: The Literal Meaning of the Boundaries

speaker to be (or to come to be) in the common ground, whoever is explicitly claimed to suppose or make it so, *regardless of whether anyone actually thinks of it that way*. If it is marked $\perp$, then it is claimed by the speaker not to be or become common ground *even if speaker and hearer both in fact believe it*.

Such apparent dissonances are very common in spoken dialog, and are a source of widespread indirection, or conversational implicature in intonational meaning. They have a lubricating effect on the dialog, eliminating friction and facilitating exchange, to which we turn next.

6.2 Intonation and Implicature

It might seem at first glance that the above is a very impoverished account of intonational meaning in comparison to the vast literature that exists on the subject. Whatever happened to notions such as "topic continuation," "other-directedness," "floor-claiming," "turn-yielding," "discourse-structuring," "evaluation with respect to subsequent material," "politeness," "face," "deixis," "commitment," "uncertainty," "affect", "ownership", "indirection," and even "questioning," all of which have been attributed to the various tones in the literature (Kruijff-Korbayová and Steedman, 2003).

The claim of the present work, following *CSEI*, is that all of these aspects of meaning arise as *conversational implicatures* of the basic information structural meanings identified in the earlier sections.

As an example of how this works, the entire range of possibilities allowed by the markers in tables 6.1 and 6.2, and some typical conversational implicatures are illustrated via some yet simpler examples in which tones including the L accents and boundaries are systematically varied across the same text.

If we limit ourselves for the sake of simplicity to tunes with a single accent, assume that H*+L and H+L* are not distinct from H* and L*, and take LL% and LH% as representative of the two S(peaker) and H(earer) responsibility classes of boundary then the classification in tables 6.1 and 6.2 allow eight tunes which exemplify the eight possible combinations of these three binary features. It is instructive to consider the conversational effect of these tunes when applied to the same sentence "I'm a millionaire," uttered in response to various prompts.

It is important to realize that all these conversational effects are indirect, and their force depends on whether the participants regard being a millionaire as counting as being rich.

(3) H: Are you rich?

S: I'm a MILLIONAIRE.
H* LL%

$\top(\rho\,(*millionaire\ me)\,S)$

"I make it common ground that I'm a millionaire."
(implicature: of course.)

(4) H: Are you poor?

S: I'm a MILLIONAIRE.
L* LL%

$\bot(\rho\,(*millionaire\ me)\,S)$

"I fail to make it common ground that I'm a millionaire."
(implicature: of course not.)

(5) H: Congratulations. You're a millionaire.

S: I'm a MILLIONAIRE?
H* LH%

$\top(\rho\,(*millionaire\,me)\,H)$

"You make it common ground that I'm a millionaire."
(implicature: really?)

(6) H: Congratulations. You're a millionaire.

S: I'm a MILLIONAIRE?
L* LH%

$\bot(\rho\,(*millionaire\,me)\,H)$

"You fail to make it common ground that I'm a millionaire."

(implicature: surely not)

The above four responses can be assumed to each consist of a solitary rheme.[5] The ones involving an L* accent mark the rheme as being not agreed. However, the accent itself does not distinguish who the opposition is coming from. This is not an ambiguity in the accent itself. Rather, the identification of the source of the conflict and the entire illocutionary force of the response depends on inference on the basis of what else is known about the participants' beliefs. Thus, in (4), the one who appears to doubt the proposition in the second utterance is the hearer, but in (6) it is the speaker. In different contexts, the difference could be reversed or eliminated.

A similar pattern can be observed for the theme accents:

(7) H: Are you rich?

S: I'm a MILLIONAIRE.
L+H* LL%

$[S^{+}]\theta\ *millionaire\,me$

"I suppose it to be common ground that I'm a millionaire."

(implicature: You know I am)

(8) H: Are you poor?

S: I'm a MILLIONAIRE.
L*+H LL%

$\bot(\theta\,(*millionaire\,me)\,S$

"I fail to suppose it to be common ground that I'm a millionaire."

(implicature: You ought to know I'm not)

5. Under the proposals in Steedman 2000a and *SCEI*, they could also be analyzed as an unmarked theme "I'm" and a rheme "a *millionaire". In this particular context it makes very little difference, and we'll ignore these readings.

(9) H: You appear to be a complete jerk.
S: I'm a MILLIONAIRE.
L+H* LH%
$\top(\theta\,(*millionaire\,me)\,H)$
"You suppose it to be common ground that I'm a millionaire."
(implicature: obviously not)

(10) H: You appear to be a complete jerk.
S: I'm a MILLIONAIRE.
L*+H LH%
$\perp(\theta\,(*millionaire\,me)\,H)$
"You fail to suppose it to be common ground that I'm a millionaire."
(implicature: Think again.)

At first encounter, it may appear that these tunes must mark rhemes, like those in (3) to (6). However, these utterances in fact seem to be solitary themes. Since uttering a truly common-ground theme with no rheme is by definition redundant, they are typically only uttered in situations where the stated theme is *not* in fact common ground. They then seem to achieve the effect of a response (as well as various other implicatures of impatience, diffidence, incompleteness, etc.) via *indirection*, by leaving the hearer to generate the implicated rheme, contradicting what they just said.

As before, the tunes involving L*+H accents imply absence from the common ground. In the case of (9) and (10), it is important to remember that the speaker's LH% boundary means only that the speaker *views the hearer as* supposing or not supposing these themes to be common ground. As far as the hearer is concerned, that is not the same as a statement about their actual beliefs. Thus the L*+H in (10) simply has the effect of correctly excluding from common ground this theme, which the boundary marks as in H, in spite of the fact that it can also be inferred to be in the speaker's own beliefs S. This is the possibility that was noticed in the discussion of tables 6.1 and 6.2: it seems a fundamental property of the system that there is a distinction between a proposition merely being in both S and H and it actually being in the common ground. The former amounts to a claim by the speaker that both participants *ought* to suppose it. The latter is a claim by the speaker that both actually *do* suppose it.

Other languages are free to draw the same distinctions syntactically, rather than intonationally, though there is a strong crosslinguistic tendency to associate accent with contrast. Thus, Hungarian appears to use fronting for both theme/topic and rheme/focus marking (É. Kiss, 1994). Italian, Spanish and Catalan use final position to mark rheme (Vallduví, 1990; Zubizarreta, 1998; Steedman, 2014). Calhoun (2015) reports a related prosodic system for the VSO Austronesian language Samoan in which rheme/focus marking is either via fronting or via final order, while Otsuka (2005) reports a similar effect of scrambling VSO order in the related VSO ergative language Tongan. Braun (2006) reports a similar system to English for German, with fronting used to mark theme/topic and distinctive accent tones for theme-contrast and rheme contrast.

6.3 Intonation Structure in Child Language Acquisition

Fisher and Tokura (1996) observed the following utterances in caretaker speech to a 13-14month-old (that is, preverbal) infant:[6]

(11) a. That looks like a DOGGY.
b. (You LIKE) (the doggy)
H* L LL%

The authors note that infant caregivers tend to use very exaggerated prosody when talking to infants, particularly at stages prior to the infant's using speech themselves, suggesting that prosodic phrasing is helpful in child language acquisition. They point out that it is therefore surprising to see that the prosodic phrase boundary in (11b) does not coincide with the traditional syntactic boundary between subject and predicate, but rather makes the subject and verb one phonological phrase, and the object the other.

Fisher and Tokura's explanation for this is that the infants are also (and at this stage perhaps *only*) learning information structure—that in the context of (11a), *the doggy* in (b) is in present terms the topic or theme (given or non-contrastive, hence without accent), while the comment or rheme is the child's liking for it.

Their claim is clearly correct, as far as the information structure goes, but it does makes the child's task of language aquisition seem unreasonably hard. Even if they only have to learn syntactic structure at a later stage, they are eventually have to learn about *two* structures, the information-structural and the syntactic, whose phrase boundaries actually intercalate or cross, and which must therefore have quite orthogonal compositional semantics.

6. The notation has been changed to be consistent with the rest of the chapter.

This does not seem reasonable. Any adult who was expected to learn a logic or programming language with two independent sets of ontercalating brackets with independent semantics would simply refuse, and tell its designer to come up with something more sensible. Why don't the children do the same, or at least grow up with undesirably cynical attitudes towards the adult world?

The resolution of this paradox should be obvious. According to the present theory, there is no such mismatch between syntactic derivation on the one hand, and information structure as marked by intonation on the other. The structure indicated in (11b) is just as legitimate a syntactic derivation as the one corresponding to the traditional subject-predicate division. It is also the derivation that partitions the logical form in a way consistent with the state of the discourse contest, thus:

(12)

$$
\begin{array}{ccccc}
\text{You} & \underset{\text{H}*}{\text{LIKE}} & \text{L} & \text{the doggy} & \underset{\text{LL\%}}{.} \\
\overline{S/(S\backslash NP) : \lambda p.p\,you} & \overline{(S\backslash NP)/NP : like} & & \overline{NP^{\uparrow} : \lambda p.p\,sk_{dog}} & \overline{} \\
\multicolumn{2}{c}{\overline{\hspace{14em}}{>}\mathbf{B}} & & & \\
\multicolumn{2}{c}{S/NP : \lambda x.like\,x\,you} & & & \\
\multicolumn{4}{c}{\overline{\hspace{26em}}{<}} & \\
\multicolumn{4}{c}{S : like\,sk_{dog}\,you} & \\
\end{array}
$$

Similarly, the non-standard constituents arising from the preposition-stranding category (65) in chapter 9 can also be marked as prosodic phrases, as in the following example (from Shattuck-Hufnagel and Turk, 1996:201):

(13)

Sesame Street is BROUGHT to you by,	the Children's TELEVISION Network.
L+H* LH%	H* LL%

Here, L+H* is a late-onset high accent, while H* is a simple high accent, LH% is a rising boundary between the two, and LL% is a final level low boundary.

All that we need to do to make this claim complete is to say how intonational markers of this degree of unambiguity limit CCG so that *only* the above derivation is allowed with this intonation.

6.4 Making CCG derivation Specify Intonation Contour

The availability of fully interpreted nonstandard derivational constituents corresponding to substrings like *you like* and *Manny married* was originally motivated by their participation in constructions like relativization and coordination and the desire to capture those constructions with a grammar obeying a very strict form of the Constituent Condition on Rules, without rules of "action at a distance", like movement and deletion.

However, a theory that allows alternative derivations like (21) and (22) in chapter 4.1 is clearly immediately able to capture the fact that prosody can make exactly the same non-standard constituents into intonational phrases, as in (14a), as easily as the standard constituents in (14b):

(14) a. Manny MARRIED   A MILLIONAIRE
  L+H* LH% H* LL%

b. MANNY married A MILLIONAIRE
  L+H* LH% H* LL%

The way that CCG derivation is made sensitive to the presence of tones is as follows (adapted from Steedman 1999 and *CSEI*).

The presence of an accent on a word infects its whole category with themehood or rhemehood, via the feature-value θ/ρ from table 6.1. (The other common-ground feature-value from table 6.1, $\top/\bot$, is suppressed in the present presentation in the interests of simplification. since all accents are H accents with the same value $\top$.)

For example the transitive verb MARRIED bearing an H* accent has the following category, in contrast to unaccented *married*:[7]

(15) $\underset{\text{H*}}{\text{MARRIED}} := (S_{\top,\theta}\backslash NP_{\top,\theta})/NP_{\top,\theta} : \lambda x.\lambda y.{*}married\,xy$

The feature θ ensures that a verb so marked can only combine with arguments that are compatible with theme marking—that is, which do not bear the rheme marking feature value ρ—and marks its result as theme marked as well. The element in the logical form correponding to the accented word itself is marked for contrast with the asterisk operator which in *CSEI* is formalized in terms of Rooth's Alternative Semantics.

Boundaries, by contrast are not properties of words or phrases, but independent string elements in their own right. They bear a category which, by mechanisms parallel to those discussed in more detail in *SP*, "freezes" θ/ρ-marked constituents as complete intonational phrases ϕ, making them unable to combine further with anything except similarly complete prosodic units.

For example, the hearer-responsibility-signalling LH% boundary bears the following category:

7. Thus, accent behaves like a morpheme at the level of the lexicon, bringing intonation structure and information structure under Chomsky's 1995b:228 Inclusiveness Condition, which says that all language- and content-specific information must be projected from the lexicon by language- and content–independent rules. (Number agreement is suppressed in the example in the interests of reducing formal clutter.)

(16) LH% := $S\$_\phi\backslash S\$_\eta : \lambda f.\eta\, f H)$

—where *S*\$ is as usual a variable ranging over *S* and syntactic function categories into S, η is a variable ranging over syntactic features θ/ρ which are also constants of the logical language defined in *CSEI* in terms of the Rooth (1992) alternative semantics, and ϕ marks the result syntactically as a complete phonological phrase.

The derivation of (14a) then appears as follows:

(17)
$$
\begin{array}{ccccc}
\text{Manny} & \text{MARRIED} & & \text{A MILLIONAIRE} & . \\
 & \text{L+H}* & \text{LH\%} & \text{H}* & \text{LL\%} \\
\cline{1-1}\cline{2-2}\cline{3-3}\cline{4-4}\cline{5-5}
S/(S\backslash NP)\ {>}\mathbf{T} & (S_\theta\backslash NP_\theta)/NP_\theta : & S\$_\phi\backslash S\$_\eta : & S_\rho\backslash(S_\rho/NP_\rho)\ {<}\mathbf{T} & S\$_\phi\backslash S\$_\eta : \\
: \lambda p.p\,manny & \lambda x.\lambda y.*married\,x\,y & : \lambda f.\eta\, f H & : \lambda p.p\,(a*millionaire) & : \lambda f.\eta\, f S \\
\cline{1-2}
\multicolumn{2}{c}{{>}\mathbf{B}\quad S_\theta/NP_\theta} & & & \\
\multicolumn{2}{c}{: \lambda x.*married\,x\,manny} & & & \\
\cline{1-3}\cline{4-5}
\multicolumn{3}{c}{{<}\quad S_\phi/NP_\phi} & \multicolumn{2}{c}{{<}\quad S_\phi\backslash(S_\phi/NP_\phi)} \\
\multicolumn{3}{c}{: \theta(\lambda x.*married\,x\,manny)\,H} & \multicolumn{2}{c}{: \rho(\lambda p.p*manny)\,S} \\
\cline{1-5}
\multicolumn{5}{c}{{<}\quad S_\phi : (\rho(\lambda p.p\,(a*millionaire)\,S))(\theta\,(\lambda x.*married\,x\,manny)\,H)} \\
\multicolumn{5}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\
\multicolumn{5}{c}{S : married\,(a\,millionaire)\,manny}
\end{array}
$$

"You suppose to be common ground *who Manny married* (as opposed to the alternatives); I make common ground that it was a *millionaire* (as opposed to the alternatives)"

In the last step of the derivation, the markers of speaker/hearer agency, positive/negative polarity, and theme/rheme are evaluated with respect to the database, to check that the associated presuppositions hold or can be accomodated. In the latter case this includes support or accommodation for the relevant alternative sets, and will include updates corresponding to the new theme and rheme. If any of these presuppositions fails, then processing will block and incomprehension will result. If it succeeds, then the two core λ-terms can β-reduce to give the canonical proposition as the result of the derivation. As noted earlier, this process (in particular, the semantics of the contrast marker * at the level of logical form) is formalized in *SCEI* using Rooth's Alternative Semantics and Stalnaker's (1978) notion of *common ground*, as informally translated below the derivations.

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

The contrastive theme tune L+H* LH% frequently appears on the first information unit in the sentence, including topicalized or fronted themes like that in (17) of chapter 9, *This movie, I like*. However, due to its otherwise rather rigid word-order, English, unlike most other languages, allows this tune to be applied anywhere in the sentence that the context allows, a phenomenon we refer to by analogy with the *in situ wh*-elements discussed in the same chapter as "*in situ* topicalization". For example, the following is a possible response to the discourse "I know Manny dated a dentist, but did he also date a millionaire?":

(18)

$$
\begin{array}{ccccc}
\text{Manny} & \text{MARRIED} & & \text{a MILLIONAIRE} & \\
 & \text{H}* & \text{L} & \text{L+H}* & \text{LH\%} \\
\hline
S/(S\backslash NP) \;{>}\mathbf{T} & (S_\rho\backslash NP_\rho)/NP_\rho : & S\$_\phi\backslash S\$_\eta : & S_\theta\backslash(S_\theta/NP_\theta)\;{<}\mathbf{T} & S\$_\phi\backslash S\$_\eta : \\
: \lambda p.p\, manny & \lambda x.\lambda y.{*}married\, x\, y & : \lambda f.\eta f\, S & : \lambda p.p\,(a\, {*}millionaire) & : \lambda f.\eta f\, H \\
\hline
\multicolumn{2}{c}{S_\rho/NP_\rho \quad {>}\mathbf{B}} & & & \\
\multicolumn{2}{c}{: \lambda x.{*}married\, x\, manny} & & & \\
\hline
\multicolumn{3}{c}{S_\phi/NP_\phi \quad <} & \multicolumn{2}{c}{S_\phi\backslash(S_\phi/NP_\phi) \quad <} \\
\multicolumn{3}{c}{: \rho(\lambda x.{*}married\, x\, manny)\, S} & \multicolumn{2}{c}{: \theta\,(\lambda p.p\,(a\, {*}millionaire))\, H} \\
\hline
\multicolumn{5}{c}{S_\phi : (\theta\,(\lambda p.p\,(a\, {*}millionaire))\, H)(\rho\,(\lambda x.{*}married\, x\, manny)\, H) \quad <} \\
\multicolumn{5}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\
\multicolumn{5}{c}{S : married\,(a\, millionaire)\, manny}
\end{array}
$$

"You suppose to be common ground *a millionaire* (as opposed to the alternatives; I make common ground that Manny *married* one (as opposed to the alternatives)"

Such *in situ* themes may also be noncontrastive, hence without accent (unaccented "a millionaire" would be equally appropriate in (18)). The earlier child-directed example (11b)/(12) can therefore be seen as a similar *in-situ* topicalization with a non-contrastive topic or theme "the doggy". Such sentences have derivations like the following, in which we follow *CSEI* in asssuming that unaccented accusatives in English are intrinsically theme-marked:

(19)

$$
\begin{array}{ccccc}
\text{You} & \text{LIKE} & & \text{the doggy} & \\
 & \text{H}* & \text{L} & & \text{LL\%} \\
\hline
S/(S\backslash NP) & (S_\rho\backslash NP_\rho)/NP_\rho & S\$_\phi\backslash S\$_\eta & NP^{\uparrow}_\theta & S\$_\phi\backslash S\$_\eta \\
: \lambda p.p\, you & : \lambda x\lambda y.{*}like\, x\, y & : \lambda f.\eta f\, S & : \lambda p.p\,(the\, dog) & : \lambda f.\eta f\, S \\
\hline
\multicolumn{2}{c}{S_\rho/NP_\rho \quad {>}\mathbf{B}} & & & \\
\multicolumn{2}{c}{: \lambda x.{*}like\, x\, you} & & & \\
\hline
\multicolumn{3}{c}{S_\phi/NP_\phi \quad <} & \multicolumn{2}{c}{NP^{\uparrow}_\phi \quad <} \\
\multicolumn{3}{c}{: \rho(\lambda x.{*}like\, x\, you)\, S} & \multicolumn{2}{c}{: \theta(\lambda p.p\,(the\, dog))\, S} \\
\hline
\multicolumn{5}{c}{S_\phi : (\theta(\lambda p.p\,(the\, dog))\, S)(\rho(\lambda x.{*}like\, x\, you)\, S) \quad <} \\
\multicolumn{5}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\
\multicolumn{5}{c}{S : like\,(the\, dog)\, you}
\end{array}
$$

"I suppose to be common ground *the dog*. I make it common ground that *you like* it (as opposed to the alternatives).

Even before the child has to deal with object *wh*-questions and relative clauses in child directed speech, it seems that it is learning about constituents like *you like* of the type S/NP that it will need in order to take these constructions in its stride when they are eventually encountered.

According to the present theory, prodosic phrases are simply surface syntactic constituents in the generalized sense of the term that is entailed by the CCG theory of grammar. Besides the standard linguist's phrasal inventory, they also include exacly the same non-standardd constituents that are seen as the residues of both moved and *in situ wh-*, and right-node-raising constructions. In particular, all three phenomena are unbounded, in the sense of being able to elide tensed clause boundaries.

This identity of domain betwen *wh* and intonation is reminiscent of the Contiguity Theoretic claim of Richards (2010, 2016) that the domain of *wh*-dependency is the prosodic phrase. The difference is that according to Richards, *in situ wh* is possible just in case the material between the wh-element and the complement forms a single phonological phrase. It is only when that material does *not* form a phonological phrase that *wh* has to move to the complement. The difference is that in CCG, the domain of *wh* is *always* a phonological phrase: the only difference between an *in situ wh* and a complement *wh* is *which side* of the phonological phrase they are contiguous *to*.[8]

6.6 Discussion

A word of caution is needed here concerning the data under discussion. Though it is common to talk in terms of English prosody in terms of tones, the realization of accent is as noted earlier extremely variable, and for some speakers tones may not involve pitch as such at all (Calhoun, 2010).

No-one in fact knows in acoustic terms exactly what invariant it is that we perceive as accent. It is possibly best to think of it as a kind of perceived vocal effort or "oomph," that is invariant across whispering and singing, as well as across speech-styles and dialects, and which may show up as a pitch maximum or minimum, or as length, or as alignment, or as increased amplitude, or as any combination of the above.

There is equally widespread confusion concerning the semantics of information structure that is anatomized here. The theme-rheme distinction assumed here is sometimes referred to as between topic and comment. The distinction between background and contrast is, as has been noted, sometimes referred to as that between given and new. Although these two dimensions are here assumed to be independent, they are unfortunately used in ways that over-

8. Since it is far from obvious that the residue of relativization in English is *not* a phonological phrase, Richards, 2016:2-3 is careful to distinguish a "rough draft" level of phonological form corresponding to Selkirk's Match Theory (see below) as the level relevant to determining movement. This rough draft may be overwritten at later stages of the derivation.

lap in the literature, often with the word "focus" substituted for "contrast" or "rheme", or both.

Within these limits, the intonational theories of Selkirk (1984), Pierrehumbert and Hirschberg (1990), and Ladd (2008), are entirely consistent with the present account of their relation to syntax and semantics. Indeed, the "edge-based" syntax-prosody interface outlined in Selkirk (1986, 1995) could be seen as attempting to derive very much the same generalized notion of consituency as CCG from a more traditional notion of syntactic structure, by introducing extra edge-related structural brackets. More recent "Match Theoretic" accounts by Kratzer and Selkirk (2007) and Selkirk (2011), Adger (2007), and Wagner (2005, 2010) are similarly related to the "phasal" version of minimalism outlined in Chomsky (2001), and (together with much older work such as that of Bresnan, 1971) can be seen as consistent with the present association of accent with lexical items. In particular, Wagner, 2005 can be seen as translating the CCG notion of constituency into Minimalist terms by the use of movement, rather than combinatory rules.

The main difference between the present account on the one hand, and these phonological theories and the related alternative-semantic accounts (Rooth, 1985, 1992; Büring, 1997b; Büring, 2016a; Truckenbrodt, 1999, 2012; Szendrői, 2001, 2004; Szendrői, 2017, and Erteschik-Shir, 2007) is that the latter all involve the idea that Focus-marking originates on a lexical item—usually, as in the present account, as a result of accent—and is then projected by an autonomous process of "F-projection" onto larger extents of the sentence, independently of syntactic structure or derivation. Since these larger extents do not, in the terms of their theories, coincide with their notion of syntactic constituency or the attendant compositional semantics, considerable complications ensue. For example, some versions engender "over-focusing", or the loss of the semantic distinction between "Fred ATE the beans" and "#Fred ate the BEANS" as answers to the question "What did Fred do with the beans?", a problem that led Schwarzschild (1999) to introduce a transderivational constraint AVOIDF to eliminate the latter in favor of the former.[9]

In the present theory, contrast is also defined in terms of alternative semantics. Lexical focus—that is, theme/rheme marking—is also projected onto larger phonologically-bounded structures. However, because CCG completely equates intonation structure with syntactic derivational structure, and distinguishes contrastive elements in the semantics, the equivalent of focus-

9. Wagner (2012c,b,a) further proposes to eschew the theme/rheme distinction as being epiphenomenal upon the relative scope of multiple "focus" operators.

projection is achieved directly by surface syntactic derivation and compositional semantics, avoiding over-focusing. That is, "Fred ATE the beans" and "Fred ate the BEANS" differ semantically, as to whether their logical forms involve alternatives to *eating* or to *beans*.[10]

The system proposed here reduces the literal meaning of the accents and boundaries to just three semantically-grounded binary oppositions. Crucially, these dimensions grammaticalize the speaker's *claims* concerning the beliefs of the participants, which may differ from those that they actually *are* committed to, including their Mutual Beliefs. It is therefore possible for the speaker to claim that either they or the hearer are or are not committed to a proposition, whether or not that is true. This is a move in the present theory that is forced by the indirect speech-acts that are implicit in examples like (3)-(10).

The theory places a correspondingly greater emphasis on the role of speaker-presupposition (and its dual, hearer-accommodation), and on inference and implicature. To that extent, the present theory follows the tradition of Halliday (1967a) and Brown (1983), in claiming that it is the speaker who, within the constraints imposed by the context and the participants' actual beliefs and intentions, determines what is theme and rheme, and what contrasts they embody, rather than the text, and that such discourse-semantic notions as "topic continuation," "other-directedness," "floor-claiming," "turn-yielding," "discourse-structuring," "evaluation with respect to subsequent material," "politeness," "face," "deixis," "commitment," "uncertainty," "affect", and "ownership" arise indirectly, as conversational implicatures arising from the hearer's fundamental need to maintain consistency in their beliefs.

Exercise : Ward and Hirschberg (1985) use the following dialog as epitomizing the use of the L*+H LH% intonation contour to convey disagreement:
H:Harry's such a klutz.

S: He's a good	BADMINTON	player.
	L*+H	LH%

Analyse this exchange in terms of the semantics of the present chapter. If you think it involves implicature, say how that works in this case.

10. Büring (2016b) offers a solution to the over-focusing problem in terms of "Unalternative Semantics", a constraint-based version of Alternative Semantics, and claims that it eliminates Focus projection. However, Unalternative Semantics includes rules of PROPAGATION and FOCUS RETRIEVAL that seem very similar to rules of focus projection.

Part II

The Basic Constructions

Rather than simply listing a semantic rule for each syntactic rule, we will develop a rule translation procedure which forms part of a mapping from grammar rules to the interpreted trees admitted by the rules. The procedure makes central use of the semantic types which are associated with constituents in the tree, and for this reason we have called it 'type-driven'. The motivation for our approach is traditional: it allows redundancy to be eliminated, and linguistically significant generalizations to be expressed.
—(Klein and Sag, 1985)

Chapter 7

The Lexicalized Constructions

There's a one-eyed yellow idol to the north of Khatmandu,
There's a little marble cross below the town;
There's a broken-hearted woman tends the grave of Mad Carew,
And the Yellow God forever gazes down.
—*The Green Eye of the Yellow God* J. Milton Hayes, 1911

The bounded constructions define relations between the arguments of a single governer, such as a verb. They are what Goldberg (2019) rfers to as the "argument structure constructions", and include passive, raising, and control, among many others.

The fundamental assumption of the chapter will be that the bounded constructions wear their hearts on their sleeves, via their syntactic category, and that notions like A-movement, Head movement, Small Clause, Exceptional Case-Marking, and the like can be entirely excluded from the theory of syntax. To the extent that such notions capture true and significant generalizations about language, they are to be seen as phenomenological generalizations concerning possible (morpho-)lexical logical forms.

7.1 Subject-Auxiliary Inversion

In English, unlike French and German, only auxiliary verbs invert with the subject:

(1) a. Does he bite?
    b. *Bites he?

However, the fact of subject auxiliary agreement, together with the semantic resemblance of the auxiliaries to the raising verbs considered below suggests the following VSX inverting category:[1]

1. The existence of coordinate sentences like *Does he bite and she kick?*, can be attributed to argument cluster coordination, discussed in chapter 3.2 and chapter (11) below. is irrelevant to this question. NOTE TO ME Reexamine this decision in the light of the analysis of Germanic.

(2) will := $(S_{inv}/VP_{inf})/NP_{3sg} : \lambda y \lambda p.[Q]will\,(p\,y)$

([Q] is a placeholder modality for question force.)

For example:

(3)
$$\begin{array}{ccc}
\text{Will} & \text{he} & \text{bite?} \\
\hline
(S_{inv}/VP_{inf})/NP_{3sg} & NP^{\uparrow}_{3sg} & VP^{\uparrow}_{inf} : \lambda p.p\,bite \\
: \lambda y \lambda p.[Q]will\,(p\,y) & : \lambda p.p\,him & : \lambda p.p\,bite \\
\hline
\multicolumn{2}{c}{S_{inv}/VP_{inf}} & < \\
\multicolumn{2}{c}{: \lambda p.[Q]will\,(p\,him)} & \\
\hline
\multicolumn{3}{c}{S : [Q]will\,(bite\,him)} <
\end{array}$$

The unaccented *do*-support auxiliary *do* is more restricted, according to the following patternl:

(4) a. Does he bite?
b. *He does bite.

we need the following :[2]

(5) does := $(S_{inv}/VP_{inf,-cop})/NP_{3sg} : \lambda y \lambda p.[Q]p\,y$

The verb categories discussed above can be seen as lexicalizing the head-movement analysis.

The restriction of inversion to the auxiliary verb is English and the idiosyncrasies of *do*-support are clearly language-specific. In French, *all* tensed verbs invert:

(6) Mord -t-il ?
bites he ?
"Does he bite?"

So we have both of the following:

(7) a. mord := $S\backslash NP_{3s} : \lambda y.bites\,y$
b. mord := $S_{inv}/NP_{3s} : \lambda y.[Q]bites\,y$

This difference between French and English has further consequences for the systems of negation and quantifier flotation, as noted by Pollock (1989), considered next.

---

2. Stressed *does* has a distinct auxiliary category:

(i) DOES := $(S\backslash NP_{3s}/VP_{inf} : \lambda p \lambda y.p\,y$
H*

The semantics of accent, omitted here, is discussed in chapter 6.

7.2 *Neg*-placement

The English sentential negation particle *not* has a category rather like certain VP adverbials, except it is limited in Englishto non-finite VP by the category below:

(8) often := $VP_x/VP_x : \lambda p\lambda y.frequent\,(p\,x)$
$(S\backslash NP)/(S\backslash NP) : \lambda p\lambda y.frequent\,(p\,x)$
not := $VP_x/VP_x : \lambda p\lambda y.\neg(p\,x)$

(9)
$$
\begin{array}{ccccc}
\text{I} & \text{have} & \text{not} & \text{often} & \text{slept} \\
\hline
NP^{\uparrow} & (S\backslash NP)/VPen & VPx/VPx & VPx/VPx & VPen \\
: \lambda p.p\,me & : \lambda p\lambda y.perf\,(p\,y) & : \lambda p\lambda y.\neg(p\,x) & : \lambda p\lambda y.frequent\,(p\,y) & : \lambda y.sleep\,y \\
 & & & \multicolumn{2}{c}{\overline{VP_{en} : \lambda y.frequent\,(sleep y)}^{>}} \\
 & & \multicolumn{3}{c}{\overline{VPen : \lambda y.\neg(frequent\,(sleep\,y))}^{>}} \\
 & \multicolumn{4}{c}{\overline{S\backslash NP : \lambda y.perf\,(\neg(frequent\,(sleep\,y)))}^{>}} \\
\multicolumn{5}{c}{\overline{S : perf\,(\neg(frequent\,(sleep\,me)))}^{>}}
\end{array}
$$

The category of *not* introduces the negation operator at the equivalent of a specifier position for its “functional projection”. However, there is no fixed position for the operator. The same categories also allow the following with a different scope for negation:

(10) a. I have often not slept.
b. $S : perf\,(frequent\,\neg((sleep\,me)))$

Although the adverbial *often* also has the category $(S\backslash NP)/(S\backslash NP)$, allowing (10a), *not* does not, so (11b) is disallowed:

(11) a. I often have not slept.
b. *I not have often slept.

Clearly, as in the case of subject inversion, another language could assign similar categories in a different way. As noted by Kayne (1975) and Pollock (1989), and much subsequent work, French is again such a language:

(12) a. J(e n)’ai pas souvent dormi.
b. *Je souvent n’ai pas dormi.

To a first approximation, we can capture the French system by assuming that the optional negating element *ne* has the following semantically vacuous category:

(13) ne := $(S\backslash NP)/(S_{neg}\backslash NP) : \lambda p.p$

—while the negating element *pas* is

(14) pas := $(S_{neg}\backslash NP)\backslash(S\backslash NP) : \lambda p\lambda y.\neg p\, y$

—and adverbials like *souvent* are VP/VP.

Other French negating elements like *point* and *jamais* have categories like *pas*, while floating quantifiers like *tous* resemble *souvent*. Zanuttini (1997) and Haegeman (1995) show that there are further subtle ordering constraints on these elements which we will pass over here, but which could be captured in present terms by a finer set of feature-values.

7.3 The Passive

The English passive construction exploits these degrees of freedom in another way, promoting a non-maximally lf-commanding patient argument to syntactic "subject" position. The necessary categories can be thought of to a first semantic appoximation as the following:[3]

(15) a. see+en := $VP_{pass} : \lambda y.see\, y\, (something\, y)$
b. persuade+en := $VP_{pass}/VP_{to} : \lambda p\lambda y.persuade(p\, y)\, y\, (something\, y)$

We defer further discussion of the semantics of the passive until the discussion of quantifiers in chapter 13, except to note that the term $(something\, y)$ in the above logical forms is a placeholder for some subtler term capturing the fact that sentences like the following entail (a) that every boy y was persuaded by a possibly *different* agent, and (b) that the unmentioned agents are of a kind than can be *expected* to persuade boys to take baths (Fodor and Fodor, 1980).

(16) Every boy was persuaded to take a bath.

We therefore assume that the long passive involves a separate category, subcategorizing for the agentive PP_{by}-phrase, noting that in this case there is a further possible reading where the *same* person persuaded every boy.

(17) Every boy was persuaded to take a bath by somebody.

The categories are introduced to the lexicon by the following morphological derivations:

3. We write stem+affix to indicate the result of morpholexical combination. The logical form of (15b) anticipates the analysis of control verbs to be developed in section **??**.

(18)
$$\cfrac{\cfrac{\text{see}}{\begin{array}{c} VP_{inf,tel}/NP \\ : \lambda x \lambda y.achievement\,(see\,x\,y) \end{array}} \quad \cfrac{+\text{en}}{\begin{array}{c} VP_{pass}\$\backslash\!\backslash VP_{inf,tel}\$/NP \\ : \lambda p \lambda y \ldots .p \ldots y\,(something\,y) \end{array}}}{VP_{pass} : \lambda y.achievement\,(see\,y\,(something\,y))}{<}LEX$$

(19)
$$\cfrac{\cfrac{\text{persuade}}{\begin{array}{c} (VP_{inf,tel})/VP_{to}/NP \\ : \lambda x \lambda p \lambda y.accomplishment\,(persuade\,(p\,x)\,x\,y) \end{array}} \quad \cfrac{+\text{en}}{\begin{array}{c} VP_{pass}\$\backslash\!\backslash VP_{inf,tel}\$/NP \\ : \lambda p \lambda y \ldots .p \ldots y\,(something\,y) \end{array}}}{VP_{pass}/VP_{to} : \lambda p \lambda y.accomplishment(persuade\,(p\,y)\,y\,(something\,y))}{<}LEX$$

Unlike the tense morpheme in (2), which applies to stems of any aspectual type *asp*, the passive morpheme is restricted to telic transitives via the feature value *tel*. It follows that examples like the following are excluded (Lakoff 1970a:19):

(20) a. #John is resembled (by his dog).
b. #A ton is weighed (by that typewriter).
c. #The finest beaches in Europe are boasted (by Skegness).

For the same reason, passivization is disallowed with atelic *promise* and *want*, despite their syntactic type-similarity in other respects to telic *persuade*:[4]

(21) a. #John was promised to leave (by the Dean).
b. #John was wanted to leave (by the Dean).

A crucial ingredient of the telicity of passivizable verbs seems to be an effect of *change* by the agent on the patient. Such effects are lacking in all of (21) and (20), and there is a complex interaction cross-linguistically between related causative and passive forms (cf. Comrie 1989).

7.4 Raising Etc.

Passive participial phrases are just one species of predicative phrase taken as complement by the copula, "be". More generally we can assume that words like "open", which are most basically adjectives, as in (22a), are also assigned the category of a predicative VP (22b):

4. But see work in the WRAP tradition stemming from Bach (1979, 1980), which assigns them a different syntactic type. It has frequently been noticed that *promise* with predicative VP complements including passives *does* passivize:
(i) i. I promised John to be allowed to leave.
ii. John was promised to be allowed to leave (by me).
However, such examples are semantically either object control or arbitrary control, and seem to arise from a distinct non-subject control lexical stem for *promise*.

(22) a. open := $N/N : \lambda n\lambda y.n\,y \wedge open\,y$
b. $AP_{adj} : \lambda y.open\,y$

The copula can then be written as follows, where $XP_{pred,agr}$ is a supertype of VP_{adj}, VP_{pass}, VP_{nom}, PA, PP, NP, etc:

(23) is := $(S\backslash NP)/VP_{pred} : \lambda p\lambda y.p\,y$

For example:

(24)
$$
\dfrac{\dfrac{\text{Every door}}{NP^{\uparrow}_{3sg} : \lambda p.\forall x[door\,x \rightarrow p\,x]} \quad \dfrac{\dfrac{\text{is}}{(S\backslash NP_{3sg})/XP_{pred,3sg} : \lambda p\lambda y.p\,y} \quad \dfrac{\text{open}}{AP^{\uparrow}_{adj} : \lambda y.open\,y}}{S\backslash NP_{3sg} : \lambda y.open\,y}>}{S : \forall x[door\,x \rightarrow open\,x]}>
$$

The copular construction is completely productive, and applies to propositional subjects, although there are of course very strong type constraints between predicates and the copular subject:

(25) a. Being green isn't easy.
b. To err is human.
c. That they won is unfortunate.

For the predicates that take propositional subjects, there is a second construction, which we will follow Chomsky (1986b) and McCloskey (1991) in involving an anaphoric relation such as extraposition or even dislocation between the proposition and a pronoun referring to an abstract situational object (Asher, 1993), rather than an expletive parallel to existential "there", considered in the next section. This category gives rise to the following paraphrases of (25):

(26) a. It isn't easy being green.
b. It is human to err.
c. It is unfortunate that they won.

7.4.1 Infinitival complementation

The varieties of discontinuity between predicates and their arguments that are bounded by the domain of a single verbal head like the following have traditionally been regarded in generative approaches as falling into two groups (Radford, 2004:268-274), exemplified by the following:

(27) a. *John* is/seems to be *nice/upstairs/sleeping*.
b. *John* hopes/tried/persuaded Mary *to win/to be upstairs/to be nice*.

The first, exemplified by (27a), consists of the "raising" constructions including the copula, which most transformational theories have described in terms of movement of subjects like *John* from predicates like *nice* to the subject position of tensed heads like *seems*. The second, exemplified by (27b), consists of the "(obligatory) control" constructions, which most have viewed as arising from an anaphoric PRO subject of non-finite predicates like "(to) go" (Chomsky, 1981; Chierchia, 1984; Landau, 2001, 2015), obligatorily bound to matrix arguments such as *John* by a variety of mechanisms. Others, including Postal (1974), Lasnik (2001), Hornstein (1999b, 2001), Boeckx, Hornstein, and Nunes (2010), and Johnson (2020), have attributed the relation, like raising, to movement. (In particular, Nunes (2004) attributes adjunct control to "sideward" movement.)

Both constructions can be analyzed by extending the set of lexical types considered so far to include certain second-order functions taking functions—more exactly, VP predicates or properties of semantic type $e \to t$ (Chierchia, 1984)—as their arguments. Here we will only consider raising in any depth, as having been attributed by all to (A-)movement.

In this connection, it will be important to recall Carlson's (1977a) distinction between "stage-level" predicates, which denote "fluents" or transient properties like *upstairs*, which are bounded in temporal extent, and "individual-level" predicates, which denote intrinsic properties with unspecified temporal extent, like *good*:[5]

While in English the stage/individual-level distinction is not marked in morpho-syntax, in languages such as West Greenlandic, it is morphologically marked (van Geenhoven, 1998), while in Spanish it is specified by different copular forms "ser" and "estar", and in Scots Gaelic (to which we return below) it is reflected structurally in two distinct subject positions for the copula, as well as in the distinction between copular and "substantive" forms "is" and "tha" (Gillies, 1993:208-211; Ramchand, 1996).[6]

We accordingly assume that the stage/individual distinction is evident in the syntactic type of all forms of infinitival *VP*, (although such details will be suppressed wherever they would merely be distracting).

5. Like other aspectual distinctions, this one is labile: the predicates usually found to be stage-level can in contexts requiring individual level predicates be "coerced" to the latter type, and vice versa.

6. Kratzer (1988/1995) grounds the stage/individual distinction in the semantics, arguing that stage-level predicates include a spatio-temporally locative Davidsonian lf event-variable, which individual-level predicates lack. The present paper obscures this distinction in the representations of logical forms, to simplify.

In English, the distinction is evident in the predicates that can occur as bare non-finite complements to "seem", which at least in some dialects are restricted to individual-level (intrinsic) predicates, unlike those of "seem to be", which are unrestricted:

(28) a. Seymour seems nice/*upstairs/*speaking.
b. Seymour seems to be nice/upstairs/speaking.

The distinction is also manifest in the existential *there*-insertion construction, to be discussed below, which is only compatible with copular verbs and stage-level (transient) predicates:[7]

(29) a. Fairies seem/try to be nice/at the bottom of our garden
b. There are/seem to be/are believed/claimed to be fairies *nice/at the bottom of our garden.
c. *There try to be fairies nice/at the bottom of our garden.

The NP complement in the *there*-construction also has to be indefinite, a restriction which we follow Bolinger (1977); Rando and Napoli (1978); Abbott (1993), and Huddleston and Pullum (2002:1392-1403) in assuming to be essentially pragmatic in origin, related to discourse "newness" (see Prince, 1981):and examples in note 7): [8]

(30) a. There are fairies/some fairies/many fairies/no fairies at the bottom of our garden.
b. There is *the fairy Paribanou/*every fairy/*it at the bottom of our garden.

In present terms, all of these distinctions must be expressed lexically, as in the following exemplars, where the syntactic type XP_{pred} schematizes over predicative $PP, AP, NP, VP_{ing}, VP_{pss}$ (excluding VP, VP_{to}, VP_{en})—roughly, the

7. The construction is often to be found in the opening lines of Edwardian dramatic monologues:
(i) *There's a one-eyed yellow idol to the North of Kathmandu,*
There's a little marble cross below the town;... (Hayes, 1911)
(ii) *There's a breathless hush in the Close to-night;*
Ten to make and the match to win; ... (Newbolt, 1898)
(iii) *There are fairies at the bottom of our garden!*
It's not so very, very far away; ... (Fyleman, 1917)

8. In the interests of brevity, we pass over the further class of verbs like "arrive", "arise", "appear", etc. that can occur in the *there*-construction when they are predicated of indefinite and stage-level complements, and appear to bear the category of the copula:
(i) a. There appeared a tall ship on the horizon.
b. There arrived a train in the station.
c. There hung a shotgun upon the wall.

See Levin (1993:§6.1), Hale and Keyser (2002), and Deal (2009) for extensive discussion.

attributive NP modifiers $NP\backslash NP$. The feature-values stg/idv respectively denote either stage-level or individual-level predicates where, crucially, raising verbs transmit that value to their result via the feature variable *pred*:[9]

(31) *Non-raising:*

think	$:= VP_{pred_{idv}}/S : \lambda s \lambda y.think\, s\, y$
Raising-to-subject:	
be	$:= VP_{pred}/XP_{pred} : \lambda p \lambda y.p\, y$
seem	$:= VP_{pred}/VP_{to,pred} : \lambda p \lambda y.seem\,(p\, y)$
	$:= VP_{pred_{idv}}/XP_{pred_{idv}} : \lambda p \lambda y.seem\,(p\, y)$
likely	$:= XP_{pred}/VP_{to,pred} : \lambda p \lambda y.probable\,(p\, y)$
to	$:= VP_{to,pred}/VP_{pred} : \lambda p \lambda y.p\, y$
Raising-to-object:	
believe	$:= (VP_{pred}/VP_{to,pred})/NP : \lambda x \lambda p \lambda y.believe\,(p\, x)\, y$
Subject-control:	
hope	$:= VP_{pred_{idv}}/VP_{to,pred} : \lambda p \lambda y.hope\,(p\, y)\, y$
Object-control:	
persuade	$:= (VP_{pred_{idv}}/VP_{to,pred})/NP : \lambda x \lambda p \lambda y.persuade\,(p\, x)\, x\, y$
Adjunct-control:	
without	$:= (VP\backslash VP)/VP_{ing} : \lambda p \lambda q \lambda y.\neg p\, y \wedge q\, y$

It will be useful in what is to follow to note that passivization morphology (50) then has the effect of mapping agentive raising-to-object and object-control verbs respectively into raising-to-subject and subject-control verbs:

(32) *Passive of Raising-to-object:*
(be) believed $:= VP_{pss,pred}/VP_{to,pred} : \lambda p \lambda x.believe\,(p\, x)\, one$
Passive of Object-control:
(be) persuaded $:= VP_{pss,pred_{idv}}/VP_{to,pred} : \lambda p \lambda x.persuade\,(p\, x)\, x\, one$

We will assume that the stage/individual-level predicate distinction is projected morpho-lexically by tense from the stem onto S in much the same way as by raising verbs, so that for example "seems to be upstairs" is $S_{pred_{stg}}\backslash NP_{3s}$. This detail is needed for the analysis of *there*-insertion below, although we will usually suppress the distinction except where raising is involved, to reduce notional clutter.

The distinction between raising and control verbs at the level of logical form in (31) is that, in the former, the raised argument occurs only once as a subject

9. Control categories are included for completeness: in contrast to raising, the present account adds little to existing accounts of control such as Landau (2001, 2015, 2021). As usual, the use of variable binding in logical forms is non-essential: all of these lexical lfs can be captured in a variable-free combinatory calculus (Steedman, 1985/1988; Szabolcsi, 1989).

variable in the predicate-argument structure, where it is bound by its λ-binder, whereas in the latter, the bound variable occurs twice at that level, once as subject or object controller and once as controllee, a distinction parallel to that between A-movement and the PRO mechanism in the Government-Binding theory.[10]

We will consider the two constructions in turn.

7.4.2 Control

Obligatory-control verbs are *second-order* functions, in the sense that their complement VP_{to} is semantically a function, which is applied at the level of lf via a λ-bound second-order variable (p in (31)) to the subject or object (y or x) of the control verb. The λ-binding of the latter does the equivalent of binding (or moving) PRO.

Control itself is mainly of interest for present purposes in contrast to raising, since unlike the latter it is not generally regarded as arising from movement

7.4.2.1 Subject and object control : If the complement VP_{to} is itself headed by a subject-control verb, binding a lower subject, the result is a "cascade" of subjacent copies at the level of lf (Ross, 1967; Sauerland, 1998):

(33) John wants to try to begin to write a play.

$$\begin{array}{ccccccc}
\text{John} & \text{wants} & \text{to} & \text{try} & \text{to} & \text{begin} & \text{to write a play.} \\
NP^{\uparrow}_{3s} & (S_{pred_{idv}}\backslash NP_{3s})/VP_{to} & VP_{to}/VP & VP/VP_{to} & VP_{to}/VP & VP/VP_{to} & VP_{to} \\
: john & : \lambda p\lambda y.pres\,(want\,(p\,y)\,y) & : \lambda p\lambda y.p\,y & : \lambda p\lambda y.try\,(p\,y)\,y & : \lambda p\lambda y.p\,y & : \lambda p\lambda y.begin\,(p\,y)\,y & : \lambda y.write\,(a\,play)\,y
\end{array}$$

$$>\mathbf{B}: \quad VP_{to}/VP_{to} : \lambda p\lambda y.try\,(p\,y)\,y$$

$$>\mathbf{B}: \quad VP_{to}/VP_{to} : \lambda p\lambda y.begin\,(p\,y)\,y$$

$$>: \quad VP_{to} : \lambda y.begin\,(write\,(a\,play)\,y)\,y$$

$$>: \quad VP_{to} : \lambda y.try\,(begin\,(write\,(a\,play)\,y)\,y)\,y$$

$$>: \quad S_{pred_{idv}}\backslash NP_{3s} : \lambda y.pres\,(want\,(try\,(begin\,(write\,(a\,play)\,y)\,y)\,y)\,y)$$

$$>: \quad S_{pred_{idv}} : pres\,(want\,(try\,(begin\,(write\,(a\,play)\,john)\,john)\,john)\,john)$$

(A similar cascade, of *chapman* objects can be seen in the logical form for the object-conttrol sentence "Keats persuded Chapman to go" in derivation (14) of chapter 2.)

10. Landau (2015), following Williams (1994), draws a number of finer semantic distinctions among subject- and object- control verbs that are passed over here, including a distinction between "predicative" ones like "manage" and "begin", and "attitudinal" ones like "hope", "persuade", and "tell", with distinctions in factivity and obligatoriness or otherwise of "de se" readings (Lewis, 1979a; Chierchia, 1989) based on scope with respect to intensional operators at the level of logical form. The lexical logical forms shown here are compatible with such finer distinctions, but are underspecified with respect to them for the present purpose at the level of logical form. Landau, 2021: 20 accounts for adjunct control by specification, as in (31) above.

Sentence (33) can also be derived by composition merger to yield non-standard constituent structures like the following, with exactly the same lf result:[11]

(34)
$$\begin{array}{ccccccc}
\text{John} & \text{wants} & \text{to} & \text{try} & \text{to} & \text{begin} & \text{to write a play.} \\
\overline{NP^{\uparrow}_{3s}} & \overline{(S_{pred_{idv}}\backslash NP_{3s})/VP_{to}} & \overline{VP_{to}/VP} & \overline{VP/VP_{to}} & \overline{VP_{to}/VP} & \overline{VP/VP_{to}} & \overline{VP_{to}} \\
:john & :\lambda p\lambda y.pres\,(want\,(p\,y)\,y) & :\lambda p\lambda y.p\,y & :\lambda p\lambda y.try\,(p\,y)\,y & :\lambda p\lambda y.p\,y & :\lambda p\lambda y.begin\,(p\,y)\,y & :\lambda y.write\,(a\,play)\,y \\
 & & \multicolumn{2}{c}{\overline{\qquad\qquad}{>}\mathbf{B}} & \multicolumn{2}{c}{\overline{\qquad\qquad}{>}\mathbf{B}} & \\
 & & \multicolumn{2}{c}{VP_{to}/VP_{to}:\lambda p\lambda y.try\,(p\,y)\,y} & \multicolumn{2}{c}{VP_{to}/VP_{to}:\lambda p\lambda y.begin\,(p\,y)\,y} & \\
 & \multicolumn{3}{c}{\overline{(S\backslash NP)/VP_{to}:\lambda p\lambda y.pres\,(want\,(begin\,(p\,y)\,y))\,y}{>}\mathbf{B}} & & & \\
 & \multicolumn{5}{c}{\overline{(S\backslash NP)/VP_{to}:\lambda p\lambda y.pres\,(want\,(try\,(begin\,(p\,y)\,y))\,y)\,y}{>}\mathbf{B}} & \\
 & \multicolumn{6}{c}{\overline{S_{pred_{idv}}\backslash NP_{3s}:\lambda y.pres\,(want\,(try\,(begin\,(write\,(a\,play)\,y)\,y)\,y)\,y)}{>}} \\
\multicolumn{7}{c}{\overline{S_{pred_{idv}}:pres\,(want\,(try\,(begin\,(write\,(a\,play)\,john)\,john)\,john)\,john)}{>}}
\end{array}$$

The generalization here is that *sequences of control verbs can compose to yield a category with the same type* $(S_{pred_{idv}}\backslash NP_{agr})/VP_{pred}$ *as a lexical control verb*. The possibility of composing the elements of the control chain, rather than simply applying them, leaves the cascade of subjects unaffected, like everything else at the level of lf. We shall see later that the involvement of such non-standard constituents allows many continuous sub-sequences of verbs and their arguments to coordinate by constituent coordination.

7.4.2.2 Adjunct Control : Another case of obligatory control that Chomsky (1981) and Williams (1992) talk of as "adjunct control" is found in examples of VP adjunction like the following, in which *harry* is not only the agent of the adjacent VP *filed the report* but also of the non-adjacent VP *reading it*: [12]

(35)
$$\begin{array}{cccccc}
\text{Harry} & \text{filed} & \text{the report} & \text{without} & \text{telling} & \text{us.} \\
\overline{NP^{\uparrow}} & \overline{(S\backslash NP)/NP} & \overline{NP} & \overline{((S\backslash NP)\backslash(S\backslash NP))/VP_{ing}} & \overline{VP_{ing}/NP} & \overline{NP^{\uparrow}} \\
:harry & :\lambda x\lambda y.past\,(file\,x\,y) & :the\,report & :\lambda p\lambda q\lambda y.\neg(p\,y)\wedge(q\,y) & :\lambda x\lambda y.tell\,x\,y & :us \\
 & \multicolumn{2}{c}{\overline{S\backslash NP:\lambda y.past\,(file\,(the\,report)\,y)}{>}} & & \multicolumn{2}{c}{\overline{VP_{ing}:\lambda y.tell\,us\,y}{<}} \\
 & & & \multicolumn{3}{c}{\overline{(S\backslash NP)\backslash(S\backslash NP):\lambda q\lambda y.\neg(tell\,us\,y)\wedge q\,y}{>}} \\
 & \multicolumn{5}{c}{\overline{S\backslash NP:\lambda y.\neg(tell\,us\,y)\wedge past\,(file\,(the\,report)\,y)}{<}} \\
\multicolumn{6}{c}{\overline{S:\neg(tell\,us\,harry)\wedge past\,(file\,(the\,report)\,harry)}{>}}
\end{array}$$

The crucial point in the above derivation is that the subject-discontiguous adjunct *without telling us* is, like a control verb, syntactically and semantically

11. Other drivations with the same result are equally possible.
12. Further categories for *without* are needed for examples like *filed the report without me/my reading it*.

a *second-order function*—that is, it takes the function *filed the report* as argument. Specifically, its category $(S\backslash NP)\backslash(S\backslash NP) : \lambda q\lambda y.\neg(tell\, us\, y) \wedge q\, y$ defines it as taking a predicate like *filed the report* and a subject as separate arguments. When it adjoins to the former in the penultimate step of the derivation, it creates a logical form that passes the value of the latter (that is, *harry*) to *both* the displaced VP *telling us and* the main-clause VP *filed the report*, via the bound variables q and y at the level of logical form. (The adjunct category is inherited from the lexical category of the adjunct head "without" in application to the predicate "telling us".)

In the terms of the movement theory of control, the above is an instance of what Nunes, 1995:93-95 called "sideward" movement, because the logical form subjects of the two predicates are first unified, and then simultaneously instantiated by the value *harry* in the last step of the derivation. Under the copy-theory, sideward movement is unusual in that the two lower copies are not in a *c*-command relation. In fact, according to Nunes (1995:94), they are not even in construction at the time of copying, a suggestion which raises the question of where in that case the copying takes place. The answer under the present proposal, as in all types of control, is that the equivalent of copying is defined "off-line", in the two (eliminable) occurrences of the bound variable y in the second-order lexical logical form $\lambda p\lambda q\lambda y.\neg(p\, y) \wedge q\, y)$ of the word "without" that heads the adjunct in (35), whence it is monotonically projected by the derivation.

However, the present account of control is more akin to other local binding accounts familiar since from Montague Grammar, G/HPSG, and LFG, as distinct from a movement account.

7.4.3 Raising

Raising-to-subject verbs are second-order functions applying their complement predicate to their subject, modally modifying the result.

(36)
$$
\begin{array}{cccc}
\text{Seymour} & \text{seems} & \text{to be} & \text{nice.} \\
\hline
NP^{\uparrow}_{3s} & (S_{pred}\backslash NP_{3s})/VP_{to,pred} & VP_{to,pred}/XP_{pred} & AP_{pred_{idv}} \\
: seymour & : \lambda p\lambda y.pres\,(seem\,(p\,y)) & : \lambda p\lambda y.p\,y & : \lambda y.nice, y \\
\cline{3-4}
 & & \multicolumn{2}{c}{VP_{to,idv} : \lambda y.nice\, y \quad >} \\
\cline{2-4}
 & \multicolumn{3}{c}{S_{pred_{idv}}\backslash NP_{3s} : \lambda y.pres\,(seem\,(nice\, y)) \quad >} \\
\hline
\multicolumn{4}{c}{S_{pred_{idv}} : pres\,(seem\,(nice, seymour)) \quad >}
\end{array}
$$

As in the case of examples (**??**, **??**), the "raising" of the subject y past tense can be thought of as mediated by the binder λy, with alignment to the left of the tensed verb at the level of pf specified in the lexical syntactic type, as with all SVO tensed verbs. Intensional scope alternation, as in examples like the following, then depends on whether the scope of "a unicorn" is evaluated before derivation begins, (the "de re" interpretation), or after "lowering" or "reconstruction" into the intensional scope of *seem* at the level of logical form ("de dicto") (Steedman, 2012:50,111):

(37) A unicorn seems to be approaching.

Unlike control, (31), raising across unboundedly many raising verbs does not create a cascade of multiple copies of the subject, which remains in situ at lf (cf. Epstein and Seely, 2006:§2.4):

(38) Seymour is believed to be dreaming

$$\begin{array}{ccccc}
\text{Seymour} & \text{is} & \text{believed} & \text{to be} & \text{dreaming} \\
\hline
NP^{\uparrow}_{3s} & (S_{pred}\backslash NP_{3s})/XP_{pss,pred} & VP_{pss,pred}/VP_{to,pred} & VP_{to,pred}/XP_{pred} & VP_{ing,stg} \\
: seymour & : \lambda p\lambda y.pres\,(p\,y) & : \lambda p\lambda y.believe\,(p\,y)\,one & : \lambda p\lambda y.p\,y & : \lambda y.prog\,(dream\,y)
\end{array}$$

$$\xrightarrow{>}\quad VP_{to,stg} : \lambda y.prog\,(dream\,y)$$

$$\xrightarrow{>}\quad VP_{pss,stg} : \lambda y.believe\,(prog\,(dream\,y))\,one$$

$$\xrightarrow{>}\quad S_{pred_{stg}}\backslash NP_{3s} : \lambda y.pres\,(believe\,(prog\,(dream\,y))\,one)$$

$$\xrightarrow{>}\quad S_{pred_{stg}} : pres\,(believe\,(prog\,(dream\,seymour))\,one)$$

Raising verbs can also merge by successive composition, when they form a category of exactly the same type as a raising verb, as in the following alternative derivation for (38):

(39) Seymour is believed to be dreaming.

$$\begin{array}{ccccc}
\text{Seymour} & \text{is} & \text{believed} & \text{to be} & \text{dreaming.} \\
\hline
NP^{\uparrow}_{3s} & (S_{pred}\backslash NP_{3s})/XP_{pred} & VP_{pss,pred}/VP_{to,pred} & VP_{to,pred}/XP_{pred} & VP_{ing,stg} \\
: seymour & : \lambda p\lambda y.pres\,(p\,y) & : \lambda p\lambda y.believe\,(p\,y)\,one & : \lambda p\lambda y.p\,y & : \lambda y.prog\,(dream\,y)
\end{array}$$

$$\xrightarrow{>\mathbf{B}}\quad (S_{pred}\backslash NP_{3s})/VP_{to,pred} : \lambda p\lambda y.pres\,(believe\,(p\,y)\,one)$$

$$\xrightarrow{>\mathbf{B}}\quad (S_{pred}\backslash NP_{3s})/XP_{pred} : \lambda p\lambda y.pres\,(believe\,(p\,y)\,one)$$

$$\xrightarrow{>}\quad S_{pred_{stg}}\backslash NP_{3s} : \lambda y.pres\,(believe\,(prog\,(dream\,y))\,one)$$

$$\xrightarrow{>}\quad S_{pred_{stg}} : pres\,(believe\,(prog\,(dream\,seymour))\,one)$$

More specifically, raising verbs can compose to yield a category of the type of the copula, $(S_{pred}\backslash NP_{agr})/XP_{pred}$. Crucially, such composite copulæ as "is/are believed to be" above transmit the stage/individual-level type of their complement to their result, unlike the corresponding control composites such as "wants/want to try to begin".

If a non-raising verb yielding an event or an individual-level stative predicate occurs anywhere in the sequence, its effect will equally be transmitted to the result:

(40)

$$\begin{array}{lllll}
\text{Seymour} & \text{seems} & \text{to want} & \text{to be} & \text{dreaming.} \\
\overline{NP^{\uparrow}_{3s}} & \overline{(S_{pred}\backslash NP_{3s})/VP_{to,pred}} & \overline{VP_{to}/VP} & \overline{VP_{to,pred}/VP_{pred}} & \overline{VP_{ing,stg}} \\
: seymour & : \lambda p\lambda y.pres\,(seem\,(p\,y)) & : \lambda p\lambda y.want\,(p\,y)\,y & : \lambda p\lambda y.p\,y & : \lambda y.prog\,(dream\,y) \\
 & \multicolumn{2}{l}{\overline{(S_{pred_{idv}}\backslash NP_{3s})/VP_{pred} : \lambda p\lambda y.pres\,(seem\,(want\,(p\,y))\,y)}{}^{>\mathbf{B}}} & & \\
 & \multicolumn{3}{l}{\overline{(S_{pred_{idv}}\backslash NP_{3s})/VP_{pred} : \lambda p\lambda y.pres\,(seem\,(want\,(p\,y))\,y)}{}^{>\mathbf{B}}} &
\end{array}$$

The significance of this fact is that the existential *there*-insertion construction applies as noted earlier, across sequences of raising verbs ending in the copula, (Schreiber, 1978) and only to transient "stage-level" predicates:

(41) There was believed/*persuaded to be an elephant in the room/*wild.

7.4.4 *There*-insertion

In present terms, Williams (1984), the related G/HPSG feature-passing approach of Gazdar, Klein, Pullum, and Sag (1985) and Levine (2017:186), and the TAG-based Minimalist approach of Frank (2002:113, account for the existential *there*-construction by assigning the raising verbs involved an additional more specialized lexical category specifying an NP_{there} subject.

The alternative approach followed here takes advantage of the fact that serial raising verbs can compose to yield a non-standard constituent with the category of the copula, as in (39), by making "there" the head of the construction, assigning it the following lexical categories selecting the copula category as argument, and further specifying it for stage-level predication:[13]

(42) Subject *there*:

$$\text{there} := ((S/XP_{pred_{stg}})/NP_{agr})/((S_{pred_{stg}}\backslash NP_{agr})/XP_{pred_{stg}}) : \lambda c\lambda y\lambda p.c\,(p\,y) \wedge \mathbf{new}\,y$$

(43) Object *there*:

$$\text{there} := ((((S\backslash NP)/XP_{pred_{stg}})/NP)/(VP_{to,pred_{stg}}/XP_{pred_{stg}}))\backslash((S\backslash NP)/VP_{to,pred})/NP)$$
$$:\lambda b\lambda x\lambda p\lambda y.b\,(p\,x)\,y \wedge \mathbf{new}\,y$$

(The element $\mathbf{new}\,y$ in the logical forms requires indefiniteness/discourse-newness of the subject, whose semantics we pass over here.)

13. A further subject-inversion category related to (43), but looking to the left for the inverting copula $((S_{inv}/XP_{pred_{stg},agr})/NP_{agr})$ as its first argument, is also needed to support questions like:

(i) a. Are there (believed to be) fairies at the bottom of our garden?
b. Where are there (believed to be) fairies?

We pass over it here, in the interests of brevity.

(42) applies to a constituent with the type of the SVX copula to its right to yield a VSX version, with the NP complement restricted to indefinites and the predicate to stage-level predicates: For example, compare the following canonical derivation with the corresponding *there*-insertion (45):

(44)
$$\begin{array}{cccc}
\text{Fairies} & \text{are} & & \text{at the bottom of our garden} \\
\hline
NP^{\uparrow}_{3p} & (S\backslash NP_{3p})/XP_{3p} & & PP^{\uparrow}_{pred_{stg}} \\
:fairies & :\lambda p\lambda y.pres\,(p\,y) & & :\lambda y.at\,(bottom\,garden)\,y \\
\end{array}$$
$$S\backslash NP_{3p} : \lambda y.pres\,(at\,(bottom\,garden)\,y) \quad <$$
$$S : pres(at\,(bottom\,garden)\,fairies) \quad >$$

(45)
$$\begin{array}{cccc}
\text{There} & \text{are} & \text{fairies} & \text{at the bottom of our garden.} \\
\hline
((S/XP_{pred_{stg}})/NP_{agr})/((S_{pred_{stg}}\backslash NP_{agr})/XP_{pred_{stg}}) & (S_{pred}\backslash NP_{3p})/XP_{pred} & NP^{\uparrow}_{3p} & PP^{\uparrow}_{pred_{stg}} \\
:\lambda c\lambda y\lambda p.c\,(p\,y)\wedge \textbf{new}\,y & :\lambda p\lambda y.pres\,(p\,y) & :fairies & :\lambda y.at\,(bottom\,garden)\,y \\
\end{array}$$
$$(S/XP_{pred_{stg}})/NP_{3p} : \lambda y\lambda p.pres\,(p\,y)\wedge \textbf{new}\,y \quad >$$
$$S/XP_{pred_{stg}} : \lambda p.pres\,(p\,fairies)\wedge \textbf{new}\,fairies \quad <$$
$$S : \lambda p.pres\,(at\,(bottom\,garden)\,fairies)\wedge \textbf{new}\,fairies \quad <$$

Since "seem", "to", and "be" can compose to yield a category of the same stage-level predicate-selecting type as the copula, there are parallel derivations for the following, in both of which long-distance agreement is just local agreement:

(46)
$$\begin{array}{ccccc}
\text{Fairies} & \text{seem} & \text{to} & \text{be} & \text{at the bottom of our garden} \\
\hline
NP^{\uparrow}_{3p} & (S\backslash NP_{3p})/VP_{to,pred} & VP_{to,pred}/VP_{pred} & VP_{pred}/XP_{pred} & PP^{\uparrow}_{pred_{stg}} \\
:fairies & :\lambda p\lambda y.pres\,(seem\,(p\,y)) & \lambda p\lambda y.p\,y & \lambda p\lambda y.p\,y & :\lambda y.at\,(bottom\,garden)\,y \\
\end{array}$$
$$VP_{to,}red_{stg}/XP_{pred_{stg}} : \lambda p\lambda y.p\,y \quad >\mathbf{B}$$
$$(S\backslash NP_{3p})/XP_{pred_{stg}} : \lambda p\lambda y.pres\,(seem\,(p\,y)) \quad >\mathbf{B}$$
$$S\backslash NP_{3p} : \lambda y.pres\,(seem\,(at\,(bottom\,garden)\,y)) \quad <$$
$$S : pres\,(seem(at\,(bottom\,garden)\,fairies)) \quad >$$

(47)
$$\begin{array}{ccccc}
\text{There} & \text{seem} & \text{to be} & \text{fairies} & \text{at the bottom of our garden.} \\
\hline
((S/XP_{pred_{stg}})/NP_{agr})/((S_{pred_{stg}}\backslash NP_{agr})/XP_{pred_{stg}}) & (S_{pred}\backslash NP_{3p})/VP_{to,pred} & VP_{to,pred}/XP_{pred} & NP^{\uparrow}_{3p} & PP^{\uparrow}_{pred_{stg}} \\
:\lambda c\lambda y\lambda p.c\,(p\,y)\wedge \textbf{new}\,y & :\lambda p\lambda y.pres\,(seem\,(p\,y)) & :\lambda p\lambda y.p\,y & :fairies & :\lambda y.at\,(bottom\,garden)\,y \\
\end{array}$$
$$(S_{pred}\backslash NP_{3p})/XP_{pred} : \lambda p\lambda y.pres\,(seem\,(p\,y)) \quad >\mathbf{B}$$
$$(S/XP_{pred_{stg}})/NP_{3p} : \lambda y\lambda p.pres\,(seem\,(p\,y))\wedge \textbf{new}\,y \quad >$$
$$S_{pred_{stg}}/XP_{pred_{stg}} : \lambda p.pres\,(seem\,(p\,fairies))\wedge \textbf{new}\,fairies \quad <$$
$$S : pres\,(seem\,(at\,(bottom\,garden)\,fairies))\wedge \textbf{new}\,fairies \quad <$$

Clearly, the raising series of type $(S_{pred}\backslash NP_{3p})/XP_{pred}$ can include unboundedly many raising verbs. However, the *there*-inserting category (42) excludes non-copular-valued serial raising verb examples like the following:

(48) a. #There seem fairies to be at the bottom of our garden.
b. #There are believed fairies to be at the bottom of our garden

The more complex category (43) for expletive "there" similarly inverts the rightward arguments of raising-to-object verbs like "believe", to yield a category much like (42), looking for a *to*-infinitival copular category such as "to be/to be certain to be believed to be ", an indefinite NP, and a stage-level predicate. as in (49).[14]

(49)
$$\begin{array}{cccccc}
\text{I} & \text{believed} & \text{there} & \text{to be} & \text{fairies} & \text{at the bottom of our garden} \\
\hline
NP^{\uparrow} & ((S\backslash NP)/VP_{to,pred})/NP & ((((S\backslash NP)/XP_{pred_{stg}})/NP)/(VP_{to,stg}/XP_{pred_{stg}}))\backslash(((S\backslash NP)/VP_{to,pred})/NP) & VP_{to,pred}/XP_{pred} & NP^{\uparrow} & PP^{\uparrow}_{pred_{stg}}
\end{array}$$
$$\underline{\qquad\qquad\qquad\qquad}_{<}\quad ((S\backslash NP)/XP_{pred_{stg}})/NP)/(VP_{to,pred_{stg}}/XP_{pred_{stg}})$$
$$\underline{\qquad\qquad\qquad\qquad}_{>}\quad ((S\backslash NP)/XP_{pred_{stg}})/NP$$
$$\underline{\qquad\qquad\qquad\qquad}_{<}\quad (S\backslash NP)/XP_{pred_{stg}}$$
$$\underline{\qquad\qquad\qquad\qquad}_{<}\quad S\backslash NP$$
$$\underline{\qquad\qquad\qquad\qquad}_{>}\quad S$$

However, "there", (42), cannot apply to "seem" or "are believed", $(S_{pred}\backslash NP_{3p})/VP_{to,pred}$, because VP_{to} is incompatible with its specification of XP in $(S_{pred_{stg}}\backslash NP_{3p})/XP_{pred_{stg}}$ (cf. Frampton and Gutmann, 2002; Stroik, 2009).

Similarly, (50) is excluded, because neither "hope", "want", nor "were persuaded" yields stage-level predicative $S_{pred_{stg}}$ (cf. 40):

(50)
$$\begin{array}{ccccc}
*\text{There} & \text{hope/try/were persuaded} & \text{to be} & \text{fairies} & \text{at the bottom of our garden} \\
\hline
((S/XP_{pred_{stg}})/NP_{agr})/((S_{pred_{stg}}\backslash NP_{agr})/XP_{pred_{stg}}) & (S\backslash NP_{3p})/VP_{to} & VP_{to,pred}/XP_{pred} & NP^{\uparrow}_{3p} & PP^{\uparrow}_{pred_{stg}}
\end{array}$$
$$\underline{\qquad\qquad\qquad\qquad}_{>\mathbf{B}}\quad (S\backslash NP_{agr})/XP$$
$$\underline{\qquad\qquad\qquad\qquad}_{***}\quad *$$

For the same reason, *there*-insertion is tense-bounded, like raising: "super *there*-insertion" into (a) below to yield (b) is predicted to be impossible, an observation that led Chomsky (2001:13) to postulate the Phase Impenetrability Condition, as arising from an asynchronous process of "transfer" to the pf and lf "modules", making finite S or T boundaries a barrier to A-movement/LDA, obviated for $\overline{\text{A}}$ movement by the COMP escape-hatch or "edge" features: (Müller, 2010):

14. We temporarily suppress logical form and details of agreement, as analogous to earlier examples.

(51) a. Fairies [(seem to) think that bicycles are] at the bottom of our garden.
b. *There [(seem to) think that bicycles are] fairies at the bottom of our garden.

In all of the above derivations, the argument of *there* bearing the raising verb type $(S|NP_{agr_i})/XP_{pred_{stg},agr_i}$ can result from composition. For example:

(52) There — seem to be — believed — to be — fairies — at the bottom of our garden

$$\text{There}: ((S/XP_{pred_{stg}})/NP_{agr})/((S\backslash NP_{agr})/XP_{pred_{stg}}) : \lambda c\lambda y\lambda p.c\,(p\,y)$$
$$\text{seem to be}: (S_{pred}\backslash NP_{3p})/XP_{pred} : \lambda p\lambda y.pres\,(seem\,(p\,y)) \quad {>}\mathbf{B}$$
$$\text{believed}: VP_{pss,pred}/VP_{to,pred} : \lambda p\lambda y.believe\,(p\,y)\,one$$
$$\text{to be}: VP_{to,pred}/XP_{pred} : \lambda p\lambda y.p\,y \quad {>}\mathbf{B}$$
$$\text{fairies}: NP^{\uparrow}_{3p} : fairies$$
$$\text{at the bottom of our garden}: XP^{\uparrow}_{pred_{stg}} : \lambda y.at\,(bottom\,garden)\,y$$
$$\xrightarrow{>\mathbf{B}} S_{pred}\backslash NP_{3p}/VP_{to} : \lambda p\lambda y.pres\,(seem\,(believe\,(p\,y)\,one))$$
$$\xrightarrow{>\mathbf{B}} (S_{pred}\backslash NP_{3p})/XP_{pred} : \lambda p\lambda y.pres\,(seem\,(believe\,(p\,y)\,one))$$
$$\xrightarrow{>} (S/XP_{pred_{stg}})/NP_{3p} : \lambda y\lambda p.pres\,(seem\,(believe\,(p\,y)\,one))$$
$$\xrightarrow{<} S/XP_{pred_{stg}} : \lambda p.pres\,(seem\,(believe\,(p\,fairies)\,one))$$
$$\xrightarrow{<} S : pres\,(seem\,(believe\,(at\,(bottom\,garden)\,fairies)\,one))$$

The result of composing *seem to believe* with *old bicycles* does not have the category of a copula because the agreement variables agr_1 and agr_2 are not identical, despite their values being accidentally the same, so the single agreement variable agr_i of the category for *there* cannot unify to yield a cetegory that would yield the meaning of (51b), *Fairies seem to believe bicycles are at the bottom of our garden*.

(53) *There — seem — to believe — bicycles — are — fairies — at the bottom of our

$$\text{*There}: ((S/XP_{pred_{stg}})/NP_{agr})/((S_{pred}\backslash NP_{agr})/XP_{pred_{stg}}) \quad \lambda c\lambda y\lambda p.c\,(p\,y)$$
$$\text{seem}: (S_{pred}\backslash NP_{3p})/VP_{to,pred} \quad \lambda p\lambda y.pres\,(seem\,(p\,y))$$
$$\text{to believe}: VP_{to}/S \quad \lambda s\lambda y.believe\,s\,y$$
$$\text{bicycles}: NP^{\uparrow} \quad \lambda p.p\,bicycles$$
$$\text{are}: (S_{pred}/NP_{3p})/XP_{pred} \quad \lambda p\lambda y.p\,y$$
$$\text{fairies}: NP^{\uparrow} \quad \lambda p.p\,fairies$$
$$\text{at the bottom of our}: PP^{\uparrow}_{pred_{stg}} \quad \lambda y.at\,(bottom\,gar$$
$$\xrightarrow{>\mathbf{B}} (S\backslash NP_{3p})/S : \lambda s\lambda y.pres\,(seem\,(believe\,s\,y))$$
$$\xrightarrow{>\mathbf{B}} S/XP_{pred} : \lambda p.p\,bicycles$$
$$\xrightarrow{>\mathbf{B}} (S\backslash NP_{3p})/XP_{pred} : \lambda p\lambda y.pres\,(seem\,(believe\,(p\,bicycles)\,y))$$
$$\xrightarrow{***}$$

The illocutionary effect of warning in intransitivized examples like the following, seems to arise from leaving the predicative element to be discovered from context:[15]

(54) There are snakes!

The class of attributive phrasal categories that are selected as $XP_{pred_{stg}}$ by the *there* categories (**??**) is somewhat mysterious. On the one hand, there seem to be some that are excluded as predicative copular complements (cf. Williams,

15. On the occasion this example came memorably to attention, the indefinite in question was a large rattlesnake, and the implicit stage-level predicate was "right where you are going to walk".

1984:133):

(55) a. There is a convict with a red shirt.
b. #A/the convict is with a red shirt.

(56) a. There have been several accidents at the factory.
b. #Several accidents have been at the factory.

It seems likely that these examples are only allowed under a reading where *with a red shirt* is a noun modifier and the stage predicate is an implicit locative deictic such as *here*.

On the other hand, attributive relative clauses are included in $XP_{pred_{stg}}$:

(57) a. There is something that doesn't love a wall.
b. Entropy is something that doesn't love a wall

The analysis of the complements of this construction as phrasal is confirmed by the fact that they can take part in $\overline{A}$- or *wh*-movement constructions, as predicted by the analysis to be presented in section **??** below:[16]

(58) a. Something there is that doesn't love a wall.
b. On yonder hill there stands a maiden.

In summary, the grammar of bounded constructions including LDA can be captured by lexicalizing control/raising verbs and adjuncts as second-order functions over predicates, thereby reducing such relations to adjacent merger, while maintaining the synchrony between agreement, merger, and semantic transfer argued for in the introduction.[17]

7.5 Light Verbs

We will assume that a substantial subclass of constructions involving "light verbs" like transitive *take*, *make*, *have*, *do*, *obtain*, etc., as in *take a walk on the wild side*, *take a good look at my face*, *take a seat*, etc., arise from "bleached" logical forms that map entities onto their *affordances* (Gibson, 1977) or characteristic events, such as *living dangerously*, *seeing clearly*, *sitting down*, etc., such as the following:

16. Example (58a) is from Frost, 1914. (58b) is traditional.

17. Certain Minimalist constraints on movement, such as Subjacency, which forbids "super-raising" or "super-control" verbs, reappear under the present proposal as limitations on the notion "possible lexical category" under the guise of the limitation to second-order functor categories (but no higher), and call for explanation in terms of a theory of lexical logical form. We will return to this question later. However, to pursue it further here would risk distraction from the current topic of how syntactic movement of all kinds can be reduced to contiguous merger.

(59) take := $(S\backslash NP)/NP : \lambda x \lambda y.\mathit{affordance}\, x\, y$

The idea of affordance and its relation to language is explored further in appendix A.

7.6 Reflexive Binding

We assume for present purposes that English reflexive pronouns are clitic, like French *se*. The boundedness of reflexivization then arises from the fact that cliticization is an essentially morpholexical process, despite the fact that the term clitic identifies them as appearing in the orthography as if separate words.

We have the following type-raised categories for clitic "himself", in which the morphological slash $\backslash\!\backslash$ restricts its application to lexical verbs:

(60) himself := $(S\backslash NP_{3sm})\backslash\!\backslash((S\backslash NP_{3sm})/NP) : \lambda p \lambda y.p\,(\mathit{self}\, y)\, y$
$VP\backslash\!\backslash(VP/NP) : \lambda p \lambda y.p\,(\mathit{self}\, y)\, y$
etc.

Syntactically, these categories are accusative instances of type-raised cased $NP^{\uparrow}_{LEX}$.[18]

The derivation for a simple reflexive transitive clause is the following, where *self harry* evaluates to *harry*:

(61)
$$\begin{array}{ccc}
\text{Harry} & \text{sees} & \text{himself.} \\
\hline
NP^{\uparrow}_{3sm} & (S\backslash NP_{3s})/NP & (S\backslash NP_{3sm})\backslash\!\backslash((S\backslash NP_{agr})/NP) \\
: \lambda p.p\, \mathit{harry} & : \lambda x \lambda y.\mathit{sees}\, x\, y & : \lambda p \lambda y.p\,(\mathit{self}\, y)\, y \\
\hline
 & \multicolumn{2}{c}{S\backslash NP_{3sm} : \lambda y.\mathit{sees}\,(\mathit{self}\, y)\, y \quad {<}LEX} \\
\hline
\multicolumn{3}{c}{S : \mathit{sees}\,(\mathit{self}\, \mathit{harry})\, \mathit{harry} \quad {>}}
\end{array}$$

For reflexive ditransitives of the kind we saw in (**??**), we have the following:

(62)
$$\begin{array}{cccc}
\text{Mary} & \text{introduced} & \text{herself} & \text{to the audience.} \\
\hline
NP^{\uparrow}_{3sf} & ((S\backslash NP_{agr})/PP_{to})/NP & ((S\backslash NP_{3sf})/PP)\backslash\!\backslash(((S\backslash NP_{3sf})/PP)/NP) & PP^{\uparrow} \\
: \lambda p.p\, \mathit{mary} & : \lambda x \lambda w \lambda y.\mathit{introduced}\, w\, x\, y & : \lambda p \lambda w \lambda y.p\, w\,(\mathit{self}\, y)\, y & : \lambda p.p\, \mathit{audience} \\
\hline
 & \multicolumn{2}{c}{(S\backslash NP_{3sf})/NP : \lambda w \lambda y.\mathit{introduced}\, w\,(\mathit{self}\, y)\, y \quad {<}LEX} & \\
\hline
 & \multicolumn{3}{c}{S\backslash NP_{3sf} : \lambda y.\mathit{introduced}\, \mathit{audience}\,(\mathit{self}\, y)\, y \quad {<}} \\
\hline
\multicolumn{4}{c}{S : \mathit{introduced}\, \mathit{audience}\,(\mathit{self}\, \mathit{mary})\, \mathit{mary} \quad {>}}
\end{array}$$

It seems reasonable to assume that *Harry talks to himself* is also a true *se*-type reflexive arising from prior lexicalization of "talks to", as in the following

18. The analysis is similar to that of Szabolcsi (1989), which is also lexicalized.

derivation:[19]

(63)
$$\begin{array}{ccc}
\text{Harry} & \text{talks to} & \text{himself.} \\
\hline
NP^{\uparrow}_{3sm} & (S\backslash NP_{3s})/NP & ((S\backslash NP_{3sm}))\backslash\!\!\backslash((S\backslash NP_{agr})/NP) \\
: \lambda p.p\,harry & : \lambda x\lambda y.talks\,(to\,x)\,y & : \lambda p\lambda y.p\,(self\,y)\,y \\
\cline{2-3}
 & \multicolumn{2}{c}{S\backslash NP_{3sm} : \lambda y.talks\,(to\,(self\,y))\,y} \quad {<}LEX \\
\hline
\multicolumn{3}{c}{S : talks\,(to\,(self\,harry))\,harry} \quad {>}
\end{array}$$

It is noteworthy that all of the above examples support deaccented “himself”

Example (64a) can be analysed similarly to (62). However, the occurrences of “himself” in (64b-f) cannot reasonably be analysed as clitic in the same way, and must be “exempt” or logophoric anaphors referring to a protagonist or “point of view” which we construe as coinciding with Harry, of a kind discussed by Pollard and Sag (1982) and Reinhart (1987), to be discussed below:

(64) a. Harry showed himself a movie.
b. Alice showed Harry to himself.
c. Alice showed himself to Harry
d. Harry showed at least a hundred movies to himself.
e. Harry talks to and about himself.
f. Harry talks to only himself.
g. Harry praises and admires himself.

The latter pronominals seem not to be compatible with deaccenting “himself”, unlike that in (a).

The following further “subject reflexive” instance of the type-raised reflexive for the non-existent “*heself”, (65) is excluded for English because it is not a possible type raised category under the schema T|(T|X), which raises only over first arguments X:

(65) *heself := $(S/NP)/\!\!/((S\backslash NP_{3sm})/NP) : \lambda p\lambda x.p\,(self\,x)\,x$

The CCG identification of a languages case-system with type-raising over its verbal categories therefore predicts the “anaphor agreement effect” of Rizzi (1990a), rather than requiring it as a stipulative constraint, thereby capturing Condition A of Chomsky (1981).

The above account almost works for Welsh, with the parallel category to for reflexives like “ei hun” (“his self”):

(66) ei hun := $NP^{\uparrow}_{LEX} : \lambda p\lambda y.p\,(self\,y)\,y$

19. This possibility may be related to the cross-linguistically unusual possibility in English of “preposition-stranding” *wh*-extraction (see section 9.8 below).

In particular, for infinitival objects, we get derivations like the following, involving the raising *do*-support-like *gwnued* construction (Borsley et al., 2007:51), which is of a kind that has been used under movement theories to argue for underlying SVO order, at least for infinitivals (Roberts, 2005):

(67)
$$\begin{array}{cccc}
\text{Gwaeth} & \text{Gwyn} & \text{weld} & \text{ei hun}\\
\text{Do.PAST.3S} & \text{Gwyn} & \text{see.INF} & \text{3MS self}\\
\hline
(S/VP)/NP & NP^{\uparrow} & VP/NP & NP^{\uparrow}_{LEX}\\
:\lambda y\lambda p.past\,(p\,y) & :\lambda p.p\,gwyn & :\lambda x\lambda y.see\,x\,y & \lambda p\lambda y.p\,(self\,y)\,y\\
\hline
\multicolumn{2}{c}{S/VP:\lambda p.past\,(p\,qwyn) \;<} & \multicolumn{2}{c}{VP:\lambda y.see\,(self\,y)\,y \;<}\\
\hline
\multicolumn{4}{c}{S:past\,(see\,(self\,gwyn)\,gwyn) \;>}
\end{array}$$

"Gwyn saw himself" (Borsley et al., 2007:51)

However. the reflexive category (66) does not support reflexivisation for VSO finite verbs, because the subject is in the way:

(68)
$$\begin{array}{lll}
\text{Gwelodd} & \text{Gwyn} & \text{ei hun}\\
\text{see.PAST.3S} & \text{Gwyn} & \text{3MS self}
\end{array}$$

"Gwyn saw himself" (Borsley et al., 2007:51)

The only ay I can see right now to handle such reflexive is via a discourse-bound lgophoric pronoun of the kind discussed next, which we could write as follows

(69) ei hun $:= NP^{\uparrow} : \lambda p\lambda y.p\,self\,y$

Such a category could compose with the subject to make it clitic:

(70)
$$\begin{array}{ccc}
\text{Gwelodd} & \text{Gwyn} & \text{ei hun}\\
\text{see.PAST.3S} & \text{Gwyn} & \text{3MS self}\\
\hline
(S/NP)/NP_{agr} & (S/NP)\backslash((S/NP)/NP_{3sm}) & S\backslash\!\!\backslash(S/NP)\\
:\lambda y\lambda x.past\,(see\,x\,y) & :\lambda p.p\,gwyn & :\lambda p.p\,self\\
\hline
 & \multicolumn{2}{c}{S\backslash\!\!\backslash((S/NP)/NP_{agr}):\lambda p.p\,self\,gwyn \;<\mathbf{B}}\\
\hline
\multicolumn{3}{c}{S:past(see\,self\,gwyn) \;<_{LEX}}
\end{array}$$

"Gwyn saw himself"

The above derivation would correspond to a movement-free analog of "head movement". This construction has been used to argue for head movement from underlying SVO order for Welsh (cf. Borsley et al., 2007:51)

As noted earlier, the presence in English and other languages of "logophoric" reflexives that are homophonous to the reflexive, but are non-clause bound, like pronouns, is a source of confusion. Such forms are exempt from the binding conditions, and refer to the individual whose viewpoint the text presents (Jackendoff, 1972; Higgins, 1973; Zribi-Hertz, 1989; Pollard and Sag, 1992), as in:

(71) a. The fact that there is a picture of himself$_i$ hanging in the post office is believed by Mary to be disturbing Tom$_i$.
b. A fear of himself$_i$ is John$_i$'s greatest problem.
c. John saw a picture of himself.

We will assume following Pollard and Sag that cases attributed to "reconstruction" like the following in fact arise from the involvement of exempt logophoric pronouns of this kind, trather than from true reflexives.

(72) a. Which pictures of himself$_i$ did Harry$_i$ see?
b. Alice wonders which pictures of himself$_i$ Harry$_i$ saw.
c. Alice wonders who$_i$ saw which pictures of himself$_i$

Further evidence for the above account can be adduced from the fact that in French, the varieties of reflexive *himself* exemplified in (**??**) and (**??**) are differentially lexicalized, as clitic *se* and *soi-même*:

(73) a. Jean se voit.
Jean self- sees
"Jean sees himself."
b. Lesquelles photos de lui-même à-t-il vu?
which photos of himself has he seen
"Which pictures of himself did he see?"

Similar reconstruction effects are observed for relativization:

(74) a. Pictures of himself that Harry saw
b. Relatives of his that every boy adores.

7.7 Discussion

The above sections show that the bounded constructions can be lexicalized, avoiding the propensity of the movement account to overgenerate word order under minimalist assumptions (Epstein and Seely, 2006).

In all cases, what we have done is to transfer the work of production rules in a base grammar like (4) in chapter**??** and transformational rules of "*A*-movement" such as raising and reflexive binding to an expanded set of language-specific lexical types, with λ-binding having the effect of movement and binding at the level of lexical logical form, and with the universal syntactic rules of functional application projecting these relations onto the sentences of the language. The advantage of this move is that we have thereby been able to bring such relations under the domain of strict locality, and bring syntactic projection under the domain of purely type-dependent binary operations over strictly contiguous constituents.

Of course, this move immediately raises the question of what is a possible lexical category. We have already noted that one can interpret much of the movement-theoretic account as addressing exactly that question. We have already mentioned subjacency as such a constraint, allowing raising and control verbs, but excluding "super-raising" and "super-control". On the basis of work in the present theory on language acquisition by child and machine discussed in appendix B, one might conjecture that there is a fixed set of semantic types, such as those of intransitive, transitive, passivized, control, etc. verbs, and that everything else is free, so that a language is free to have subject-, object-, etc., control of the VP, whether or not that VP is semantically ergative or accusative. We will return to this question in section 12.5.

It follows from the above that notions like "A-movement", "chain", feature "checking", "probe", "goal" and "long-range agreement" are redundant. However, the way we have proposed to eliminate chains and the attendant extended projection principle (EPP) differs from that of Epstein and Seely (2006), who make a similar point about the redundancy of A-chains, but reintroduce long-range A-movement. It also differs from the related methods CHECK of Grohmann, Drury, and Castillo (2000); Manzini and Roussou (2000). The analysis presented above is essentially an extension of the categorial approach of Jacobson (1992b) and Carpenter (1997).

Apart from showing that such extra non-semantic baggage is unnecessary, the analyses presented here are not different in any important respect from analyses of the same phenomena in other frameworks, such as HPSG or LFG. As remarked in the preceding chapter, it is in the mechanisms they apply to unbounded movement that the theories differ significantly, and it is to them that we turn next.

The CCG morpholexicon described in this chapter is construction-based, in the sense that the passive and the various raising and control constructions are defined in the morpholexicon, under the principle (**??**) of radical lexicalism. The question of what is a possible construction, which the minimalist program attempts to answer in terms of significant generalizations concerning limitations on A-movement and head-movement, therefore reduces in CCG, as in G/HPSG and LFG, to the question of what is a possible morpholexical category.

The answer to this question appears to lie in the semantics, and in particular (since languages may place the mopholexical/syntactic boundary differently) in the *types* that the semantics deals in and their mapping onto directional syntactic categories. We return to this question in a later chapter.

Exercise : Take your CCG analysis of the passive from the exercise for chapter 2 and use it as a basis for an explanation for the following asymmetry:

a. Frankie persuaded Albert to take a bath/Albert was persuaded to take a bath (by Frankie)

b. Frankie promised Albert to take a bath/#Albert was promised to take a bath (by Frankie)

Consider syntactic versus semantic bases for the asymmetry. Consider evidence from other non-passivizable complements if necessary:

a. #A ton is weighed by that typewriter.

b. #The finest beaches in the country are boasted by Skegness.

c. #Keats is resembled by his dog.

Exercise : The above account assumes that English verbs have a separate lexical entry for reflexive/reciprocal verbs, despite the lack of an explicit morphological marker. Defend (or attack) this assumption, possibly on the basis of evidence from another language or languages.

Chapter 8

Are All Constructions Lexically Governed?

I can't get my winkle out
Isn't it a sin?
The more I try to get it out,
The further it goes in.
—Picking All The Big Ones Out (The Winkle Song) (trad.)

We have assumed up to this point that all constructions are lexically governed, and that the lexical heads have an obvious semantics. In the case of constructions like raising, control, passive, coordination, and the various unbounded constructions such as relativization and em tough-movement, we could get away with this (at lest as long as we ignored the semantics of tense and aspect, and assumed that the semantic content of verbs like *promised* and *persuaded* were unanalysed *promised* and *persuaded*).

However, the illusion begins to break down when we consider the semantics of content words more seriously, and in particular when we consider the apparently systematic ways in which different categories can alternate for the same verbs.

8.1 Dative alternation

We usually think of the ditransitive/prepositional forms of the verb *give* as independent lexical items that simply share a semantics *give* via different λ-bindings:

(1) a. $((S\backslash NP)/NP)/NP : \lambda w\lambda x\lambda y.give\,w\,x\,y$
b. $((S\backslash NP)/PP)/NP : \lambda x\lambda w\lambda y.give\,w\,x\,y$

However, Oehrle (1976) pointed out that the two constructions are not in fact semantically equivalent. Compare:

(2) a. Nixon gave Mailer an idea for a novel.
b. Nixon gave an idea for a novel to Mailer.

The first sentence has a meaning according to which the idea originated with Mailer, which the second only has the reading on which the idea originated with Nixon.

At this point it is tempting to try to replace the constant *give* in the logical forms of (1) with more elaborate formulæ to capture the distinction between the fomer meaning as something like *Nixon caused Mailer to begin to possess an idea* and the latter more specific meaning as something like *Nixom transfered possession of an idea from himself to Mailer*.

Many have tried to formalize such distinctions in terms of some fized set of underlying semantic relations such as *cause*, *begin*, etc., or in terms of some fixed set of thematic relations according to which varbs can be typed (Fillmore, 1968; Schank, 1972; Dowty, 1979b; Pinker, 1989; Grimshaw, 1990; Jackendoff, 1990; Van Valin, 1993; Levin, 1993; Levin and Rappaport-Hovav, 1995; Williams, 2015). The difficulty is that there seem to be rather a lot of different ways in which the relevant distinctions could be drawn (Dowty, 1991b; Goldberg, 2006). The method followed has been to seek a set of common concepts underlying a wider range of alternations, of which the following are typical.

8.2 The "Unaccusative" alternation

English "unaccusative" intransitive verbs in sentences like (3b) are so called because of their supposed derivation from transitives like (a) via a lexical rule:

(3) a. Gabbitas opened the safe.
    b. The safe opened.
    c. The safe was open.

However, intransitive predicates like (3b) might equally plausibly be assumed to be derived from predicatives like (3c), where $VP_{pred_{idv}}$ is the syntactic type of the (individual-level) predicate in "The door is open", and '*become s*' is an inchoative event that initiates a state *s*, via the following lexical rule (cf. Levin and Rappaport-Hovav, 1995:23):[1]

(4) $VP_{pred_{idv}} : \lambda y.\mathrm{pred}\, y \Rightarrow_{LEX} VP_{inch} : \lambda y.become\,(\mathrm{pred}\, y)$

This would suggest lexical entries like the following for "unaccusative" verbs

1. I make no special claim for linguistic or psychological reality of such logical forms, or for any differences between the present notation and those used in the related accounts of Levin and Rappaport-Hovav (1995); Dowty (1979b); Jackendoff (1990) and Pustejovsky (1991). The present notation is offered only as a convenient placeholder for a semantics for content words, homomorphic to a universal language of mind that we have no direct access to as linguists, and to whose nature we will return in appendix D.

(5) open $:= VP_{inch} : \lambda y.become\,(open\,y)$
break $:= VP_{inch} : \lambda y.become\,(broken\,y)$
die $:= VP_{inch} : \lambda y.become\,(dead\,y)$
cool $:= VP_{inch} : \lambda y.become\,(cold\,y)$

Thus, such verbs may not be semantically unaccusative in the strong sense of being derived from transitive objects, as passive subjects could be said to be. Nor do they implicate null indefinite agents, as do "short" passives like *Mistakes were made*. Indeed, we shall see that they essentially lack agents.

As Dowty (1979b) and Grimshaw (1990) pointed out, the ability of verbs to occur in both unaccusative and transitive propositions seems to reflect their underlying semantic character. Thus we do not see such alternation for the transitive verb *please*, as in the following:

(6) a. The story pleased the child.
b. *The child pleased.

Grimshaw suggests that the difference between (3b) and (6b) reflects the fact that propositions like (3a) seem at the level of logical form to denote a Vendlerian accomplishment paraphraseable as *Gabbitas caused the safe to become open*, defining a change of state, whereas *The story pleased the child* seems to denote a state *The story was pleasing to the child*. This claim is supported by the differential compatibility of the two propositions with the progressive, which in the terms of Moens and Steedman (1987) and Pustejovsky (1991) "coerces" telic verbs to the activity that leads to the change of state:

(7) a. Gabbitas is opening the safe.
b. ?The story is pleasing the child

Jackendoff (1990), Dowty (1991a), and Levin and Rappaport-Hovav (1995) offer more extensive lexicalist theories of the unaccusatives and related constructions, for which the present analyis is intended as no more than a placeholder.

The so-called unaccusatives should not be confused with a variety of "middle" constructions exemplified by the following:

(8) a. John pleases easily.
b. The safe opens with difficulty.
c. This dress cleans easily.

Such sentences seem to be closely paraphraseable by other constructions like the following involving transitive *please* (Lees 1960; Chomsky 1964:66; Partee 1977; Newman 2020):

(9) a. John is easy to please. ("tough movement")
b. It is easy to please John. (expletive)
c. John is easily pleased. (passive)
d. Pleasing John is easy. (nominal)

Examples (8) and (9) are related to the causatives considered next.

8.3 The Causative Alternation

By a standard lexical process (cf. Lakoff 1970a; Dowty 1979b), inchoative intransitives like *open*, (5a), could via the following further lexical rule acquire further lexical entries like (11), corresponding to a causative transitive, supporting sentences like (3a), *Gabbitas opened the safe*:

(10) $VP_{inch} : \lambda y.become\,(\mathrm{pred}\,y)$
$\Rightarrow_{LEX} VP_{ptve}/NP : \lambda x \lambda y.cause\,(act\,y)\,(become\,(\mathrm{pred}\,x))$

(11) open := $VP_{ptve}/NP : \lambda x \lambda y.cause\,(act\,y), (become\,(open\,x))$

($act\,y$ represents the agent y of the causative doing something (cf. Levin and Rappaport-Hovav 1995:83), and is a placeholder for a more explicit Davidson/Reichenbachian event semantics. VP_{ptve} denotes a telic or perfective verb-phrase, a distinction that is morphologically realized in other languages, such as Russian.)

Similarly, intransitive *break* (5b) acquires a further lexical entry corresponding to a causative transitive, supporting sentences like *Floyd broke the glass*:

(12) break :=$VP_{ptve}/NP : \lambda x \lambda y.cause\,(act\,y), (become\,(broken\,x))\,y)$

Of course, not all causatives are lexically derived from inchoatives. The causative *kill* means something like $\lambda x \lambda y.cause\,(act\,y)\,(become\,(dead\,x))$, while *die* itself does not undergoe transitive causativization via rule (10), presumably because it is not an inchoative VP_{inch}.[2]

The lexical mapping from intransitives to causative transitives is strictly one-way. The fact that English has a (different) transitive verb "break" applying to deontic entities like laws and promises does not entail the existence of a corresponding intransitive, any more than the transitive (6a).

Nor do intransitive verbs like *arrive* do not transitivize, as in **The postman*

2. Because the inchoative intransive lexicalization for $\lambda x.become\,(cold\,x)$ is "cool", the corresponding causative is also "cool." Children overgeneralize such causatives, to produce utterances like *#It colds my bottom* (Bowerman 1973b).

arrived the package. We will assume that this is because they are *already* causative and perfective, and bear categories like the following, where *one* is a Skolem term dependent on *y* of a kind to be defined in chapter 13:

(13) arrive := $VP_{ptve}/PP_{loc} : \lambda loc \lambda y.cause(become\,(at\,loc\,y))\,one$

This analysis seems to be equivalent to Levin and Rappaport-Hovav's (1995:90) distinction between *internal and external causation*. That is, intransitive *arrive* is internally causative in the sense of having an internally specified agent, whereas intransitive *break* (5b) has no agent until it is transitivized as (12).[3]

8.4 The Resultative alternation

A wide variety of transitve and intransitive verbs support a resultative construction:

(14) a. The river froze solid. (unaccusative)
b. The door opened wide. (causative)
c. The platter was licked clean. (passive)
d. They licked the platter clean. (transitive)

Oddly, the construction is possible with "objects" that are not subcategorised for by the verb:

(15) a. Dora shouted. (unergative)
b. *Dora shouted herself.
c. Dora shouted herself hoarse.

(16) a. They drank. (intransitivized)
b. *They drank the pub.
c. They drank the pub dry.

(17) a. Frank sneezed. (intransitivized)
b. *Frank sneezed the napkin.
c. Frank sneezed the napkin off the table.

3. Levin and Rappaport-Hovav identify the "agents" of internal causation (1995:174) as essentially equivalent to Dowty's 1991 "Proto-Agent" role. That is, they constitute a role that is in part dependent for its specification upon the predicate. Thus, trains can be proto-agents internally causative of *arriving*, flowers are proto-agents internally causative of *blooming*, and so on, none of which causatively transitivize. As noted before, all of these theories are only partially successful attempts to formalize a universal language-independent conceptual "language of mind" to which we have no direct access.

The result predicate is clearly a causative adjunct of some kind—most straightforwardly, in the case of the unaccusative and passive resultatives (14a,b):[4]

(18)
$$
\begin{array}{ccc}
\text{The river} & \text{froze} & \text{solid.} \\
\hline
S/(S\backslash NP) & S\backslash NP & (S\backslash NP)\backslash(S\backslash NP) \\
:\lambda p.p\,\mathit{the\,river} & :\lambda y.\mathit{become}\,(\mathit{frozen}\,y)) & :\lambda p\lambda y.\mathit{cause}\,(p\,y)\,(\mathit{solid}\,y) \\
\cline{2-3}
 & \multicolumn{2}{c}{S\backslash NP \quad <} \\
 & \multicolumn{2}{c}{:\lambda y.\mathit{cause}\,(\mathit{become}\,(\mathit{frozen}\,y))\,(\mathit{solid}\,y)} \\
\hline
\multicolumn{3}{c}{S:\mathit{cause}\,(\mathit{become}\,(\mathit{frozen}\,(\mathit{the\,river})))\,(\mathit{solid}\,(\mathit{the\,river})) \quad >}
\end{array}
$$

However, in the case of transitive resultatives like (14c), the adjunct modifies the accusative argument, rather than the VP:

(19)
$$
\begin{array}{cccc}
\text{They} & \text{licked} & \text{the platter} & \text{clean.} \\
\hline
S/(S\backslash NP) & (S\backslash NP)/NP & (S\backslash NP)\backslash((S\backslash NP)/NP) & ((S\backslash NP)\backslash((S\backslash NP)/NP))\backslash((S\backslash NP)\backslash((S\backslash NP)/NP)) \\
:\lambda p.p\,\mathit{them} & :\lambda x\lambda y.\mathit{licked}\,x\,y & :\lambda p\lambda y.p\,(\mathit{the\,platter})\,y & :\lambda p\lambda q\lambda y.\mathit{cause}\,(p\,q\,y)\,(p\,\mathit{clean}) \\
\cline{3-4}
 & & \multicolumn{2}{c}{(S\backslash NP)\backslash((S\backslash NP)/NP) \quad <} \\
 & & \multicolumn{2}{c}{:\lambda q\lambda y.\mathit{cause}\,(q\,(\mathit{the\,platter})\,y)\,(\mathit{clean}\,(\mathit{the\,platter}))} \\
\cline{2-4}
 & \multicolumn{3}{c}{:\lambda y.\mathit{cause}\,(\mathit{licked}\,(\mathit{the\,platter})\,y)\,(\mathit{clean}\,(\mathit{the\,platter})) \quad <} \\
\hline
\multicolumn{4}{c}{S:\mathit{cause}\,(\mathit{licked}\,(\mathit{the\,platter})\,\mathit{them})\,(\mathit{clean}\,(\mathit{the\,platter})) \quad >}
\end{array}
$$

(Thus, the ill-formedness of (**??**b), "*They darnk the pub" is predicted.)

The case of the paradoxical intransitivized resultatives like like (16c), the analysis is the same, except that the intransitivized verb is necessarily not predicated of the accusative:[5]

(20)
$$
\begin{array}{cccc}
\text{They} & \text{drank} & \text{the pub} & \text{dry.} \\
\hline
S/(S\backslash NP) & S\backslash NP & (S\backslash NP)\backslash((S\backslash NP)/NP) & ((S\backslash NP)\backslash(S\backslash NP))\backslash((S\backslash NP)\backslash((S\backslash NP)/NP)) \\
:\lambda p.p\,\mathit{them} & :\lambda y.\mathit{drank\,something}\,y & :\lambda p\lambda y.p\,(\mathit{the\,pub})\,y & :\lambda p\lambda q\lambda y.\mathit{cause}\,(q\,y)\,(p\,\mathit{dry}) \\
\cline{3-4}
 & & \multicolumn{2}{c}{(S\backslash NP)\backslash(S\backslash NP) \quad <} \\
 & & \multicolumn{2}{c}{:\lambda q\lambda y.\mathit{cause}\,(q\,y)\,(\mathit{dry}\,(\mathit{the\,pub}))} \\
\cline{2-4}
 & \multicolumn{3}{c}{:\lambda y.\mathit{cause}\,(\mathit{drank\,something}\,y)\,(\mathit{dry}\,(\mathit{the\,pub})) \quad <} \\
\hline
\multicolumn{4}{c}{S:\mathit{cause}\,(\mathit{drank\,something\,them})\,(\mathit{dry}\,(\mathit{the\,pub})) \quad >}
\end{array}
$$

The fact that the derivation for the transitive (19) only goes through if the adjunct applies first to the accusative to yield a complex adjunct, and is blocked if the accusative first combines with the transitive verb, is consistent with the insight of Jackendoff, 1990:228, who notes that "the fixed syntax of this con-

4. Tense is omitted to ease the notational burden on the reader.
5. The semantics shown for intransitivized "drank" is a placeholder. We shall return to the semantics of intransitivization in Chapter 13

struction suggests that *even in the transitive cases [in (14c)]*, the direct object as well as the predicate AP is actually an adjunct" (emphasis added). Under the present analysis, all arguments adjoin to the verb, by virtue of being type-raised, a fact which will later be called on in Chapter refsecn:unbounded to explain their status as islands to *wh*-extraction.

Not all transitives and intransitives support the resultative construction

(21) a. *Mary arrived her car home.
b. *Harry drew a circle perfect.
c. *Harry broke the door ruined.
d. *Harry wrote the novel finished

The exceptions seem to be telic or perfective, including the causativized unaccusatives. As in the case of the causative lexical rule, we therefore assume that the lexical entries for resultative adjuncts specify *non* telic or imperfective verbs/verb-phrases VP_{-ptve}, as in the following schemata:

(22) a. $VP_{ptve} \backslash VP_{-ptve} : \lambda p \lambda y.cause\,(p\,y)\,(\mathbf{pred}\,y)$
b. $(VP_{ptve} \backslash (VP_{-ptve}/NP)) \backslash (VP \backslash (VP/NP))$:
$\lambda p \lambda q \lambda y.cause\,(p\,q\,y)\,(p\,\mathbf{pred})$
c. $(VP_{ptve} \backslash VP_{-ptve}) \backslash (VP \backslash (VP/NP)) : \lambda p \lambda q \lambda y.cause\,(q\,y)\,(p\,\mathbf{pred})$

The resultative adjuncts should be distinguished from "depictives" which seem to be extraposed adjuncts (cf. Jackendoff, 1990:278).

(23) a. Jack ate the steak raw.
b. Jack ate the steak naked.
c. Mary arrived breathless.
d. Harry wrote the novel drunk.

8.5 Some More Idiosyncratic Constructions

Construction Grammarians such as Goldberg (1995) and Kay and Fillmore (1999) have drawn attention to a number of more idiosyncratic constructions that have not received much attention recently within mainstream linguistics.

8.5.1 The Ethic Dative construction

The ethic dative adds a benefactive first argument to a wide variety of verbs:[6]

6. (24a) is famously from West (1933)—cf. Frishberg (1962) and O'Day (1962) on the productivity of this and other constructions discussed here. (24b) is archaic fromShakespeare (c.1592), and illustrates its ambiguity.

(24) a. Peel me a grape.
b. I'm going to catch me/myself a trout
c. Knock me at this gate.

The construction is probably the origin of the lexicalized dative alternation for verbs like "give" and "tell" discussed in the last chapter. But in its productive form it is mostly restricted in Modern English to pronominal datives, and oddly restricted as to the verbs it can occur with (Goldberg, 2019):

(25) a. #Peel the man in the Brooks Brothers' shirt a grape.
b. #Explain me what you mean.
c. #You can donate me £10 to the cause.

We will assume that such pronouns head the construction via morpho-lexlcial clitic categories like the following, and that their non-productivity is a result of other factors which we return to in the Discussion section **??** below.[7]

(26) me := $((S\backslash NP)/NP)\backslash_{\diamond}((S\backslash NP)/NP) : \lambda v\lambda x\lambda y.v\,x\,y \wedge benefit\,(v\,x\,y)\,me$

8.5.2 The *way* construction

The following examples suggest that the "*way* construction" analyzed by Jackendoff (1990) and (Goldberg, 1995, 2006, 2019:35), which in other respects resembles the resultatives discussed in section 8.4, is specifically governed by reflexives like *his way*:

(27) a. Marcel bribed his way to the top.
b. # Marcel bribed a/the/Anna's way to the top.
c. # Marcel bribed his path/career to the top.

Thus, it can only apply to intransitives, including those like *ate* and *bribed* that result from intransitivization (whose semantics is to be discussed in chapter 14):

(28) *Marcel bribed judges his way to the top

This seems to indicate that it is a reflexive of the kind seen in 7.6 differing from standard accusatives in bearing *3sm* subject agreement.[8]

(29) his way := $(S\backslash NP_{3sgm})\backslash((S\backslash NP_{3sgm})/NP) : \lambda p\lambda y.p\,(his\,way)$

In all other respects, the derivation is parallel to the intransitivized resultative (20):[9]

7. The semantics could be made less clunky with a Davidsonian event variable.
8. The semantics is a placeholder to which we return in chapter 14.
9. This proposal is not the same as one made in Steedman and Baldridge 2011.

(30)
$$\begin{array}{cccc}
\text{Marcel} & \text{bribed} & \text{his way} & \text{to the top.}\\
\hline
S/(S\backslash NP_{3sgm}) & S\backslash NP_{agr} & (S\backslash NP_{3sgm})\backslash((S\backslash NP_{3sgm})/NP) & ((S\backslash NP_{agr})\backslash(S\backslash NP_{agr}))\backslash((S\backslash NP_{agr})\backslash((S\backslash NP_{agr})/NP))\\
:\lambda p.p\,marcel & :\lambda y.bribed\,y & :\lambda p\lambda y.p\,(his\,way)\,y & :\lambda p\lambda q\lambda y.cause\,(q\,y)\,(p\,to\,(the\,top))\\
\cline{3-4}
 & & \multicolumn{2}{c}{(S\backslash NP_{3sgm})\backslash(S\backslash NP_{3sgm})\quad :\lambda q\lambda y.cause\,(q\,y)\,(to\,(the\,top)\,(his\,way))} \quad < \\
\cline{2-4}
 & \multicolumn{3}{c}{S\backslash NP_{3sgm} : \lambda y.cause\,(bribed\,y)\,(to\,(the\,top)\,(his\,way))} \quad < \\
\hline
\multicolumn{4}{c}{S : cause\,(bribed\,marcel)\,(to\,(the\,top)\,(his\,way))} \quad >
\end{array}$$

(We defer discusssion of how *his* gets bound to *Marcel* until chapter 14 on binding.)

This also seems to be the way that the agreement works in (15), repeated here:

(31) Dora yelled herself hoarse.

8.5.3 The *doing* construction

As with the bounded lexically-governed constructions, many more unbounded constructions offer themselves as lexicalizable in this way. For example, the following (from Kay and Fillmore 1999; Kay 2002) seems a suitable case for treatment with "doing" as head:[10]

(32) a. What's this fly doing in my soup?
b. What do you think this fly is doing in my soup?
c. What's this fly think it's doing in my soup?

(33)
$$\begin{array}{ccccc}
\text{What} & \text{is} & \text{this fly} & \text{doing} & \text{in my soup}\\
\hline
S_{whq}/(S_{inv}/NP) & (S_{inv}/VP_{ing})/NP_{3s} & NP^{\uparrow}_{3s} & (VP_{ing}/VP_{pred})/NP & PP^{\uparrow}\\
:\lambda p\lambda x.[Q]p\,x & :\lambda y\lambda p.p\,y & :\lambda p.p\,fly & :\lambda x\lambda p\lambda y.cause\,(p\,y)\,x & \lambda x.insoup\,x\\
\cline{2-3}\cline{4-5}
 & \multicolumn{2}{c}{S_{inv}/VP_{ing} : \lambda p.p\,fly} \quad < & \multicolumn{2}{c}{VP_{ing}/NP : \lambda x\lambda y.cause\,(insoup\,y)\,x} \quad <\mathbf{B}_{\times}\\
\cline{2-5}
 & \multicolumn{4}{c}{S_{inv}/NP \quad S : \lambda x.cause\,(insoup\,fly)\,x} \quad >\mathbf{B}\\
\hline
\multicolumn{5}{c}{S_{whq} : \lambda x.[Q]cause\,(insoup\,fly)\,x} \quad >
\end{array}$$

8.5.4 The *so big a mess* construction

The following construction seems to have been first noticed by Berman (1974), who called it the "big mess" construction (Van Eynde, 2007; Kim and Sells,

10. I am grateful to Amir Zeldin for proposing a related analysis.

2011; Kay and Sag, 2012):[11]

(34) a. This administration is so big a mess that I ran away
b. Citi is too big a bank to fail

The construction is choosy in the same sense as the *way* construction: in particular, unlike its uninverted relative, the *mess* NP has to be indefinite:

(35) a. A/the/some/every mess so big that it will never be cleaned up
b. So big a/*the/*some/*every mess that it will never be cleaned up

We will assume that degree modifiers like *so* have as basic categories the following:

(36) a. so := $(AP/S')/AP : \lambda p \lambda t \lambda y.cause\,(p\,y)\,t$
b. too := $(AP/VP_{to})/AP : \lambda p \lambda q \lambda y.cause\,(p\,y)\,(\neg q\,y)$

The basic categories give rise to examples like the following

(37) a. This mess is so big that I ran away.
b. That bank is too big to fail.

To cover (34) we need the following further special category for indefinite articles to govern the construction:[12]

(38) a/an := $((NP/XP)\backslash(AP/XP))/N : \lambda n \lambda p \lambda q \lambda y.n\,y \wedge p\,q\,y$

This category takes a noun such as *mess* and a degree modifier such as *too* with one of the categories (36), and yields the latter category with a logical form that will yield properties equivalent to the following:

(39) a. $\lambda y.mess\,y \wedge cause\,(big\,y)\,(ran\,away\,,\,me)$
b. $\lambda y.bank\,y \wedge cause\,(big\,y)\,\neg fail\,y)$

For example, we have the derivation in figure 8.1 for (34b).[13]

Thus, the sentences (34) are roughly paraphrasable as in (40):

11. We omit further cases involving degree comparisons discussed by these authors.

12. The fact that in English this function is carried by the indefinite seems to be a historical accident. (In some dialects, the indefinite in this construction is preceded by *of*, as in *too wide of a margin* (Kennedy and Merchant, 2000:143; Kim and Sells, 2011:339). In other languages such as French, this is not a possible meaning of the indefinite, and must be lexicalized differently CHECK:

(i) *Citi est trop grande comme/*une banque pour échouer.

13. In making *a* an non-determiner and the sister of the noun, and in making the degree marker *too* the sister of the adjective, this analysis resembles the HPSG analyses of Van Eynde (2007) and Kay and Sag (2012). In other respects, the analyses are different.

$$
\begin{array}{ccccccc}
\text{Citi} & \text{is} & \text{too} & \text{big} & \text{a} & \text{bank} & \text{to fail} \\
\hline
NP^{\uparrow} & (S\backslash NP)/NP & (AP/VP_{to})/AP & AP & ((NP/XP)\backslash(AP/XP))/N & N & VP_{to} \\
:\lambda p.p\,citi & :\lambda p\lambda y.p\,y & :\lambda p\lambda q\lambda y.cause\,(p\,y)\,(\neg q\,y) & :big & :\lambda n\lambda p\lambda q\lambda y.n\,y\wedge p\,q\,y & :citi & :\lambda y.fail\,y \\
\end{array}
$$

$$
\begin{array}{l}
\xrightarrow{>}\ AP/VP_{to} : \lambda q\lambda y.cause\,(big\,y)\,(\neg q\,y) \\
\xrightarrow{>}\ (NP/XP)\backslash(AP/XP) : \lambda p\lambda q\lambda y.bank\,y\wedge p\,q\,y \\
\xrightarrow{<}\ NP/VP_{to} : \lambda q\lambda y.bank\,y\wedge cause\,(big\,y)\,(\neg q\,y) \\
\xrightarrow{>}\ NP : \lambda y.bank\,y\wedge cause\,(big\,y)\,(\neg fail\,y) \\
\xrightarrow{>}\ S\backslash NP : \lambda y.bank\,y\wedge cause\,(big\,y)\,(\neg fail\,y) \\
\xrightarrow{>}\ S : bank\,citi'\wedge cause\,(big\,citi)\,(\neg fail\,citi)
\end{array}
$$

Figure 8.1:

(40) a. This administration is a mess whose large size caused me to run away.
b. Citi is a bank whose large size will cause it to not fail.

The construction can give the appearance of interacting with the *tough*-movement construction (section 9.10):

(41) a. The view is so lovely to look at.
b. The problem is too hard to solve.
c. The box is too heavy to lift

However, the class of predicates that take part is broader than just the *tough*-adjectives. We assume that these examples involve a further category for the comparative.

(42) a. so := $(AP/(VP_{to}/NP))/AP : \lambda p \lambda t \lambda y.cause\,(p\,y|)\,(q\,y)$
b. too := $(AP/(VP_{to}/NP))/AP : \lambda p \lambda q \lambda y.cause\,(p\,y)\,(\neg q\,y)$

Since these categories are of the form $(AP/X)/AP$ the category (38) can apply to them to allow the following:

(43) a. It is so lovely a view to look at.
b. It is too hard a problem to solve.
c. It is too heavy a box to lift.

The generalization underlying many of the above analyses of otherwise puzzlingly idiosyncratic constructions is that, where their idiosyncrasy is such as to make it appear that the construction requires as an argument a category that is not normally subcategorized for, such as the definite or indefinite NP, or the main verb *doing*, then we suspect the specifier of that category to be the governor of the construction.

A number of other constructions identified by Goldberg can be lexicalized with similar apparatus, from idioms like “kick the bucket” and a number of causatives like “hammer the metal flat” to fully productive constructions such as the “ethic dative” exemplified in examples like the follwoing (cf. Abeille and Schabes 1989 and Kay 2002):

(44) Cry me a river.

8.6 Discussion

The fact that there have been so many different approaches to these constructions raises the question of whether any of them, including the present one, are getting anywhere near the truth.

Goldberg (1995, 2006, 2013a) suggests that approaches like those of Levin (1993), Pinker (1989), and Rappaport Hovav and Levin (1998) based on lexical rules or related derivational argument-structural templates are implausible, because: (a) they overgeneralize constructions to verbs that don't allow them (*#it colds my bottom*); (b) they require the postulation of implausible specialized verb senses (*sing me back home* $\vdash$ sing : *cause to move somewhere by singing*), or implausibly specific subcategorizations (*bribed his way to the top* $\vdash$ bribe := $VP/PP_{loc})/NP_{poss,way}$; (c) they do not extend to idioms (*look on the bright side* $\vdash$ *# look on the side*).

Some of these apparent overgeneralizations stem from incorrect specification of the lexical derivation. For example, we saw the causative transitive *open* (3a) derived from stative "open" via the inchoative intransitive. Since there is no inchoative *cold* (*#My bottom colds*), there is no corresponding causative transitive (*#It colds my bottom*).

We have also seen that some of these overgeneralizations arise from specifying alternations in terms of verbal syntactic types, in the style of Levin (1993). It seems likely that there is no set of verb types that is fine-grained enough to capture these generalizations completely (Dang, Rosenzweig, and Palmer, 1997). However, we have also seen that some of the more productive constructions should probably be governed by elements other than the verb. Thus *sing me back home* involves the standard meaning of *sing*, the standard benefactive meaning of the ethic dative *me*, and a special causative/resultative meaning available for all locative adverbials like *back home*. Similarly, *sneezing your way across the room* is headed by a special category for *way* reflexives, rather than by a semantically or categorially unreasonable category for *sneeze*. Moreover, since the lexicon clearly contains many multi-word items with specialized meanings, such as *Monday Night Football* (Jackendoff, 2002), there is nothing to prevent idioms like *looking on the bright side* being lexicalized with an interpretation equivalent to *take an optimistic view*.

Thus, despite its radical lexicalism, CCG is not subject to the criticisms Goldberg levels against type-based accounts of category alternations. Indeed, despite frequent denials of lexicalism, and invocations of a "constructicon" whose main identifying characteristic is that it is not the lexicon or associated lexical rules, the Construction Grammatical account seems essentially categorial in its assumptions.

Nevertheless, Goldberg (2013b,a) has a point when she notes that many of these constructions are idiosyncratic and only partially productive, in a way that mere lexicalism leaves unexplained. In CCG terms, we still need to ask ourselves why phrases like *his/your way* can take on this particular function in the lexicon.

Goldberg's answer, following Langacker (1988) and Bybee (2013), is that such constructions are "usage-based". In part, this means that, though in many cases they were at some earlier stage of the language's history entirely compositional (*Knock me at this gate, He went his weary way to London*), they have become lexicalized, and are learned as non-productive or semi-productive idioms from exemplars during the process of first language acquisition by the child, perhaps by a mechanism such as the one discussed in appendix B, which learns a statistical model (of a kind familiar from the computational literature on wide-coverage parsing) of all grammatical possibilities that it has seen evidence for in the form of potential meanings associated with strings of the language, rather than a grammar as such. Such semi-productive systems are extremely common in the lexicon. The English past tense-system is a case in point, in which a small number of highly frequent irregular forms coexists with an open-ended productive morphology. Yang (2016) discusses an elegant mathematical model of the statistical dynamics of such systems.

Seen in this light, the Construction Grammarians insistence on a usage-based Constructicon as the repository of constructions, as distinct from the lexicon (Goldberg, 2019:34-40), seems really to be an insistence on the centrality of what a computational linguist would call "the parsing model".

The computational linguists would certainly agree that the statistical model is at least as important a part of the problem as the grammar. Ineed, there are some who think that the model, in the form of a neural-computional "Sequence-to-Sequence" transducer from strings to meaning representations, directly trained on string-meaning pairs, is the only component that is needed. However, it is hard to believe that any linguist could accept such a position.

The weakness of such models lies their limited ability to compositionally generalize beyond the training data. It is also unclear how meaning representations adequate for full language understanding can be obtrained to support training such transducers. As we have noted at a number of points, What is lacking from all of these accounts is a systematic framework of rules of paradigmatic lexical inheritance across lexical entries. It is quite possible in principle and entirely desirable to add such a framework to the present account, as Beavers (2004) has proposed. But we do not really know how to do that.

Exercise : Identify a construction that you feel sure cannot be lexicalized. Then lexicalize it, and criticize the result. (You can even use one of the constructions in this chapter, so long as you do the last part.)

Chapter 9

Wh-Constructions

Le silence éternel de ces espaces infinis m'effraie.[1]
—*Pensées*, ¶ 206 Blaise Pascal, 1671

The preceding chapters have argued that all bounded constructions, including those like raising and *there*-insertion that allow dependencies to grow across unboundedly many non-finite clauyse boundaries, are lexically headed.

The unbounded constructions such as relativization are those in which the elements of a semantic dependency are separated by unboundedly many finite-clause boundaries without being semantically arguments or dependents of the intervening verbs, as the long-distance raised/controlled arguments are in (**??**) and (**??**) of chapter 9. They have widely been held to require rules of discontinuity or "action at a distance, such as *wh*-movement. The present chapter argues that they too are lexically headed, this time to a first approximation by the *wh*-item.

The account proposed here adheres to the following principle, which further constrains the lexicon to which combinatory projection applies:[2]

(1) *The Projective Dependency Principle (PDP)*
A single non-disjunctive lexical category for the governor of a given construction specifies both the bounded dependencies that arise when its complements are in canonical position and the unbounded dependencies that arise when those complements are displaced under relativization, coordination, and the like.

The latter principle does not prevent a given word from heading more than one construction, and hence being associated with more than one category. Nor does it exclude the possibility that a given word-sense pair may permit more than one canonical order, and hence have more than one category per sense. The claim is simply that, normally, a single category specifies *both* local *and* long-range syntactic and semantic dependencies for the construction

1. "The eternal silence of these infinite spaces terrifies me."
2. Cf. *SP*:33, where the same principle is called "Head Categorial Uniqueness".

that it governs. It follows that the dependencies that in the terms of Dependency Grammar (Hays, 1964, 1967) are called "non-projective" cannot exist: all dependency is projective.

For example, we shall see in later sections that the single lexical syntactic category $VP/NP : \lambda x \lambda y.see\,x\,y$ that we have assumed for the transitive verb "see", which simply specifies an object to its right and does not distinguish between "antecedent," "θ," or any other variety of government, is involved in all of the dependencies underlined in (2):

(2) a. Gabbitas sees a bird.
 b. Gabbitas sees every bird.
 c. the bird that I believe that Gabbitas sees
 d. I believe that Gabbitas sees and you believe that Thring hears, a bird.

The principle (1) has the effect of keeping the grammar relatively small. However, just as the definition above allows the exceptional possibility that an argument *only* appears in situ, it also allows categories to specify arguments as exceptionally *only* supporting extraction. The impossibility in general of subordinate subject extraction in rigid SVO languages like English, and the existence of a small class of English bare-complement verbs like *believe* that *do* allow extraction of embedded subjects means that those verbs require an additional special case lexical category to license the extraction. These and other cases are discussed in Steedman 1991, and more briefly in section 9.6 of chapter 9.

Like the Minimalist assumption of the Copy Theory of Movement, the present Projective Dependency Principle (1) makes redundant Chomsky's 1981 Empty Category Principle (ECP), which said that the trace residues of movement had to be in positions where the verb involved would normally govern its arguments.

The Projective Dependency Principle (1) distinguishes CCG and the Minimalist Program from both TAG and GPSG, in which local and long-distance dependencies are mediated by different initial trees, rules, and/or categories, and from HPSG, in which they are mediated by a disjunctive category using SLASH features.[3]

3. HPSG SLASH features are a descendant of GPSG's slash notation for passing long-range dependency information through the CF derivation. While there is a historical reason for the GPSG/HPSG slash features being so named (Gazdar 1981:159), they are, as noted earlier, interpreted quite differently from categorial slashes, as specifically denoting extraction, rather than general-purpose specification. LFG represents extraction by "functional uncertainty", defining an f-structural path between source and target of what is in other respects a movement analysis.

It is the inclusion of the rules of composition introduced as (6) in chapter 4.1 that will allow capture of relativization and other unbounded phenomena in CCG without resorting to movement or any related notion of "action at a distance". The first such construction we will consider is the *Wh*-question.

9.1 *Wh*-questions

The following observations concerning the formation of *Wh*-questions by Koopman (1983) can be explained in terms of the earlier discussion of Subject-AUX inversion in section 7.1 of the last chapter and the categories shown for the *Wh*-element. EnglishObject *wh*-questions require subject-aux-inversion, as in (3), and are not equivalent to noninverting object indirect questions like (15):[4]

(3)
$$\begin{array}{cccc}
\text{What} & \text{did} & \text{Mary} & \text{buy?} \\
\hline
S_{whq}/_{\diamond\star}(S_{inv}/NP) & (S_{inv}/VP)/NP & NP^{\uparrow} & VP/NP \\
 & \multicolumn{2}{c}{\underline{\qquad\qquad\qquad}<} & \\
 & \multicolumn{2}{c}{S_{inv}/VP} & \\
 & \multicolumn{3}{c}{\underline{\qquad\qquad\qquad\qquad}>\mathbf{B}} \\
 & \multicolumn{3}{c}{S_{inv}/NP} \\
\multicolumn{3}{c}{\underline{\qquad\qquad\qquad\qquad\qquad}>} & \\
\multicolumn{3}{c}{S_{whq}} &
\end{array}$$

(4)
$$\begin{array}{ccc}
\text{What} & \text{Mary} & \text{bought} \\
\hline
S_{wh}/_{\diamond\star}(S/NP) & NP^{\uparrow} & VP/NP \\
 & \multicolumn{2}{c}{\underline{\qquad\qquad}>\mathbf{B}} \\
 & \multicolumn{2}{c}{S/NP} \\
\multicolumn{3}{c}{\underline{\qquad\qquad\qquad\qquad}>} \\
\multicolumn{3}{c}{S_{wh}(\neq S_{whq})}
\end{array}$$

Subject *wh*-questions require the *non*-inverting verb category, and exclude *do*-support:

(5)
$$\begin{array}{cc}
\text{Who} & \text{bought } \textit{Ulysses}? \\
\hline
S_{whq}/_{\diamond\star}(S\backslash NP) & S\backslash NP \\
\multicolumn{2}{c}{\underline{\qquad\qquad\qquad\qquad}>} \\
\multicolumn{2}{c}{S_{whq}}
\end{array}$$

(6)
$$\begin{array}{ccc}
*\text{Who} & \text{did} & \text{buy } \textit{Ulysses}? \\
\hline
S_{whq}/_{\diamond\star}(S\backslash NP) & (S_{inv}/VP)/NP & VP \\
\multicolumn{2}{c}{\underline{\qquad\qquad\qquad\qquad}*} &
\end{array}$$

4. We suppress details of the semantic side of the derivation until the discussion of the related relative clause construction in section 9.4.

(The latter is ungrammatical in the absence of contrastive accent on *did*.)[5]

Wh-determiners project these categories onto more complex *wh*-items:

(7)
$$
\begin{array}{ccccc}
\text{What} & \text{book} & \text{did} & \text{Mary} & \text{buy?} \\
\overline{(S_{whq}/_{\diamond\star}(S_{inv}/NP))/N} & \overline{N} & \overline{(S_{inv}/VP)/NP} & \overline{NP^{\uparrow}} & \overline{VP/NP} \\
\multicolumn{2}{c}{\overline{}{>}} & \multicolumn{2}{c}{\overline{}{<}} & \\
\multicolumn{2}{c}{S_{whq}/_{\diamond\star}(S_{inv}/NP)} & \multicolumn{2}{c}{S_{inv}/VP} & \\
 & & \multicolumn{3}{c}{\overline{}{>}\mathbf{B}} \\
 & & \multicolumn{3}{c}{S_{inv}/NP} \\
\multicolumn{5}{c}{\overline{}{>}} \\
\multicolumn{5}{c}{S_{whq}}
\end{array}
$$

(8)
$$
\begin{array}{ccc}
\text{Which} & \text{woman} & \text{bought } \textit{Ulysses}\text{?} \\
\overline{(S_{whq}/_{\diamond\star}(S\backslash NP))/N} & \overline{N} & \overline{S\backslash NP} \\
\multicolumn{2}{c}{\overline{}{>}} & \\
\multicolumn{2}{c}{S_{whq}/_{\diamond\star}(S\backslash NP)} & \\
\multicolumn{3}{c}{\overline{}{>}} \\
\multicolumn{3}{c}{S_{whq}}
\end{array}
$$

Crucially, the inclusion of raised (cased) argument types in the morpholexicon and rules of functional composition in syntax means that object *Wh*-questions are immediately predicted to be *unbounded*—that is, an unbounded number of embeddings may separate the *wh*-element from the verb it depends on.

(9)
$$
\begin{array}{ccccccc}
\text{What} & \text{did} & \text{Mary} & \text{say} & \text{that} & \text{she} & \text{bought?} \\
\overline{S_{whq}/_{\diamond\star}(S_{inv}/NP)} & \overline{(S_{inv}/VP_{inf})/NP} & \overline{NP^{\uparrow}} & \overline{VP_{inf}/_{\diamond\star}S'} & \overline{S'/_{\diamond\star}S} & \overline{NP^{\uparrow}} & \overline{(S\backslash NP)/NP} \\
 & \multicolumn{2}{c}{\overline{}{<}} & & & \multicolumn{2}{c}{\overline{}{>}\mathbf{B}} \\
 & \multicolumn{2}{c}{S_{inv}/VP_{inf}} & & & \multicolumn{2}{c}{S/NP} \\
 & \multicolumn{3}{c}{\overline{}{>}\mathbf{B}} & \multicolumn{3}{c}{\overline{}{>}\mathbf{B}} \\
 & \multicolumn{3}{c}{S_{inv}/_{\diamond\star}S'} & \multicolumn{3}{c}{S'/NP} \\
 & \multicolumn{6}{c}{\overline{}{>}\mathbf{B}} \\
 & \multicolumn{6}{c}{S_{inv}/NP} \\
\multicolumn{7}{c}{\overline{}{>}} \\
\multicolumn{7}{c}{S_{whq}}
\end{array}
$$

Once again, it emphasised that this mechanism for establishing unbounded dependency is quite different from the notationally similar gap-feature-passing mechanism of GPSG and HPSG. Whereas GPSG and HPSG pass an NP-gap feature from *What* to *bought* at each production in the derivation, the composition of X/Y and Y/Z to yield X/Z is entirely independent of the presence elsewhere of a gap of type Z—or indeed of a gap of any type, as inspection of the many applications of composition in (9) and (**??**) will make clear.

In fact, the establishment of the dependency in the last step of (9) is more

5. We shall see in chapter 6 on spoken intonation that words with and without accent count as distinct lexical items, much as stems with and without tense morphology do.

akin to the uniform "swoop" account of movement of Ross (1967) than it is like the standard punctuated successive cyclic movement theory, much less the "hyper-cyclic" feature-passing at every rule-application of G/HPSG or Neeleman and van de Koot (2010), a point to which .

When more than one *Wh*-element is questioned, then only one of them is fronted. (Chomsky, 1973) noted the "superiority effect" in English illustrated by the following apparent asymmetry:[6]

(10) Who bought what?

(11) #What did who buy?

The possibility of *in situ* wh-elements in examples like (10) is accounted for if they are allowed to have order-preserving *in situ* *Wh*-question-forming categories like the one illustrated for *what* in the following derivation:

(12)
$$\begin{array}{ccc}
\text{Who} & \text{bought} & \text{what?} \\
\overline{S_{whq}/_{\diamond\star}(S\backslash NP)} & \overline{(S\backslash NP)/NP} & \overline{S_{whq}\backslash(S_{whq}/NP)} \\
\multicolumn{2}{c}{\underline{\hphantom{S_{whq}/_{\diamond\star}(S\backslash NP)\ (S\backslash NP)/NP}}{>}\mathbf{B}} & \\
\multicolumn{2}{c}{S_{whq}/NP} & \\
\multicolumn{3}{c}{\underline{\hphantom{S_{whq}/_{\diamond\star}(S\backslash NP)\ (S\backslash NP)/NP\ S_{whq}\backslash(S_{whq}/NP)}}{<}} \\
\multicolumn{3}{c}{S_{whq}}
\end{array}$$

In the absence of a further in situ category $(S_{inv}/VP)\backslash((S_{inv}/VP)/NP)$ for nominative *who*, raised over the inverting auxiliary, (11) cannot be derived. However, the inclusion of such a category would immediately overgenerate *in situ* *Wh*-questions like the following with the meaning as (5):

(13) *Did who buy *Ulysses*?

There could not, therefore, be a language that just like English except for including that further category for *who*. The superiority asymmetry seems to be a side-effect of other idiosyncrasies of the English question-formation system. In other languages allowing multiple *Wh*-elements, superiority effects are not forced, as Bošković (2002) has shown for Slavic languages.

In Japanese, all *wh*-question forming expressions are *in situ*. Thus we have:

6. The superiority effect is weak for some speakers, and the judgements correspondingly uncertain.

(14)

$$
\begin{array}{cccc}
\text{John}-\text{wa} & \text{nani}-\text{o} & \text{kaimasita} & \text{ka?} \\
\text{John.TOP} & \text{what.ACC} & \text{bought} & \text{Q} \\
\overline{S/(S\backslash NP_{nom})} & \overline{(S_{whq}\backslash NP_{nom})/((S_q\backslash NP_{nom})\backslash NP_{acc})} & \overline{(S\backslash NP_{nom})\backslash NP_{acc}} & \overline{S_q\$\backslash S\$} \\
 & & \multicolumn{2}{c}{\underline{\hspace{8em}}{<}} \\
 & & \multicolumn{2}{c}{(S_q\backslash NP_{nom})\backslash NP_{acc}} \\
 & \multicolumn{3}{c}{\underline{\hspace{20em}}{>}} \\
 & \multicolumn{3}{c}{S_{whq}\backslash NP_{nom}} \\
\multicolumn{3}{c}{\underline{\hspace{24em}}{>}} & \\
\multicolumn{3}{c}{S_{whq}} &
\end{array}
$$

"What did John buy?"

It should be remarked that the English fronting *Wh*-question–forming categories given above, like all fronting categories, are semantically type-raised, but are not order-preserving, unlike the above Japanese *wh*-item. They conform to the Combinatory Projection Principle (CPP, (5) of chapter (2)) by marking their result S_{whq} as a different type of clause from the result S of the the functor that they are raised over.

9.2 Indirect questions

Embedded or "indirect" questions of the kind already seen as (15) occur as the complement to verbs like "ask", "know", and "wonder":

(15)

$$
\begin{array}{ccccc}
\text{I} & \text{wonder} & \text{what} & \text{Mary} & \text{bought} \\
\overline{S/(S\backslash NP_{1s})} & \overline{(S\backslash NP)/S_{wh}} & \overline{S_{wh}/_{\diamond\star}(S/NP)} & \overline{NP^{\uparrow}} & \overline{VP/NP} \\
 & & & \multicolumn{2}{c}{\underline{\hspace{5em}}{>}\mathbf{B}} \\
 & & & \multicolumn{2}{c}{S/NP} \\
 & & \multicolumn{3}{c}{\underline{\hspace{12em}}{>}} \\
 & & \multicolumn{3}{c}{S_{wh}} \\
 & \multicolumn{3}{c}{\underline{\hspace{14em}}{>}} & \\
 & \multicolumn{3}{c}{S\backslash NP} & \\
\multicolumn{3}{c}{\underline{\hspace{18em}}{>}} & & \\
\multicolumn{3}{c}{S} & &
\end{array}
$$

Interestingly, indirect multiple *wh*-questions do not appear to suffer from superiority effects (or at least they are much reduced), supporting the earlier suggestion that such effects arise from other idiosyncrasies of English question forms, in particular *do*-support:

(16) a. I asked which woman bought which book.
b. I asked which book which woman bought.

9.3 Topicalization

Topicalization, as in (17), requires a "fronting" category similar to that of a *wh*-question or indirect question item such as *what*, as in (18):

(17) This movie, I like.

(18) a. What := $S_{whq}/_{\diamond\star}(S_{inv}/NP) : \lambda p\lambda x.[Q]p\,x$
b. what := $S_{wh}/_{\diamond\star}(S/NP) : \lambda p\lambda n\lambda x.[Q]p\,x$
c. This movie := $S_{top}/_{\diamond\star}(S/NP) : \lambda p.p(\mathit{this\,movie}) \wedge \mathit{topic}\,(\mathit{this\,movie})$

Since more or less any argument can be topicalized, this is a further source of lexical ambiguity in arguments and their specifiers such as determiners and prepositions. As in the case of cased raised types, we can think of this either in terms of ambiguity "compiled out" into the lexicon, or in terms of underspecified categories like determiners $NP^{\uparrow}/N$. Since topicalization is a root transformation and can only occur in sentence intitial position in English, disambiguation by the processor is easy in either case.

Such raised categories are again order-changing, and again mark their result as a clause-type distinct from *S*, in conformity to CPP, (5) of chapter 2. Languages are free to include them or not, as we saw for the English "superiority effect" contrast in (10) and (11). Similarly, French lacks any equivalent of the topicalization category (18c), although in other respects its *wh*-constructions are quite similar to English.

The category (18c), lexicalized via determiners like *this* and certain intonational markers, maps similar residues onto English topicalized sentences like *This movie, (I think that) you will like*, with a side-effect asserting $\mathit{topic}\,(\mathit{this\,movie})$, making *this movie* the discourse topic.[7]

If we can exploit the degrees of freedom in CCG to assign a distinct topicalizing category to phrases that act otherwise as *in situ* objects, as in (18c), then the same degree of freedom allows us to specific similar lexically distinct categories that are *only* allowed in fronted position. The following exclamatory construction (Huddleston and Pullum, 2002:918; Collins, 2005), which involve fronted elements like *what a difference* and *How insensitive* seems to involve such categories:

(19) a. What a difference a day makes.
b. How insensitive I must have seemed.

Thus, they are unbounded:

(20) a. What a difference she said that a day makes.
b. How insensitive I fear that I must have seemed.

7. Halliday, 1967b and Steedman, 2014 refer to topic in this sense as "theme", the latter offering a dynamic logic account of the side-effect.

However, unlike the apparently similar *such a difference* and *so insensitive*, they cannot occur as *in situ* objects:

(21) a. *A day makes what a difference.
b. *I must have seemed how insensitive.

Thus, they seem to involve the following additional categories for *what* and *how*:

(22) a. What := $(S_{xcl}/(S/NP))/NP : \lambda p.extremely(p\,x)$
b. How := $(S_{xcl}/(S/AP))/AP : \lambda p.extremely(p\,x)$

To the extent that both Topicalization and *Wh*-questions resist embedding as complements in English, the fact can be captured by ensuring that no verb's subcategorization is compatible with the *top* and *whq* features they bear. However, this detail is language-specific, and contested even for English.[8]

9.4 Relativization

For English we can assume on semantic grounds that object relative pronouns have the following category, in which "|" is a slash whose value is either $/$ or $\backslash$:[9]

(23) who(m), which, that := $(N_{agr}\backslash_{\diamond\star}N_{agr})/_{\diamond\star}(S \mid NP_{agr}) : \lambda q\lambda n\lambda y.n\,y \wedge q\,y$

Like other "*wh*-moved" types, such relativized elements bear lexicalized non-order-preserving type-raised categories, which conform to the Combinatory projection principle (CPP), (5) of chapter 2, by yielding a type $N\backslash N$ that is distinct from S.

They support derivations like the following:

8. Miyagawa (2017) shows that embedded topicalization is possible under certain circumstances in Japanese.
9. The significance of the agreement features on the category when it is $/$ will become apparent when we consider subject extraction in section 9.6.

(24)

$$
\begin{array}{ccccc}
\text{A} & \text{man} & \text{that} & \text{Harry} & \text{detests} \\
\hline
NP^{\uparrow}_{agr}/_{\diamond\star}N_{agr} & N_{3s} & (N_{agr}\backslash_{\diamond\star}N_{agr})/_{\diamond\star}(S/NP_{agr}) & NP^{\uparrow}_{3s} & (S\backslash NP_{3s})/NP \\
 & :man & :\lambda q\lambda n\lambda y.ny\wedge qy & :\lambda p.p\ harry & :\lambda x\lambda y.detests'xy \\
\end{array}
$$

$$
\begin{array}{ll}
\text{Harry detests} \xrightarrow{>\mathbf{B}} & S/NP \\
 & \lambda x.detests'x\ harry \\
\text{that Harry detests} \xrightarrow{>} & N_{agr}\backslash_{\diamond\star}N_{agr} : \lambda n\lambda y.ny\wedge detests'y\ harry \\
\text{man that Harry detests} \xrightarrow{<} & N_{3s} : \lambda y.man'y\wedge detests'y\ harry \\
\text{A man that Harry detests} \xrightarrow{>} & NP^{\uparrow}_{3s}
\end{array}
$$

As in the case of *Wh*-questions (9), the inclusion of type-raising (case) and composition in the grammar immediately predicts that the construction is unbounded:

(25)

$$
\begin{array}{cccccccc}
\text{A} & \text{man} & \text{that} & \text{Harry} & \text{said} & \text{that} & \text{he} & \text{detests} \\
\hline
NP^{\uparrow}_{agr}/_{\diamond\star}N_{agr} & N_{3s} & (N_{agr}\backslash_{\diamond\star}N_{agr})/_{\diamond\star}(S/NP_{agr}) & NP^{\uparrow}_{3s} & (S\backslash NP_{agr})/_{\diamond\star}S' & S'/_{\diamond\star}S & NP^{\uparrow}_{3s} & (S\backslash NP_{3s})/NP
\end{array}
$$

$$
\begin{array}{ll}
\text{Harry said} \xrightarrow{>\mathbf{B}} & S/_{\diamond\star}S' \\
\text{he detests} \xrightarrow{>\mathbf{B}} & S/NP \\
\text{Harry said that} \xrightarrow{>\mathbf{B}} & S/_{\diamond\star}S \\
\text{Harry said that he detests} \xrightarrow{>\mathbf{B}} & S/NP \\
\text{that Harry said that he detests} \xrightarrow{>} & N_{agr}\backslash_{\diamond\star}N_{agr} \\
\text{man that \dots} \xrightarrow{<} & N_{3s} \\
\text{A man that \dots} \xrightarrow{>} & NP^{\uparrow}_{3s}
\end{array}
$$

It is worth noting that the syntactic type of the relative pronoun resembles that of a type-raised NP determiner or generalized quantifier determiner, except that it is non-order-preserving and has a noun modifier $N\backslash N$ as its result, rather than S.

It should be apparent from derivations like (24) and (25) that it is only *arguments* like the object of *detests* or the complement of *said* that can be extracted or extracted out of. It follows that backward adjuncts (including many type-raised NPs) are predicted to be “islands” in the sense of Ross (1967) .

The precise form of the present version of the slash-typing restrictions on rules (6) (which is slightly different from previous versions) is dictated by the interaction of the “Heavy Shift” (mediated by crossed composition) with extraction (mediated by harmonic composition) in examples like the following:[10]

10. Details of agreement are suppressed to limit visual clutter.

(26)
$$
\begin{array}{llllll}
\text{people} & \text{that} & \text{I} & \text{heard} & \text{yesterday} & \text{that you like} \\
\overline{N} & \overline{(N\backslash_{\diamond\star}N)/_{\diamond\star}(S/NP)} & \overline{NP^{\uparrow}} & \overline{(S\backslash NP)/_{\diamond\star}S'} & \overline{(S\backslash NP)\backslash(S\backslash NP)} & \overline{S'/NP} \\
 & & & \multicolumn{2}{c}{\overline{\qquad (S\backslash NP)/_{\diamond\star}S' \qquad}^{<\mathbf{B}_{\times}}} & \\
 & & & \multicolumn{3}{c}{\overline{\qquad (S\backslash NP)/NP \qquad}^{>\mathbf{B}}} \\
 & & \multicolumn{4}{c}{\overline{\qquad S/NP \qquad}^{>\mathbf{B}}} \\
 & \multicolumn{5}{c}{\overline{\qquad N\backslash N \qquad}^{>}} \\
\multicolumn{6}{c}{\overline{\qquad N \qquad}^{<}}
\end{array}
$$

The inclusion of second-level composition (8) of chapter 4.1 allows related heavy-shifting derivations like the following:

(27)
$$
\begin{array}{llllll}
\text{people} & \text{that} & \text{I shall} & \text{persuade} & \text{tomorrow} & \text{to vote for me} \\
\overline{N} & \overline{(N\backslash_{\diamond\star}N)/_{\diamond\star}(S/NP)} & \overline{S/VP}^{>\mathbf{B}} & \overline{(VP/VP_{to})/NP} & \overline{VP\backslash VP} & \overline{VP\backslash(VP/VP_{to})}^{<\mathbf{T}} \\
 & & & \multicolumn{2}{c}{\overline{\qquad (VP/VP_{to})/NP \qquad}^{<\mathbf{B}^{2}_{\times}}} & \\
 & & & \multicolumn{3}{c}{\overline{\qquad VP/NP \qquad}^{<\mathbf{B}_{\times}}} \\
 & & \multicolumn{4}{c}{\overline{\qquad S/NP \qquad}^{>\mathbf{B}}} \\
 & \multicolumn{5}{c}{\overline{\qquad N\backslash N \qquad}^{>}} \\
\multicolumn{6}{c}{\overline{\qquad N \qquad}^{>}}
\end{array}
$$

The fact that in many dialects of English (including the author's), "inner" arguments of ditransitives can extract, means that we need further categories for the relative pronoun, schematized over ditransitive types, as in the derivation in floated figure 9.1a, crucially involving second-level composition $\mathbf{B}^2$, rule (8a) of chapter 4.1.

It is interesting to compare relativization in "structurally" cased English with Latin, because of the relatively unambiguous morphological case system of the latter. In floated figure 9.1b,unambiguous cased type-raised categories are spelled out, while $NP^{\uparrow}_{m}$ schematizes as usual over all cased masculine NP types.

English relative pronouns like (23) and Latin relative pronoun categories like *quem*, $(NP^{\uparrow}_{m,3s}\backslash NP^{\uparrow}_{m,3s})/(S\backslash NP_{acc})$, above are examples of a general relative pronoun category of the form (28), mapping functions like $S\backslash NP$ into N and NP modifiers, in both cases agreeing in case with the residue of extraction:[11]

(28) $(N_{agr1}|N_{agr1})/(S|NP_{agr2}) : \lambda p\lambda n\lambda x.p\,x \wedge n\,x$

11. See section 9.6 below for details of English embedded subject extraction.

a. (a)

man	that	I	gave	flowers
N $: man$	$((N\backslash N)/X)/((S/X)/NP)$ $: \lambda p\lambda n\lambda x.p\,x \wedge n\,x$	$NP^{\uparrow}$ $: \lambda p.p\,me$	$((S\backslash NP)/NP)/NP$ $: \lambda x\lambda w\lambda y.gave\,w\,x\,y$	$NP^{\uparrow}$ $: \lambda p.p\,flowers'$

$$\text{I gave} \Rightarrow_{>\mathbf{B}^2} (S/NP)/NP : \lambda x\lambda w.gave\,w\,x\,me$$

$$\text{that I gave} \Rightarrow_{>} (N\backslash N)/NP : \lambda w\lambda n\lambda x.gave\,w\,x\,me' \wedge n\,x$$

$$\text{that I gave flowers} \Rightarrow_{<} N\backslash N : \lambda n\lambda x.gave\,flowers\,x\,me' \wedge n\,x$$

$$\text{man that I gave flowers} \Rightarrow_{<} N : \lambda x.gave\,flowers\,x\,me' \wedge man\,x$$

b.

Murus *Wall*	quem *that.ACC*	Balbus *Balbus.NOM*	dificare *build.INF*	vult *wants.pres.3s*
$S/(S\backslash NP_{m,3s,nom})$	$(NP^{\uparrow}_{m,3s}\backslash NP^{\uparrow}_{m,3s})/(S\backslash NP_{acc})$	$S/(S\backslash NP_{m,3s,nom})$	$VP_{inf}\backslash NP_{acc}$	$(S\backslash NP_{3s,nom})\backslash VP_{inf}$

$$\text{dificare vult} \Rightarrow_{<\mathbf{B}} (S\backslash NP_{3s,nom})\backslash NP_{acc}$$

$$\text{Balbus dificare vult} \Rightarrow_{>\mathbf{B}_{\times}} S\backslash NP_{acc}$$

$$\text{quem Balbus dificare vult} \Rightarrow_{>} NP^{\uparrow}_{m,3s}\backslash NP^{\uparrow}_{m,3s}$$

$$\text{Murus quem Balbus dificare vult} \Rightarrow_{<} S/(S\backslash NP_{m,3g,nom})$$

"A/The wall that Balbus wants to build″

Figure 9.1:

The crossing dependencies between verbs and arguments in figure 9.1b arise from crossed composition reordering the nominative and accusative. Of course, this same reordering can be observed in Latin in non-relativized sentences, as in the scrambled version (7) of Chapter (**??**) of Figure 3.1 of Chapter 3.

Any such generalization will be subject to certain constraints on free order and combinatory rules that amount to saying that you can't do anything with an underspecified slash that you couldn't have done by proliferating distinct lexical entries with fully specified slashes (Baldridge 2002). In particular, the Combinatory Projection Principle (5) of chapter 2 requires that the interpretation of such underspecified slashes in the result of applying the composition rules (6) to such underspecified categories be confined to the alternatives allowed by those rules for the fully specified categories

As will be apparent from the Latin examples above, it is the inclusion of crossing composition rules in CCG that allows a limited amount of "scrambling" of word order, and hence allows non-peripheral arguments to extract. Together with the generalization of composition to boundedly polyvalent dependent functions such as $(Y/Z)/W$, the related valency limit on the coordinating categories, and with the limitations on the variety of such rules imposed by the Combinatory Projection Principle (5), it is the source of the "near-context-free" expressive power of CCG (see Joshi et al. 1991; Koller and Kuhlmann 2009; Kuhlmann, Koller, and Satta 2010, 2015, and appendix C).

9.5 Bare relatives

We will assume that English bare relatives like (30, 31) are headed by a special relative pronoun like category for the subject NP and its determiners, by including the following categories, as well as the standard order-prederving ones, making them the head of a noun-modifier, $N\backslash_{\diamond\star}N$ rather than S, and applying to a constituent of the type of a transitive verb rather than :

(29) I $:= (N\backslash_{\diamond\star}N)/((S/NP)/NP) : \lambda p\lambda n\lambda x.nx \wedge pxy$
the $:= ((N\backslash_{\diamond\star}N)/((S/NP)/NP))/N : \lambda m\lambda p\lambda n\lambda x.nx \wedge \mathit{said}\,(px)\,(\mathit{them})$

(30)
$$
\begin{array}{cccc}
\text{The} & \text{man} & \text{Harry} & \text{detests} \\
\hline
NP^{\uparrow}/_{\diamond\star}N & N & (N\backslash N)/((S\backslash NP)/NP) & (S\backslash NP)/NP \\
 & & \multicolumn{2}{c}{\overline{\quad N\backslash_{\diamond\star}N \quad}{>}} \\
 & \multicolumn{3}{c}{\overline{\quad N \quad}{<}} \\
\multicolumn{3}{c}{\overline{\quad NP^{\uparrow} \quad}{>}} &
\end{array}
$$

(31)
$$
\begin{array}{llllll}
\text{The} & \text{man} & \text{Harry} & \text{said} & \text{he} & \text{detests} \\
\hline
NP^{\uparrow}/_{\diamond\star}N & N & (N\backslash_{\diamond\star}N)/_{\diamond\star}((S/NP)/NP) & ((S\backslash NP)/S) & NP^{\uparrow} & (S\backslash NP)/NP
\end{array}
$$
$$
\begin{array}{l}
\text{he detests: } S/NP \quad {>}\mathbf{B} \\
\text{said he detests: } (S\backslash NP)/NP \quad {>}\mathbf{B} \\
\text{Harry said he detests: } N\backslash_{\diamond\star}N \quad {>} \\
\text{man Harry said he detests: } N \quad {<} \\
\text{The man Harry said he detests: } NP^{\uparrow} \quad {>}
\end{array}
$$

Since objects do not similarly head bare relatives in English, this analysis immediately captures the fact that bare subject relatives do not in general exist in English:[12]

(32) *This is the man likes Harry.

It may seem to proliferate nominal lexical category types recklessly. However, the ambiguity of the subject category is uniform and can be specified as a lexical rule, or be left unspecified as $NP^{\uparrow}$ in favor of (**??**) on the partial criterion of left-adjacent N under the guidance of the parsing model. (This is in fact how such ambiguities are handled in practical wide coverage CCG parsers such as Hockenmaier and Steedman (2002a), Clark and Curran (2007b), and Lewis and Steedman (2014a) and the treebanks such as CCGbank that they are modeled on (Hockenmaier and Steedman, 2007).)[13]

A related analysis offers itself for VSO languages like Modern Welsh (Borsley et al., 2007) and Mayan (Ponvert, 2008), and for SOV languages like Japanese (Kuno, 1974, 1973a), in which there are no relative pronouns, and in that sense, *all* relative clauses are bare. The difference is that bare subject relatives are allowed in these languages, so that analysing them analogously to English bare relatives requires allowing all NP arguments of verbs to similarly head relative clauses.

Relative modifiers in these languages obey a generalization stated by Lehmann (1973):48, following Greenberg (1963), as follows:

(33) In VO languages, nominal modifiers such as relative, adjectival, and genitival expressions follow nouns; in OV languages they precede nouns.

For example, in Japanese we have:

12. They are occasionally found in poetic registers—see the epigraph to Chapter 7 for an example.
13. Another analysis is possible in which it is the *verb* that heads bare relative clauses, via categories like $(N\backslash N)/NP)$ for the transitive. This also could be handled via underspecification and/or unary rules.

(34) a. [watakusi-ga kaita] hon
[I wrote] book
N/N N

"book (that) I wrote"

b. [hon-o kaita] hito
[book wrote] person
N/N N

"person *(that) wrote a book"

(However, the object relativization in (34a) is only felicitous if the book is topic, either by virtue of *wa*-marking or by virtue of coreference with the topic, as in the following (Kuno, 1973b:234):)

(35) Kore-wa watakusi-ga kaita hon desu
This-TOP I-NOM wrote book is

"This is a book that I wrote"

See Kuroda (1965), Hasegawa (1985:n1), and below for further discussion, including the possiblity that Japanese relatives are mediated by pro-drop anaphora, as argued by Kuno (1973b). Accordingly, we remain for the moment uncommitted on exactly how $N\backslash N$ is derived in (34).)

In the case of Welsh, we can assume that for every raised NP type $NP^{\uparrow}$, schematized in (11) of chapter 3 as $\mathrm{T}\backslash(\mathrm{T}/NP)$, there is one or more corresponding types in which the argument T/NP has one fewer NP arguments and the result is (a function into) $N\backslash N$. These functions can only apply to the soft-mutated form of the verb, which we assume is distinguished by delivering S_{rel} rather than S as a result, since the soft-mutated form can head interrogatives in its own right. For example,

(36) a. dynes welodd cath

$$\dfrac{\dfrac{}{N : woman} \quad \dfrac{\dfrac{}{(S_{rel}/NP)/NP : \lambda y\lambda x.past\,(see\,x\,y)} \quad \dfrac{}{(N\backslash N)\backslash((S_{rel}/NP)/NP) : \lambda p\lambda n\lambda x.n\,x \wedge p\,x\,(a\,cat)}}{N\backslash N : \lambda n\lambda x.n\,x \wedge past\,(see\,x\,(a\,cat))}<}{N : \lambda x.woman\,x \wedge past\,(see\,x\,(a\,cat))}<$$

"woman (that) a cat saw"

b. dynes welodd gath

$$\dfrac{\dfrac{}{N : woman} \quad \dfrac{\dfrac{}{(S_{rel}/NP)/NP : \lambda y\lambda x.past\,(see\,x\,y)} \quad \dfrac{}{(N\backslash N)\backslash((S_{rel}/NP)/NP) : \lambda p\lambda n\lambda x.n\,y \wedge p\,(a\,cat)\,x}}{N\backslash N : \lambda n\lambda x.n\,x \wedge past\,(see\,(a\,cat)\,x)}<}{N : \lambda x.woman\,x \wedge past\,(see\,(a\,cat)\,x)}<$$

"woman *(that) saw a cat"

Helpfully, as well as marking the relativized verb by soft mutation where applicable, as we saw in chapter4.1 Welsh distinguishes subjects from non-subjects, including those that head bare-relative clauses, by marking the latter with the soft-mutated form, a distinction that would be unmarked in English if it allowed bare subject-relativization.

It would be possible in principle to propose a mirror-image account of bare relativization for Japanese. However, we defer discussion of this question until we have dealt with some more complex relativization phenomena in English.

9.6 Embedded Subject Extraction

The fact that embedded subject extraction from *that*-complements is impossible in English, unlike extraction of other arguments of the verb, follows immediately from the present account, without further stipulations such as the Empty Category Principle (ECP) of Chomsky (1981) or the Generalized Left-Branch Condition (GLBC) or the related SLASH Termination Metarule of GPSG (Gazdar, 1981; Gazdar et al., 1985:161) and Trace Condition of HPSG (Pollard and Sag, 1994:172-3)

(37) a. a man who(m) [I think (that)]$_{S/_{\diamond\star}S}$ [Chester likes]$_{S/NP}$

b. *a man who(m) [I think that]$_{S/_{\diamond\star}S}$ [likes Lester]$_{S\backslash NP}$

c. a man who(m) [I think]$_{S/_{\diamond\star}S}$ [likes Lester]$_{S\backslash NP}$

While, as we have seen, extraction of an object (37a) is accomplished by the standard forward harmonic composition rule (6a), as in (25), extraction of a subject (37b) would require an unrestricted slash type S'/S in the category of verbs like "think" and/or the complementizer "that", allowing the forward crossed compostion rule (6c) to apply to yield **I think that likes Lester*$_{S\backslash NP}$.

However, as pointed out in *SS&I*, if we made that change to the grammar of English to allow (37b), it would immediately *also* allow examples like the following:

(38) *I Chester think that likes Lester.

For the same reason, we cannot capture the exceptional cases of subject extraction that English allows for the small class of verbs that take bare complements, as in (37c), by making the appropriate entry for *think* compatible with both varieties of forward composition. Instead, we must add the following exceptional category for the bare-complement verbs:

(39) think := $((S\backslash NP)/NP_{+WH,agr})/(S\backslash NP_{agr}) : \lambda p\lambda x\lambda y.think\,(px)y$

This category is exceptional, in the sense that it violates the Projective Dependency Principle (1) (PDP) of chapter 2, in being a category that supports extracted but not *in situ* arguments. That is to say, it supports derivations like floated figure 9.2 for (37c), in which for the first time the agreement features on the English relative pronoun category(23) do important work.

The reason the extracting argument $/NP_{+WH,agr}$ has to be *rightward*, will become apparent in the next chapter 11, in connection with example (7), where it supports across-the-board extraction of *accusative* arguments.

But it does not support a corresponding derivation with an "exceptionally case-marked" *in situ* accusative, because all lexical NPs are $-WH$:

(40) a. *I think likes Lester my very heavy friend
b. *I think him likes Lester.

The reader is directed to *SS&I* for further details, where it is shown, following Kayne 1983, that certain predictions that must follow if such PDP-exceptional categories are allowed—for example, that similar constructions might exist that *only* allow extracted subjects, and disallow those in situ—are confirmed in French and English. For example, we have the following asymmetry in English:

(41) a. A man that I assure you to be a genius
b. *I assure you him to be a genius

The earlier work points out that the existence of dialects (including this author's) in which the following asymmetry is exhibited for the non-nominative relative pronoun *whom* is support for the analysis of embedded subject extraction as essentially accusative, an account which bears some relation to the account of Chomsky (1981) in terms of "antecedent government" of subject traces.

(42) a. A man who likes me
b. *A man whom likes me
c. A man who/whom I think likes me

See É. Kiss (1991) for discussion of related asymmetries in Hungarian.

If we can assume a special complement-subject-extracting category for certain verbs then we must predict that the same degree of freedom could be exploited for other categories taking bare S complements, in particular complementizers themselves, in other languages if not in English.

French provides a case in point. First, it strictly distinguishes the nominative

$$
\begin{array}{cccccc}
\text{man} & \text{who} & \text{I} & \text{think} & \text{likes} & \text{Lester} \\
\hline
N_{3s} & (N_{agr}\backslash N_{agr})/(S/NP_{+WH,agr}) & NP^{\uparrow}_{1s} & ((S\backslash NP)/NP_{+WH,agr})/(S\backslash NP_{agr}) & (S\backslash NP_{3s})/NP & NP^{\uparrow} \\
: man & \lambda p\lambda n\lambda x.p\,x \wedge n\,x & \lambda p.p\,me & \lambda p\lambda x\lambda y.think\,(p\,x)\,y & \lambda x\lambda y.likes\,x\,y & lester' \\
 & & \multicolumn{2}{c}{\xrightarrow{\hspace{6em}} {>}\mathbf{B}^2} & \multicolumn{2}{c}{\xrightarrow{\hspace{6em}} {<}} \\
 & & \multicolumn{2}{c}{(S/NP_{+WH,agr})/(S\backslash NP_{agr}) : \lambda p\lambda x.think\,(p\,x)\,me} & \multicolumn{2}{c}{S\backslash NP_{agr} : \lambda y.likes\,lester\,y} \\
 & & \multicolumn{4}{c}{\xrightarrow{\hspace{14em}} {>}} \\
 & & \multicolumn{4}{c}{S/NP_{+WH,3s} : \lambda x.think\,(likes\,lester\,x)\,me} \\
 & \multicolumn{4}{c}{\xrightarrow{\hspace{16em}} {>}} & \\
 & \multicolumn{4}{c}{N_{3s}\backslash N_{3s} : \lambda n\lambda x.think\,(likes\,lester\,x)\,me' \wedge n\,x} & \\
\multicolumn{4}{c}{\xrightarrow{\hspace{16em}} {>}} & & \\
\multicolumn{4}{c}{N_{3s} : \lambda x.think\,(likes\,lester\,x)\,me' \wedge man\,x} & &
\end{array}
$$

Figure 9.2:

relative pronoun *qui* from accusative *que*, as in the following contrast:

(43) a. l'homme *qui/que les oiseaux voient ("the man that the birds see')
b. l'homme qui/*que voit les oiseaux ("the man that sees the birds')

Second, French has a lexically distinct complementiser *qui* that allows *only* embedded subject extraction for *all* verbs taking complements of that type:

(44) l'homme que je t'ai dit qui/*que voit les oiseaux" ("*the man that I told you that sees the birds").

Thus where Englsh has the categories in (45), French has the categories (46):[14]

(45) a. who, that := $(N_{agr}\backslash N_{agr})/(S\backslash NP_{agr})$
b. whom, who, that := $(N_{agr}\backslash N_{agr})/(S/NP_{agr})$
c. that := S'/S
d. think := $((S\backslash NP)/NP_{+WH,agr})/(S\backslash NP_{agr})$

(46) a. qui := $(N_{agr}\backslash N_{agr})/(S\backslash NP_{agr})$
b. que := $(N_{agr}\backslash N_{agr})/(S/NP_{agr})$
c. que := S'/S
d. qui := $(S'/NP_{+WH,agr})/(S\backslash NP_{agr})$

Thus, French displays many of the alternative choices implicit within the degrees of freedom exploited in accounting for the idiosyncrasies of English in the possibiites allowed for subject extractions from tensed complements. The prediction is that the many other fixed word-order verb-medial languages that exhibit similar general constraints on subject extraction will be found the allow similar exceptions within the same degrees of freedom.

The fact that French does so in a way that allows embedded subjects to extract for all verbs taking *qui*-complements raises the possibility that there might be dialects or idiolects of English including a subject extracting category like (46d) for the complementizer *that*, as is suggested by the work of Sobin (1987) and Cowart (1997)—see Pesetsky (2017) for discussion.

The related question of how the child language learner can learn such lexicalized exceptions, rather than engendering collapse of the fixed word-order property by induction of over-general category types, as we noted would be appropriate in a veriable word-order language like Hebrew, which does allow

14. The agreement-passing variable feature-value *agr* is shown for once, since it matters for subject-extraction. Recall that lexical NPs bear the feature $-WH$, and cannot combine with NP_{+WH}.

general subject extraction, is deferred until appendix B.

The **that-t* phenomenon has been claimed above to be a consequence of English rigid word order and the differential directionality of the subject argument of the verb in SVO languages. It is therefore unsurprising to find that parallel constraints do not apply to verb-initial and verb-final languages and constructions. Thus in Italian, the subjects of exactly those verbs that allow subject inversion also allow embedded subject extraction (Perlmutter, 1971; Rizzi, 1982, 1990b; Ishii, 2004):

(47) a. Gianni abbia telephonato. (“Gianni has telephoned”)
b. Abbia telephonato Gianni. (“It is Gianni who has telephoned”)
c. Chi credi che abbia telephonato?
Who.NOM think.2SG that has.PRES telephoned.PPL

(“*Who do you think that has telephoned?”)

This phenomenon is accounted for on the assumption that verbs like *abbia* have the VXS category $(S/NP)/VP_{ppl}$ as well as the standard SVO category, allowing subject extraction in (47c) via harmonic composition:[15]

(48)
$$\begin{array}{ccccc}
\text{Chi} & \text{credi} & \text{che} & \text{abbia} & \text{telephonato?}\\
\hline
S_{whq}/(S/NP) & S/S' & S'/S & (S/NP)/VP_{ppl} & VP_{ppl}\\
\cline{2-3}\cline{4-5}
 & \multicolumn{2}{c}{S/S \quad {>}\mathbf{B}} & \multicolumn{2}{c}{S/NP \quad {>}}\\
\cline{2-5}
 & \multicolumn{4}{c}{S/NP \quad {>}\mathbf{B}}\\
\hline
\multicolumn{5}{c}{S_{whq} \quad {>}}
\end{array}$$

Support for this explanation comes from the fact that VSO Welsh also allows embedded subjects to extract, assuming the account of Welsh relative clauses as headed by root NP categories, parallel to English bare relatives, introduced in section 9.5.

15. Modern Hebrew, which is generally regarded as SVO but both allows post-vverbal subjects and subject-extraction from complements, seems to be a similar case to Italian (Amir Zeldes p.c.).

(49)

$$
\begin{array}{ccccc}
\text{dynes/ddynes} & (*\text{a})\ \text{wyddai} & \text{Gwyn} & (\text{y})\ \text{gweliff} & \textbf{gath} \\
\hline
N_{+fs} & (S_{rel}/S)/NP & ((N\backslash N)/(S_{rel}/NP))\backslash((S_{rel}/S)/NP) & (S/NP)/NP & (S_{rel}/NP)\backslash((S/NP)/NP) \\
:woman & :\lambda y\lambda s.past\,(know\,s\,y) & :\lambda p\lambda q\lambda n\lambda x.n\,x\wedge p\,(q\,x)\,gwyn & :\lambda y\lambda x.will\,(see\,x\,y) & :\lambda p\lambda y.p\,(a\,cat)\,y
\end{array}
$$

$$(N\backslash N)/(S_{rel}/NP) : \lambda q\lambda n\lambda x.n\,x\wedge past\,(know(q\,x)\,gwyn) \quad <$$

$$S_{rel}/NP \quad \lambda y.will\,(see\,(a\,cat)\,y) \quad <$$

$$N\backslash N : \lambda n\lambda x.n\,x\wedge past\,(know(will\,(see\,(a\,cat)\,x))\,gwyn) \quad >$$

$$N_{+fs} : \lambda x.woman\,x\wedge past\,(know\,(will\,(see\,(a\,cat)\,x)\,gwyn)) \quad <$$

("woman that Gwyn knew will see a cat")

(50)

$$
\begin{array}{ccccc}
\text{dynes/ddynes} & (*\text{a})\ \text{wyddai} & \text{Gwyn} & (\text{y})\ \text{gweliff} & \textbf{ca} \\
\hline
N_{+fs} & (S_{rel}/S)/NP & ((N\backslash N)/(S_{rel}/NP))\backslash((S_{rel}/NP)/NP) & (S/NP)/NP & (S_{rel}/NP)\backslash(\\
:woman & :\lambda y\lambda s.past\,(know\,s\,y)x & :\lambda p\lambda q\lambda n\lambda x.n\,x\wedge p\,(q\,x)\,gwyn & :\lambda y\lambda x.will'\,(see'\,x\,y) & :\lambda p\lambda x.p
\end{array}
$$

$$(N\backslash N)/(S_{rel}/NP) : \lambda q\lambda n\lambda x.n\,x\wedge past\,(know(q\,x)\,gwyn) \quad <$$

$$S_{rel}/NP \quad \lambda x.will\,(see\,x\,(a\,cat))$$

$$N\backslash N : \lambda n\lambda x.n\,x\wedge past\,(know(will\,(see\,x\,(a\,cat)))\,gwyn)$$

$$N_{+fs} : \lambda x.woman\,x\wedge past\,(know\,(will\,(see\,x\,(a\,cat))\,gwyn)) \quad <$$

("woman that Gwyn knew a cat will see")

Again, it is helpful that Welsh disambiguates the verb and the residual NP in embedded S_{rel}/NP by soft-mutation. The subscript *rel* is a purely syntactic feature that limits overgeneralization. We shall see it again in the relative constructions of the other Celtic languages Scots Gaelic and Modern Irish. Note that the fragment "wyddai Gwyn" ("Gwyn knew") has the same category as the English object relative pronoun, while the remainder of the relative clause "(gwybodd e) gweliff cath", "(he knew) a cat will see", in which the verb(s) have the standard unmuted form, has the same category as the corresponding residues of English object relativization.

As predicted, the verb-final language Japanese also shows no assymetry in extraction of embedded subjects and objects. In particular it too allows embedded subject-relativization under the conditions noted in connection with (35) (Kuroda, 1965; Kuno, 1973b):

(51) Kore-wa John-ga hon-o kaita to itta hito desu
This-sc top John-NOM book-ACC wrote that said person is
"This is the person who John said (*that) wrote the book"

However, we continue to defer discussion of Japanese relativeization.

There is considerable variation across Germanic dialects and individual speakers as to whether they allow any extraction of embedded arguments, with Northern German speakers tending to disallow any extraction from embedded clauses, while Southern speakers tend to allow extraction of both subjects and objects from embedded clauses. This observation suggests that in the former dialects complementizers such as "daß" have the same category as Irish "go", while in the latter they have the same category as Irish "a(L)".

Bresnan (1977:173;194,n.7), Culicover (1993), and Browning (1996:237,n.1) discuss some cases of English embedded subject extraction where the inclusion of a sentential adjunct adjunct seems to facilitate subject extractions that are otherwise disallowed. The status of many of these examples is rather unclear, as Bresnan and Browning point out, but the most convincing examples involve the negative adverbials that precipitate inversion, as in (52b).

(52) a. a person who(m) I said that under no circumstances would run for any public office
b. I said that under no circumstances would this person run for any public office.

Such examples are parallel to the possibility of complement subject extraction in Welsh discussed above, since under present assumptions negative fronted items require an inverting VSX verb such as the following, which do not require crossed composition for the subject to be accessible:

(53) would := $(S_{inv}/VP)/NP$

Other extraction-facilitating adjuncts discussed by these authors such as (54b) are harder to explain:

(54) a. a person who(m) I said that *(in my opinion) was unfit for any public office.
b. a person who(m) I wonder whether *(in your opinion) is fit for public office.

SS&I:61 suggests that such adverbials carry a *wh*-extraction only category $(S/NP_{+WH,agr})/(S\backslash NP_{agr}$, a suggestion that is supported by the fact that the effect seems to hold for verbs that do not support bare complements, as in (54b).

Many analyses of the constraints on complement subject-extraction and the exceptions to those constraints have been proposed, and are reviewed by Pesetsky (2017). It seems fair to say that none of them are entirely satisfactory. Under the present account, the possibility of asymmetry in extractability of subjects and objects in rigid-order SVO languages like English follows from their lexical specification in their governing category as different in directionality, from which it follows that distinct combinatory rules must apply. Object extraction requires rightward harmonic composition of the complement-taking VP/S into the tensed verb category, so it is potentially allowed. Embedded subject extraction would require forward *crossed* composition into the same category. Since that is a different rule a language is free to independently disallow that extraction. What is more, allowing such crossed composition would immediately allow very free word-order, so we would no longer be able to talk of the language as SVO in the first place.

Conversely, the tendency towards symmetry of embedded subject and object extraction in verb initial (VSO) and verb final (SOV) languages follows from their lexical specification in their governing category as the *same* in directionality, from which it follows that the *same* combinatory rules must apply to them. Thus if a rule of composition allows extraction of objects, then it must apply to subjects. It follow that such languages can exhibit symmetry in extraction, either allowing or disallowing both. In fact, the only way that such languages could exhibit extraction asymmetry would be via the agreement system, say by restricting relative pronouns to nominative, absolutive, etc. agreement, as in Latin (Figure 9.1, Kennedy, 1882:§330), and other cases discussed by Keenan and Comrie (1977).

9.7 Island Constraints

The various types of long-range dependency considered in the preceding sections have one characteristic in common: In every case the displaced element is an argument, such as *NP*, and the domain with respect to which it is displaced is also an argument, such as *VP*, or S/NP, itself composed from heads and complements of arguments. The displaced element may bear its standard, case-raised category, as in Japanese scrambling. Or it may be a special category-changing type, like an English topic. But in all cases it is a (raised) argument.[16]

If a category is an adjunct, such as the adverbials "naked" and "whistling *Dixie*", $VP\backslash VP$, then it is defined as selecting, rather than selected-for, thereby rendering inapplicable all the mechanisms above for extraction. Moreover, incomplete adjuncts such as "whistling" $(VP\backslash VP)/NP$ cannot combine to the

16. The same holds for "remnant" movements, such as Germanic transitive verb-topicalization, as in "Essen wird er Æpfel" ("*Eat will he apples"), (Müller, 1996), since $VP\backslash NP$ is the argument of the German accusative NP, allowing the movement residue $S_{inv}/(VP\backslash NP)$ to form, unlike the corresponding residue in English.

left with VP in advance of rightward with NP.[17]

Under present assumptions adjuncts are therefore predicted to be islands to extraction, as observed by Huang, 1982a:505, following Cattell (1976):[18]

(55) a. *How [do you regret that Sally met Harry?]$_{S_{inv}}$
b. *Who did Harry [file the reports]$_{VP}$ [without telling]$_{(VP\backslash VP)/NP}$?
c. *Who do you know a [man]$_N$ [that met]$_{(N\backslash N)/NP}$?

Since we have type-raised all arguments including indirect questions, in most cases excluding the corresponding unraised category from the lexicon, and type-raised categories are VP-adjunct-like, in the sense that they select verbs and verb-phrases, as in $VP|(VP|X)$ etc, we correctly predict that raised arguments including $NP^{\uparrow}$ and indirect questions are also islands:

(56) a. *What did you [doubt]$_{VP/NP}$ the claim$_N$ [that he has read]$_{(N\backslash N)/NP}$?
b. *Which woods do you think you [know]$_{VP/S_{iq}}$ [who owns]$_{(VP\backslash(VP/S_{iq}))/NP}$?

The possibility of explaining island effects in this way is another desirable consequence of specifying linear correspondence in the lexicon.

Certain constructions that look like extractions from NP and PP like the following possibly arise from lexicalized multi-word expressions or verb-particle constructions, predicting that such expressions as subjects are islands like all arguments, because noun-modifiers are adjuncts $N\backslash N$:[19]

(57) a. Who did they take advantage of?
b. What will you paint a picture of?

In English, S, $\overline{\text{S}}$, and the various kinds of VP complements are the only phrasal arguments that exist in their unraised form. However, their participation in argument-cluster coordinations like the following, analogous to (**??**), suggests that they also bear adjunct or raised categories:[20]

17. Other, that is, than via the parasitic gapping **S** substitution rules we have passed over in section **??** (Szabolcsi, 1983/1992; Steedman, 1987, 1996).
18. Substituting “bridge” verbs like “think” for non-bridge “regret” considerably improves (55a), because bridge-verbs actually do optionally subcategorize for adjuncts.
19. The relative weakness of “derived” subject islands such as unaccusatives found by Jurka (2013) and Polinsky, Gallo, Graff, Kravtchenko, Morgan, and Sturgeon (2013) is not explained in these terms. The latter authors, citing Chung and McCloskey (1983) and Kluender (1998), note definiteness effects that suggest that discourse processing factors may be at work here too.
20. It seems likely that as adjuncts they are extraposed, with an anaphoric relation to an argument of verbs like “ask” and “tell” at the level of logical form.

(58) I will tell$_{(VP/S)/NP}$ [[[Donald] [(that) he is fired]] and [[Ronald] [(that) he is hired]]]$_{VP\backslash((VP/S)/NP)}$.

The categorial ambiguity claimed here for English complements is clearly a lexical degree of freedom upon which languages can be expected to differ, some including sentential complements that bear only adjunct or type-raised categories, making them islands, as appears to be the case for *daß*-complements in many dialects of German, or including different complement types, some of which are adjuncts, and some subcategorized-for arguments, as is the case for *wh/that*-complements in English:

(59) a. *Who were you surprised when you saw?
b. Who were you surprised that you saw?

If a language like English can arrange its lexicon so as to make certain components such as *that*-complements bear both adjunct/type-raised (island) and complement categories, it is clear that we must expect islands in general to appear to exhibit a continuum of extractability, from "strong" islands bearing only adjunct categories that are not subcategorized-for and completely block extraction, to "weak" islands bearing both adjunct and argument categories, the latter sometimes subcategorized for, allowing extraction (Cinque, 1990; Szabolcsi, 2007; Truswell, 2007b,a; Boeckx, 2012:16). Truswell illustrates the strong/weak distinction in minimal pairs like the following, among many others:

(60) a. *What tune does John work whistling?
b. What tune did John drive Mary crazy whistling?

Example (60a) shows that VP-modifiers like "whistling *Dixie*" are not subcategorized-for by predicates like "work" VP: the only way such modifiers can combine with them is as adjuncts $VP\backslash VP$, which are islands. However, (60b), shows that they also carry the argument category VP_{ing}, allowing extraction past subcategorizing verbs, as in "What tune is John whistling?". In the case of (60b), this implies a category for "drive" and related causatives like "make" of $((VP/VP_{ing})/XP_{pred_{idv}})/NP$, subcategorizing for VP_{ing} and allowing (60b) by composition.

These observations mean that when we talk of modifiers like "whistling *Dixie*" as "weak islands", we simply mean that they are lexically ambiguous. They are strong islands under their adjunct category, and non-islands under their argument category with suitable matrix verbs. This means of course that "John drives Mary crazy whistling *Dixie*" is ambiguous between a (preferred) argument reading in which it is specifically John's whistling *Dixie* that drives Mary crazy, and an adjunct reading analogous to "John works whistling *Dixie*", under which John merely happens to whistle that tune while he does whatever it is that actually drives her crazy.

The exact conditions under which weak island ambiguities are resolved in favor of the complement to permit extractions depend upon the matrix-verb's subcategorization, the parsing model, and/or world knowledge, essentially as proposed in neural-computational terms by Dowty (2003) and in event-semantic terms by Truswell, rather than upon syntax alone.

Kuno (1973b) noted that Japanese allowed certain relative clauses that appear to violate such island constraints. An example is the following:[21]

(61) [[katteita] inu-ga sinde simatta] kodomo
[[kept] dog-NOM die-PERF] child
"child$_i$ who the dog that (#he$_i$) kept died."

This possibility is not predicted by the analysis of Japanese relatives floated in section 9.5, in which relative clauses were hypothesised to be bare adjuncts N/N, and has been analysed extensively by Hasegawa (1985) and Richards (2001) in terms of very powerful movement theories that cannot be simulated in CCG terms.[22]

However, Kuno and Hasegawa point out that such examples are only allowed under the condition that the apparently island-violating extraction (here, headed by the child) is the subject, theme or topic of the inner clause, ("kept"), as it is here. Since Japanese is both topic-prominent and pro-drop, it seems possible that the subject long-range dependency in question is anaphorically mediated, as it is by a resumptive pronoun in the corresponding English given above. Indeed it seems possible that *all* Japanese relativization is mediated by pro-drop, hence an anaphoric rather than syntactic phenomenon. (Kuno's own "topic deletion" account is compatible with this proposal.)

9.8 Preposition Stranding and Relativization

The possibility exhibited by English and some Celtic and Scandinavian languages of extracting complements of prepositions, as in the following, is cross-linguistically extremely rare (van Riemsdijk, 1978):

21. We noted earlier that Japanese, being SOV, allows embedded subjects to extract, unlike English.
22. Richards (2000, 2001, 2002) relates such examples to the possibility of *in situ* multiple *wh* elements, which are briefly discussed in non-movement terms in section 9.1 above.

(62) a. Who did you buy the bicycle from?
b. I [sold a book to and bought a bicycle from] a very nice man,
c. I took [the bus to and the train from] 30th Street station.

No other Germanic or Latinate language, including Latin itself, strands adpositions with anything like this generality. Prepositions in such languages behave much like determiners, suggesting that in most languages they are case-like operators turning their complements into type-raised categories, which, being adjuncts, cannot be extracted out of. For example, the following seems to be the category of the Dutch preposition *op*, "on/in", as the specifier of a verbal argument:

(63) op := $PP^{\uparrow}_{op}/NP$

Like all the prepositions considered here, *op* has another category as the head of a sentential adverbial adjunct:

(64) op := $(VP\backslash VP)/NP$

as well as that of a particle in constructions like *opbellen* "ring up".

In English, by contrast, argument prepositions like "to" arguably do not bear any category related to (63) in alternation with the adjunct-heading category (64). Instead, they seem to bear ab adjunct-particle-like type-changing category like the following, restricted in the lexicon by the $\diamond\star$ slash-type.

(65) to := $(VP/_{\diamond\star}NP)\backslash_{\diamond\star}(VP/PP_{to})$

This category allows the following derivation:

(66)
$$\begin{array}{cccc}
\text{give} & \text{flowers} & \text{to} & \text{Henry}\\
\hline
(VP/PP_{to})/NP & (VP/PP)\backslash_{\diamond\star}((VP/PP)/NP) & (VP/NP)\backslash_{\diamond\star}(VP/PP_{to}) & VP\backslash_{\diamond\star}(VP/NP)\\
\cline{1-2}
\multicolumn{2}{c}{VP/PP_{to} \quad <} & & \\
\cline{1-3}
\multicolumn{3}{c}{VP/NP \quad <} & \\
\hline
\multicolumn{4}{c}{VP \quad <}
\end{array}$$

It also allows the following alternative derivation, among others, for the same sentence:

(67)
$$\begin{array}{cccc}
\text{give} & \text{flowers} & \text{to} & \text{Henry}\\
\hline
(VP/PP_{to})/NP & (VP/PP)\backslash_{\diamond\star}((VP/PP)/NP) & (VP/NP)\backslash_{\diamond\star}(VP/PP_{to}) & VP\backslash_{\diamond\star}(VP/NP)\\
\hline
\multicolumn{2}{c}{VP/PP_{to} \quad <} & \multicolumn{2}{c}{VP\backslash_{\diamond\star}(VP/PP_{to}) \quad <\mathbf{B}}\\
\hline
\multicolumn{4}{c}{VP \quad <}
\end{array}$$

Coordinate sentences like the following are thereby allowed as a form of argument-adjunct cluster coordination, introduced in Chapter 4.1:

(68) I will give flowers [to Henry and to Sam]$_{VP\backslash_{\diamond\star}(VP/PP_{to})}$.

The category (65) allows preposition-stranding extraction:

(69)
$$\begin{array}{ccccccc}
\text{man} & \text{who} & \text{I} & \text{will} & \text{give} & \text{that book} & \text{to} \\
\overline{N} & \overline{(N\backslash N)/(S/NP)} & \overline{NP^{\uparrow}} & \overline{(S\backslash NP)/VP} & \overline{(VP/PP_{to})/NP} & \overline{NP^{\uparrow}} & \overline{(VP/NP)\backslash_{\diamond\star}(VP/PP_{to})} \\
 & & & & \multicolumn{2}{c}{\underline{\hspace{6em}}_{<}} & \\
 & & & & \multicolumn{2}{c}{VP/PP_{to}} & \\
 & & & & \multicolumn{3}{c}{\underline{\hspace{12em}}_{<}} \\
 & & & & \multicolumn{3}{c}{VP/NP} \\
 & & & \multicolumn{4}{c}{\underline{\hspace{16em}}_{>\mathbf{B}}} \\
 & & & \multicolumn{4}{c}{(S\backslash NP)/NP} \\
 & & \multicolumn{5}{c}{\underline{\hspace{16em}}_{>\mathbf{B}}} \\
 & & \multicolumn{5}{c}{S/NP} \\
 & \multicolumn{6}{c}{\underline{\hspace{16em}}_{>}} \\
 & \multicolumn{6}{c}{N\backslash N} \\
\multicolumn{7}{c}{\underline{\hspace{16em}}_{<}} \\
\multicolumn{7}{c}{N}
\end{array}$$

(70)
$$\begin{array}{ccccccc}
\text{land} & \text{that} & \text{I} & \text{will} & \text{travel} & \text{to} & \text{tomorrow} \\
\overline{N} & \overline{(N\backslash N)/(S/NP)} & \overline{NP^{\uparrow}} & \overline{(S\backslash NP)/VP} & \overline{VP/PP_{to})} & \overline{(VP/NP)\backslash_{\diamond\star}(VP/PP_{to})} & \overline{VP\backslash VP} \\
 & & & & & \multicolumn{2}{c}{\underline{\hspace{12em}}_{\mathbf{B}_{\times}^{2}}} \\
 & & & & & \multicolumn{2}{c}{(VP/NP)\backslash_{\diamond\star}(VP/PP_{to})} \\
 & & & & \multicolumn{3}{c}{\underline{\hspace{14em}}_{<}} \\
 & & & & \multicolumn{3}{c}{VP/NP} \\
 & & & \multicolumn{4}{c}{\underline{\hspace{16em}}_{>\mathbf{B}}} \\
 & & & \multicolumn{4}{c}{(S\backslash NP)/NP} \\
 & & \multicolumn{5}{c}{\underline{\hspace{16em}}_{>\mathbf{B}}} \\
 & & \multicolumn{5}{c}{S/NP} \\
 & \multicolumn{6}{c}{\underline{\hspace{16em}}_{>}} \\
 & \multicolumn{6}{c}{N\backslash N} \\
\multicolumn{7}{c}{\underline{\hspace{16em}}_{<}} \\
\multicolumn{7}{c}{N}
\end{array}$$

As predicted by the earlier account of argument/adjunct cluster coordination in chapter 4.1, preposition-stranding sequences like "to tomorrow" can freely coordinate, as in the following example:

(71) A land that I will travel [to today and from tomorrow]$_{(VP/NP)\backslash_{\diamond\star}(VP/PP_{to})}$

However, the $\diamond$ modality on (65) prevents violation of the constraint that in the heyday of Constraints on Transformations Kuno (1973a) jokily entitled the "Clause Non-final Incomplete Constituent Constraint (CNFICC)" on preposition stranding out of the Heavy NP Shift construction analysed in example (9) of Chapter **??**:

(72)
$$\begin{array}{lccc}
{*}\text{a man who I will} & \text{give} & \text{to} & \text{that very heavy book} \\
 & \overline{(VP/PP_{to})/NP} & \overline{(VP/NP)\backslash_{\diamond\star}(VP/PP_{to})} & \\
 & \multicolumn{2}{c}{\underline{\hspace{14em}}_{<\mathbf{B}_{\times}}} & \\
 & \multicolumn{2}{c}{***} &
\end{array}$$

The modality that prevents preposition stranding of this category via crossed composition would also block Heavy NP Shift itself contrary to fact.. It follows that Heavy NP Shift over PP must involve the other, adverbial, category of PP and syntactically transitive *sent*, analogous to the derivation in (9), where the latter category semantically includes an argument that we write $sk_{recipient}$ which is anaphoric to something with the property of being a recipient, the property which the adjunct predicates of *Harry*:[23]

(73)
$$\begin{array}{cccc}
\text{I} & \text{sent} & \text{to Harry} & \text{a very heavy book} \\
\hline
NP^{\uparrow}_{1s} & (S\backslash NP)/NP & (S\backslash NP)\backslash(S\backslash NP) & NP^{\uparrow} \\
: \lambda p.p\ i & : \lambda x\lambda y.sent\ sk_{recipient}\ x\ y & : \lambda p\lambda y.recipient\ harry \wedge p\ y) & : \lambda p.p\ sk_{very\,heavy\,book} \\
\hline
 & \multicolumn{2}{c}{(S\backslash NP_{agr})/NP : \lambda x\lambda y.recipient\ harry \wedge sent\ sk_{recipient}\ x\ y \quad {<}\mathbf{B}_{\times}} & \\
\hline
 & \multicolumn{3}{c}{S\backslash NP_{agr}) : \lambda y.recipient\ harry \wedge sent\ sk_{recipient}\ sk_{very\,heavy\,book}\ y \quad >} \\
\hline
\multicolumn{3}{c}{S : recipient\ harry \wedge sent\ sk_{recipient}\ sk_{very\,heavy\,book}\ i) \quad >} &
\end{array}$$

The above analysis of stranding prepositions resembles a lexicalized version of Hornstein and Weinberg's and Kayne's (1981) "reanalysis" accounts of preposition stranding. The paradigm discussed in this section was also a major movation for Pesetsky's 1995:176 notion of "cascade" structure—see Phillips (1996, 2003) for an extended comparison between Pesetsky's approach and CCG. We shall see in chapter 11 further evidence in support of this account of English preposition stranding from adjunct/argument cluster coordination.[24]

9.9 On "Remnant Movement"

Both Heavy NP Shift and crossing dependencies in serial verb constructions have been analysed under the movement theory in terms of remnant movement—that is, the movement of constituents that already include a trace as a result of some other movement. The effect of crossed composition is to accept exactly those word orders that if derived from German- style verb-final embeddings to general the "English" orders available in Dutch and Hungarian, give the appearance of moving clauses with gaps in them.

23. We defer discussion of such terms until chapter 13. We could achieve the same effect with transitive semantics and a Davidsonian event variable, at the cost of some notational clutter.

24. Contrary to the prodictions of the above analysis, Pesetsky claims (1995:249) that Heavy Shift is clause bounded, claiming an asymmetry in acceptability with leftward extraction in examples like the following:

(i) a. What gift will Mary [claim that she gave to Harry]$_{(S\backslash NP)/NP}$ TOMORROW$_{(S\backslash NP)\backslash(S\backslash NP)}$?
   b. Mary [will claim that she gave to Harry]$_{(S\backslash NP)/NP}$ tomorrow$_{(S\backslash NP)\backslash(S\backslash NP)}$ a very heavy BOOK.

However, any such asymmetry seems to arise from the greater demands that heavy NP shift makes on the preceding context: (ib) seems perfectly acceptable as an answer to the question (ia).

Erzählen	wird	er	seiner Tochter	ein Märchen	können
Tell	will	he.NOM	his daughter.DAT	a story : ACC	be able
$S_{top}/(S/((VP\backslash NP_{dat})\backslash NP_{acc}))$	$(S/VP)/NP_{3s}$	$NP^{\uparrow}_{nom}$	$NP^{\uparrow}_{dat}$	$NP^{\uparrow}_{acc}$	$VP\backslash VP$

$$\dfrac{(S/VP)/NP_{3s} \quad NP^{\uparrow}_{nom}}{S/VP}<$$

$$\dfrac{NP^{\uparrow}_{dat} \quad NP^{\uparrow}_{acc}}{VP/((VP\backslash NP_{dat})\backslash NP_{acc})}>\mathbf{B}$$

$$\dfrac{VP/((VP\backslash NP_{dat})\backslash NP_{acc}) \quad VP\backslash VP}{VP/((VP\backslash NP_{dat})\backslash NP_{acc})}<\mathbf{B}_{\times}$$

$$\dfrac{S/VP \quad VP/((VP\backslash NP_{dat})\backslash NP_{acc})}{S/((VP\backslash NP_{dat})\backslash NP_{acc})}>\mathbf{B}$$

$$\dfrac{S_{top}/(S/((VP\backslash NP_{dat})\backslash NP_{acc})) \quad S/((VP\backslash NP_{dat})\backslash NP_{acc})}{S_{top}}>$$

"He will be able to tell his daughter a story"

Figure 9.3:

The availability of crossed composition is also both necessary and sufficient to support the analysis in figure 9.3 of “verb fronting” or verb topicalization in German, which has been held to provide major evidence for rules of remnant movement or “discontinuous constituency” (Reape 1994; Müller 1998; De Kuthy and Meurers 2001; see Müller 2007; Wechsler 2015:234-6 for discussion and a literature review).

This analysis, in which the ditransitive infinitival main verb, like all topics, receives a lexicalized exotypic order-changing type-raised function into topicalized sentences S_{top} (cf. English 18b), is equivalent to that of Nerbonne (1986) and Johnson (1986), which as they point out are essentially combinatory-categorial.[25]

The following asymmetry discussed by Johnson, which is left as an exercise, is also predicted:

(74) a. Erzählen können hat er seiner Tochter ein Märchen.
b. *Können hat er seiner Tochter ein Märchen erzählen.

9.10 *Tough*-movement

Consider the examples in the following alternation:

(75) a. To (imagine that we could) please John is easy
b. It is easy to (imagine that we could) please John.
c. John is easy to (imagine that we could) please.

The first two sentences (a,b) are just an example of the copular propositional subject construction and its alternation with extraposition or dislocation that was discussed in section 7.4:

(76) a. Being green isn’t easy/much fun/a cakewalk.
b. To err is human/embarassing/a pity
c. That they won is unfortunate/surprising/a triumph.

(77) a. It isn’t easy/much fun/a cakewalk being green.
b. It is human/embarassing/a pity to err.
c. It is unfortunate/surprising/a triumph that they won.

25. The significance of non-standard constituents like *seiner Tochter ein Mürchen*$_{VP/((VP\backslash NP_{dat})\backslash NP_{acc})}$ is discussed in section 11.4. Any corresponding verb topicalization in English **tell he will his daughter a story* is prevented by the same independent features of English grammar that both contribute to its rigid word-order and in general prevent subject extraction, namely $\diamond$ modality on complement taking verbs including the modals (*SP*:53-4).

The categories for *easy* are as follows:[26]

(78) a. easy := $XP_{pred_{idv}} : \lambda y.easy\,y$

b. easy := $VP_{pred_{idv}}/(VP_{to}/NP) : \lambda p\lambda x.easy\,(p\,x\,one)$

As in the case of the short passives discussed in the last chapter, the constant *one* is a placeholder representing an arbitrary agent whose semantic nature need not concern us here, but to which we will return in section 13.6.3:

The derivation of (75c) is then as in (79):

(79)
$$
\begin{array}{ccccc}
\text{John} & \text{is} & \text{easy} & \text{to} & \text{please.} \\
\hline
NP^{\uparrow} & (S\backslash NP)/VP_{pred} & VP_{pred}/(VP_{to}/NP_{-wh}) & VP_{to}/VP & VP/NP \\
:\lambda p.p\,john & :\lambda p\lambda y.p\,x & :\lambda p\lambda x.easy\,(p\,x\,someone) & :\lambda p.p & :\lambda x\lambda y.please\,x\,y \\
\hline
 & & & \multicolumn{2}{c}{>\mathbf{B}} \\
 & & & \multicolumn{2}{c}{VP_{to}/NP : \lambda x\lambda y.please\,x\,y} \\
\hline
 & & \multicolumn{3}{c}{>} \\
 & & \multicolumn{3}{c}{VP_{pred} : \lambda x.easy\,(please\,x\,someone)} \\
\hline
 & \multicolumn{4}{c}{>} \\
 & \multicolumn{4}{c}{S\backslash NP : \lambda y.easy\,(please\,y\,someone)} \\
\hline
\multicolumn{5}{c}{>} \\
\multicolumn{5}{c}{S : easy\,(please\,john\,someone)}
\end{array}
$$

This analysis in essence follows those of Carpenter (1992) and Jacobson (1992a), who point out that functional composition allows the infinitival transitive verb to be an unboundedly large fragment such as *to imagine that we could please*, accounting for the unbounded character of the dependency involved, while maintaining the Principle of Combinatory Projection.[27]

The limitation of the *tough* construction to nesting its dependency in the minimum pair shown in figure 9.4a,b (Chomsky, 1977b) follows immediately from the lexical analysis and the mechanism for extracting inner arguments (cf. figure 9.1a):

The intended reading with *sonatas played upon the violin* cannot be obtained from figure 9.3b, because the stranded preposition *upon* cannot combine with *are easy to play* until the relative pronoun has combined with the latter. However, that cannot happen until *violins* has combined, and the only way for *that* to happen is for the latter to have the category of a subject, forcing the unintended reading with *#the violin played upon sonatas*. Nor is there any other assignment of legal CCG categories that will allow the intended meaning to be derived

Thus, CCG offers a solution to the problem of an asymmetry which appears to remain open or stipulative in solution in other theories of grammar including

26. Further categories for *easy* are needed to capture the related sentences *It is easy for us to ... please John* and *John is easy for us to ... please*. See Jacobson 1992b for discussion.

27. Jacobson points out that the analysis has earlier antecedents in unpublished work by Gazdar, and in Chomsky 1977a.

a. (a) violin which sonatas are easy to play upon

violin	which	sonatas	are	easy	to	play	upon
N : $violin$	$((N\backslash N)/X)/((S/X)/NP)$: $\lambda p\lambda n\lambda x.n\,x\wedge p\,x$	$NP^{\uparrow}_{3pl}$: $\lambda p.p\,sonatas$	$(S\backslash NP_{3pl})/VP_{pred}$: $\lambda p\lambda y.p\,y$	$(VP_{pred}/X)/((VP_{to}/X)/NP_{-wh})$: $\lambda p\lambda x\lambda w.easy\,(p\,w\,x\,someone)$	VP_{to}/VP : $\lambda p.p$	$(VP/PP_{upon})/NP$: $\lambda x\lambda w\lambda y.play\_upon\,w\,x\,y$	$((S\backslash NP)/NP)\backslash_{\diamond\star}((S\backslash NP)/PP_{upon})$ $\lambda p\lambda x.instrument\,x\wedge p\,x$

$$(VP_{to}/PP_{upon})/NP : \lambda x\lambda w\lambda y.play\_upon\,w\,x\,y \quad {>}\mathbf{B}^2$$

$$(VP_{pred}/PP_{upon})/NP : \lambda x\lambda w.easy\,(play\_upon\,w\,x\,someone) \quad {>}\mathbf{B}^2$$

$$(S\backslash NP_{3pl})/PP_{upon} : \lambda w\lambda y.easy\,(play\_upon\,w\,y\,someone) \quad {>}\mathbf{B}^2$$

$$(S\backslash NP_{3pl})/NP : \lambda w\lambda y.instrument\,w\wedge easy\,(play\_upon\,w\,y\,someone) \quad {<}$$

$$S/NP : \lambda w.instrument\,w\wedge easy\,(play\_upon\,w\,sonats\,somone) \quad {>}\mathbf{B}$$

$$N\backslash N : \lambda n\lambda x.n\,x\wedge instrument\,x\wedge easy\,(play\_upon\,x\,sonatas\,somone) \quad {>}$$

$$N : \lambda x.violin\,x\wedge instrument\,x\wedge easy\,(play\_upon\,x\,sonatas\,somone) \quad {<}$$

b. #(a) sonata which violins are easy to play upon

sonata	which	violins	are	easy	to	play	upon
N : $sonata$	$((N\backslash N)/X)/((S/X)/NP)$: $\lambda p\lambda n\lambda x.n\,x\wedge p\,x$	$NP^{\uparrow}_{3pl}$: $\lambda p.p\,violins$	$(S\backslash NP_{3pl})/VP_{pred}$: $\lambda p\lambda y.p\,y$	$(VP_{pred}/X)/((VP_{to}/X)/NP_{-wh})$: $\lambda p\lambda x\lambda w.easy\,(p\,w\,x\,someone)$	VP_{to}/VP : $\lambda p.p$	$(VP/PP_{upon})/NP$: $\lambda x\lambda w\lambda y.play\_upon\,w\,x\,y$	$((S\backslash NP)/NP)\backslash_{\diamond\star}((S\backslash NP)/PP_{upon})$ $\lambda p\lambda x.instrument\,x\wedge p\,x$

$$(S\backslash NP_{3pl})/PP_{upon} : \lambda w\lambda y.easy\,(play\_upon\,w\,y\,someone)$$

$$\text{———}\;*$$

$$\text{———}\;*$$

Figure 9.4:

G/HPSG and MP (Chomsky, 1977b; Gazdar et al., 1985:150-153; Pollard and Sag, 1994:169; Hornstein, 2001; Hicks, 2009),

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

Following Morrill 1994, 1995 and Steedman 1987, *TS*:89–91 and *SS&I*:50–51 propose an extra lexical category for pied-piping *wh*-items such as *which* and *who(m)* in NPs such as *reports [the height of the lettering on the covers of which] the government prescribes*, which can under present assumptions concerning the type-raising of NPs be written as in (80):

(80) who(m), which := $((N\backslash N)/(S|NP))\backslash(NP^{\uparrow}/NP) : \lambda p\lambda q\lambda r\lambda x.q(px)\wedge rx$

Apart from the fact that relativizers are functors into $N\backslash N$, rather than S, this category is simply that of a type-raised NP raised over functions into type-raised NPs—that is, $NP^{\uparrow}\backslash((S/(S\backslash NP))/NP)$.

We know that composition can form constituents like *the government prescribes* and *the height of the lettering on the covers of*, since we can extract out of them:

(81) Reports which the government prescribes the height of the lettering on the covers of

So the category in (80) allows the following:

(82) [[Reports]$_N$ [the height of the lettering on the covers of]$_{NP^{\uparrow}/NP}$ [which]$_{((N\backslash N)/(S/NP))\backslash(NP^{\uparrow}/NP)}$ [the government prescribes]$_{S/NP}$]$_N$

The pied-piping *wh* category $NP^{\uparrow}\backslash(NP^{\uparrow}/NP)$ is also the type we need for the related category of in situ *wh*-items like *which reports* in “quiz show” questions like the following:

(83) The government prescribes [[the height of the lettering on the covers of]$_{NP^{\uparrow}/NP}$ [which reports]$_{NP^{\uparrow}\backslash(NP^{\uparrow}/NP)}$]$_{NP^{\uparrow}}$?

The availablity of a special-case “antecedent-controlled” subject extracting category (39) for bare-complement verbs like *know*, $(VP/NP_{+WH,agr})/(S\backslash NP_{agr})$, does not support *in situ* complement subject *wh*-items for those verbs. *In situ* *wh*-embedded subjects are in all cases just as unacceptable as subject extractions over complementizers like *that*:

(84) a. *Who believes which candidate will win the election?
b. *Who believes that which candidate will win the election?

The *in situ wh*-elements discussed above are well known for being immune in other respects to island constraints, as shown in the following comparison for English:

(85) a. #Which famous movie did Mary marry the man who directed?
b. Mary married [the$_{NP\uparrow/N}$ [man$_N$ [[who directed]$_{(N\backslash N)/NP}$ [which famous movie]$_{(N\backslash N)\backslash((N\backslash N)/NP)}$]$_{N\backslash N}$]$N$]$_{NP\uparrow}$?

The *in situ wh*-item can have a category type-raised over the type $(N\backslash N)/NP$ of *who directed*, and apply to it to yield the noun-modifier *who directed which musical* $N\backslash N$, which then combines in the standard way with the noun *man* and determiner *the* to yield a raised object NP, which can combine with the matrix in the standard way to yield *S*. (The question illocutionary force appears to stem from intonational accent on the *wh* element.)

We noted earlier in section 9.5 that the same immunity from islands is known to apply quite generally in languages like Japanese, in which all overt *wh*-elements are normally in situ (Kuno, 1973b). For example:

(86) Mary-wa [[John-ni nani-o ageta]$_{S'}$ hito-ni]$_{NP}$atta-no
Mary-TOP [[John-DAT what-ACC gave]$_{S'}$man-DAT]$_{NP}$met-Q
'#What did Mary meet the man who gave to John/Mary met the man who gave what to John??'

Choe (1987), Nishiguchi (1990), Pesetsky (1987), Watanabe (1992, 2001), Brody (1995), von Stechow (1996), and Richards (2000) have variously argued, against Huang (1982b) and Lasnik and Saito (1984), that in situ *wh*-constructions escape island constraints because of the possibility of covert "large-scale" pied-piping of entire phrases like *John-ni nani-o ageta hito-ni*, *the man who gave what to John*.

Although the present account of pied-piping does not involve syntactic movement, those arguments, and in particular the "across-the-board" condition on the scopes of multiple in situ *wh*-items inside islands discussed by Watanabe and Richards, appear to support it.

In particular, as in the case of the basic English relative pronoun (23), the pied-piping category (80), lexicalizes exactly the same insight as covert large-scale pied-piping movement, albeit statically at the level of lexical logical form, using only the standard variable-binding apparatus of the λ-calculus.

In the light of the homomorphic relation of prosody and intonation structure to syntax propsed in chapter 6, it should be clear at this point that the theory of relativization proposed in this chapter is closely related to the insight of Richards (2010, 2016) relating the scope of *wh* to a contiguous prosodic phrase. The only difference is that under the present theory this identity holds for both *wh in situ* and *wh*-movement, with combinatory derivation providing the domain for both, as well as for the formation of intonational phrases discussed there.

That is simply to say that the domain of *in-situ wh* and pied-piping *wh* are both, like ordinary pronominal relativization, defined by combinatory *constituency*. The only difference between the two forms of *wh*-construction and the bare forms of relativization considered in section 9.5 above is the language-specific lexical specification of wh-elements as either leftward- or rightward-combining.

9.12 Celtic relativization

Interestingly, in terms of the distinction within transformationalist theory from the "swoop" theory of unbounded movement over multiple tensed boundaries as a single operation of *Aspects* to the "cyclic" theory of unbounded dependencY as multiple local movements within successive tensed domains introduced by Chomsky (1973), the present non-movement theory is closest to a swoop theory, since the *wh*-element combines with the entire combinatory residue in a single merger.

The resemblance of CCG to a swoop, rather than a cyclic, account of movement might appear to conflict with the claims of McCloskey (1979); Adger (2003) and Boeckx (2008) to have proven on the basis of a number of constructions in a number of typologically distinct languages that movement is necessarily cyclic, and *not* swooping.

Adger 2003:376-386 provides a convenient summary of the argument based on *wh*-constructions in Scottish Gaelic, which is essentially parallel to McCloskey, 1979:150 for Irish, and to Welsh, seen earlier.

All *wh*-questions in Gaelic include an overt complementizer "a":

(87) Cò an duine **a** tha thu a' pòsadh
Who the man **that** are you ing- marry
"Which man are you marrying?"

When the *wh*-dependency is embedded, the "a" complementizer is obligatorily repeated:

(88) Cò an duine **a** bha thu ag ràdh *(**a**) bhuail i
Who the man **that** were you ing- say **that** hit she
"Which man were you saying that she hit?"

The "a" complementizer involved in long-range *wh*-constructions is in complementary distribution to the standard complementizer "gu(n)", which specifies the "dependent" form of its complement verb, here "do bhuail" ("hit"). The "gu(n)" complementizer acts as a barrier to extraction:

(89) *Cò an duine **a** bha thu ag ràdh **gun** do bhuail i
Who the man **that** were you ing- say **that** prt hit she
"Which man were you saying that she hit?"

In general, Adger notes the following patterns for embedded complements:

(90) a. $\ldots gu(n) \ldots gu(n) \ldots$ (complementation)
b. $*\ldots gu(n) \ldots a \ldots$ (*)
c. $\ldots a \ldots a \ldots t$ (*wh*-relativization)
d. $*\ldots a \ldots gu(n) \ldots t$ (*)

These authors claim that if the movement from "bhuail i" ("hit she") to the root "Có" were to take place as a unitary swoop, then we would have no way to prevent (89), since the movement would not be blocked by being unable to land at the illegal intermediate complementizer. Only if movement is cyclic can the moving element notice that its upward passage is blocked.

However, this claim is clearly theory-internal, and only makes sense on the assumption that unbounded dependency is literally mediated by a rule of displacement. We can assume instead that the relativizing complementizer "a" is a standard complementizer marking its clause as of the special type S_{rel} that was first encountered in the analysis of Welsh bare relative clauses in section 9.5, while "gu(n)" is like Northern German "daß" the specifier of a type-raised island complement, implying the following lexical categories:

(91) a $:= S_{rel}/S : \lambda s.s$
gu(n) $:= (S\backslash(S/S'))/S_{dep} : \lambda s\lambda p.ps$

On the assumption that *Wh*-questions in Gaelic depends upon the combination of a *wh*-element such as "Cò"of type $S_{whq}/(S_{rel}/NP)$ with a constituent of type S_{rel}/NP formed by combinatory composition, differing only from the corresponding English categories in specifying the relativized form in the residue, then the contrast between (88) and (89) can be captured without any explicitly cyclic operations or any other kind of action-at-a-distance:

(92)

$$
\begin{array}{cccccc}
\text{Cò } \textit{an duine} & \textbf{a} & \text{bha thu ag rà}\textit{dh} & \textbf{a} & \text{bhuail i} \\
\text{Who the man} & \textbf{that} & \text{were you ing- } \textit{say} & \textbf{that} & \text{hit she} \\
\overline{S_{whq}/(S_{rel}/NP)} & \overline{S_{rel}/_{\diamond\star}S} & \overline{S/_{\diamond\star}S_{rel}} & \overline{S_{rel}/_{\diamond\star}S} & \overline{S/NP} \\
 & \multicolumn{2}{c}{\overline{\qquad\qquad\qquad}{>}\mathbf{B}} & & \\
 & \multicolumn{2}{c}{S_{rel}/_{\diamond\star}S_{rel}} & & \\
 & \multicolumn{3}{c}{\overline{\qquad\qquad\qquad\qquad}{>}\mathbf{B}} & \\
 & \multicolumn{3}{c}{S_{rel}/_{\diamond\star}S} & \\
 & \multicolumn{4}{c}{\overline{\qquad\qquad\qquad\qquad\qquad}{>}\mathbf{B}} \\
 & \multicolumn{4}{c}{S_{rel}/NP} \\
\multicolumn{5}{c}{\overline{\qquad\qquad\qquad\qquad\qquad\qquad}{>}} \\
\multicolumn{5}{c}{S_{whq}}
\end{array}
$$

"Which man were you saying that she hit?"

(93)

$$
\begin{array}{ccccc}
{*}\text{Cò } \textit{an duine} & \textbf{a} & \text{bha thu ag rà}\textit{dh} & \textbf{gun} & \text{do bhuail i} \\
\text{Who the man} & \textbf{that} & \text{were you ing- } \textit{say} & \textbf{that} & \text{prt hit she} \\
\overline{S_{whq}/(S_{rel}/NP)} & \overline{S_{rel}/S} & \overline{S/S'} & \overline{(S\backslash(S/_{\diamond\star}S'))/S_{dep}} & \overline{S_{dep}/NP} \\
 & \multicolumn{2}{c}{\overline{\qquad\qquad\qquad}{>}\mathbf{B}} & \multicolumn{2}{c}{\textit{fcomp2}} \\
 & \multicolumn{2}{c}{S_{rel}/_{\diamond\star}S'} & \multicolumn{2}{c}{S\backslash(S/_{\diamond\star}S')} \\
 & \multicolumn{4}{c}{\overline{\qquad\qquad\qquad\qquad\qquad\qquad}{*}}
\end{array}
$$

"Which man were you saying that she hit?"

Relative clauses in Scots Gaelic also involve the relative complementizer "a" and exclude "gu(n)". However, as for other Celtic languages, all authorities insist that Gaelic, Irish, and Welsh "a" is not a relative pronoun, parallel to English *wh*, but a complementizer (McCloskey, 1979; Gillies, 1993; Borsley et al., 2007).

We therefore make the same assumption as for Welsh, that one of the NP complements of the matrix verb is the head of the relative clause of type $N\backslash N$, with the additional assumption that the Scots "a" complementizer has a further category mapping that category to itself

(94) $\text{a} := S_{rel}/S : \lambda s.s$
$\quad := (N\backslash N)/(N\backslash N) : \lambda p\lambda n.pn$

We then have the following pattern of derivations for Gaelic long-range relativization (90c):[28]

28. It will be clear from the derivation why it is frequently possible to elide "a".

(95)

$$
\begin{array}{ccccccc}
\text{duine} & \textbf{a} & \text{thuirt} & \text{e} & \textbf{a} & \text{bhuail} & \text{i} \\
\text{man} & \textbf{that} & \text{said} & \text{he} & \textbf{that} & \text{hit} & \text{she} \\
\hline
N & (N\backslash N)/(N\backslash N) & (S/S')/NP & ((N\backslash N)/(S_{rel}/NP))\backslash((S/S')/NP) & S_{rel}/S & (S/NP)/NP & NP^{\uparrow} \\
: man & : \lambda p\lambda n.p\,n & : \lambda y\lambda s.past\,(said\,s\,y) & : \lambda q\lambda p\lambda n\lambda x.n\,x \wedge q\,(p\,x)\,him & : \lambda s.s & : \lambda y\lambda x.past\,(hit\,x\,y) & \lambda p.p\,her
\end{array}
$$

$$(N\backslash N)/(S_{rel}/NP) : \lambda p\lambda n\lambda x.n\,x \wedge past\,(said\,(p\,x)\,him) \quad <$$

$$S/NP : \lambda x.past\,(hit\,x\,her) \quad <$$

$$S_{rel}/NP : \lambda x.past\,(hit\,x\,her) \quad >\mathbf{B}$$

$$N\backslash N : \lambda n\lambda x.n\,x \wedge past\,(said\,(past\,(hit\,x\,her))\,him) \quad >$$

$$N\backslash N : \lambda n\lambda x.n\,x \wedge past\,(said\,(past\,(hit\,x\,her)\,him)) \quad >$$

$$N : \lambda x.man\,x \wedge past\,(said\,(past\,(hit\,x\,her)\,him)) \quad <$$

"man that he said that she hit"

Note that "thuirt e" ("said he") in the above carries the same category as a matrix English object relative pronoun. Embedded verbs like "bhuail" carry the standard independent S type.)

By contrast, "gu(n)" acts as an island-inducing type-raised argument head, requiring a dependent S_{dep} complement:[29]

(96)

$$
\begin{array}{cccccc}
\text{Thuirt} & \text{e} & \text{gun} & \text{do bhuail} & \text{i} & \text{e} \\
\text{Said} & \text{he} & \text{that} & \text{hit} & \text{she} & \text{him} \\
\hline
(S/S')/NP & NP^{\uparrow} & (S\backslash(S/S'))/S_{dep} & (S_{dep}/NP)/NP & NP^{\uparrow} & NP^{\uparrow}
\end{array}
$$

$$S/S' \quad < \qquad S_{dep}/NP \quad <$$

$$S_{dep} \quad <$$

$$S\backslash(S/S') \quad >$$

$$S \quad <$$

"He said that she hit him"

However, if the relative complementizer "a" intervenes no similar derivation of a complete complement is possible:

(97)

$$
\begin{array}{cccccc}
{*}\text{Thuirt} & \text{e} & \text{a} & \text{bhuail} & \text{i} & \text{e} \\
\text{Said} & \text{he} & \text{that} & \text{hit} & \text{she} & \text{him} \\
\hline
(S/S')/NP & NP^{\uparrow} & S_{rel}/S & (S/NP)/NP & NP^{\uparrow} & NP^{\uparrow}
\end{array}
$$

$$S/S' \quad < \qquad S/NP \quad <$$

$$S_{rel}/NP \quad >$$

$$*** \quad !!!$$

Similarly, embedded island barrier "gu(n)" blocks the formation of "said he that hit she", the argument of long range relativizer "a":

29. Semantics is suppressed as similar to English.

(98)
$$
\begin{array}{ccccccc}
*\text{duine} & \textbf{a} & \text{thuirt} & \text{e} & \textbf{gun} & \text{do bhuail} & \text{i} \\
\text{man} & \textbf{that} & \text{said} & \text{he} & \textbf{that} & \text{hit} & \text{she} \\
\hline
N & (N\backslash N)/(N\backslash N) & \multicolumn{2}{c}{S/S' \textbf{ or } S_{rel}/S_{rel}} \; < & (S\backslash(S/S'))/S_{dep} & (S_{dep}/NP)/NP & NP^{\uparrow} \\
 & & & & & \multicolumn{2}{c}{\overline{\quad S_{dep}/NP \quad}<} \\
 & & & & \multicolumn{3}{c}{\overline{\quad (S\backslash(S/S'))/NP \quad}>\mathbf{B}} \\
 & \multicolumn{5}{c}{\overline{\qquad\qquad * * * \qquad\qquad}!!!} &
\end{array}
$$

Like any island effect, movement is blocked because the residue of relativization cannot form in the first place. The mover "a" has no need to "notice" why not.[30]

9.13 Discussion

The combinatory treatment of the relative clause proposed here is syntactically mediated solely by combinatory reductions operating on adjacent typed constituents. CCG can therefore be seen as reducing the transformationalists' operation MOVE to (external) MERGE.

The long-range dependency at the heart of the construction is established via the lexical logical form $\lambda q\lambda n\lambda y.qy \wedge ny$ of the relative pronoun (23), and in particular by the use there of the second-order variable q.

One might also view this second-order λ-term as formalizing some version of the "copy theory of movement," with the λ-bound variables doing the work of "copying" declaratively, at the level of static logical form. In that sense one could also see CCG as reducing "overt" syntactic movement to "covert" LF movement, as does the "Trace Conversion Rule" of Fox (2002).

However, there is no process of copying as such. This is simply the standard apparatus of binding a value to a variable that occurs, possible more than once, in a logical formula. The relation of the head noun and subordinate clause is defined once and for all, in the lexical logical form, which the combinatory syntax merely projects by adjacent merger onto sentential logical form. Any constraints on possible projections can only arise from the combinatorics of syntactic projection.

Exercise The pattern in Irish is similar to Gaelic, but complicated by the possibility of resumptive as well as *wh*-relativization. McCloskey (2002:193-200) identifies the following more complex pattern, of which the first three are the most frequent: (See also McCloskey, 2017.)

30. A related analysis to that of Gaelic above appears to be applicable to the Germanic "*wh*-copying" phenomenon discussed by Felser (2004).

(99) a. $\ldots go \ldots go \ldots$ (complementation)
b. $\ldots a^L \ldots a^L \ldots t$ (*wh*-relativization)
c. $\ldots a^N \ldots go \ldots pro$ (resumptive-relativization)
d. $\ldots a^N \ldots a^N \ldots pro$ (resumptive-relativization)
e. $\ldots a^L \ldots a^N \ldots pro$ (rarely, resumptive-relativization)
f. $\ldots a^N \ldots a^L \ldots t$ (rarely, *wh*-relativization)

Try to extend the analysis of Scots Gaelic relatives in section 9.12 to Irish.

Chapter 10

Parasitic *Wh*-dependencies

The Introduction noted that sentences involving parasitic gaps or dependencies of a single relativized *wh*-element on more than one verb presented a problem for all movement-based accounts of long-range dependencies. The example there was (25), repeated here:

(1) 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 second *wh*-dependency on "reading" in (1a) is referred to as "parasitic" because, unlike the normal dependency in (1b) it is not allowed on its own, because it is in an adjunct island.

10.1 Combining Categories III: Substitution

In terms of the Combinatory Logic of Curry and Feys (1958), they tell us that the combinatory rules must inclde a "duplicator", of which the simplest examples are W and S. Szabolcsi (1983, 1989) was the first to propose combinatory rules of the following form, which, like the composition rules (6), come in harmonic and crossing forms, with corresponding slash-types that may limit their applicability to any given category (Steedman, 1987, SS&I).

(2) MERGE IIIA: THE SUBSTITUTION RULES

a. Forward Substition:

$$(X/_{\diamond}Y)/Z : f \quad Y/Z : g \quad \Rightarrow_{\mathsf{S}} \quad X/Z : \lambda z.f\,z(g\,z) \qquad (>\mathsf{S}\;)$$

b. Backward Substition:

$$Y\backslash Z : g \quad (X\backslash_{\diamond}Y)\backslash Z : f \quad \Rightarrow_{\mathsf{S}} \quad X\backslash Z : \lambda z.f\,z(g\,z) \qquad (<\mathsf{S}\;)$$

c. Forward Crossing Substition:

$$(X/_{\times}Y)\backslash Z : f \quad Y\backslash Z : g \quad \Rightarrow_{\mathsf{S}} \quad X\backslash Z : \lambda z.f\,z(g\,z) \qquad (>\mathsf{S}_{\times})$$

d. Backward Crossing Substition:

$$Y/Z : g \quad (X\backslash_{\times}Y)/Z : f \quad \Rightarrow_{\mathsf{S}} \quad X/Z : \lambda z.f\,z(g\,z) \qquad (<\mathsf{S}_{\times})$$

As in the case of composition, further rules generalize substitution to higher valency dependent functors $(Y/Z)|W$ and $(Y\backslash Z)|W$.

For example, we shall occasionally need the following instance of second-level substitution:[1]

1. Our use of the S^n notation is different to that of Curry and Feys (1958) but is consitent with the $\mathbf{B}^n$ convention.

(3) Backward Crossing Second-level Substition:

$$(Y/W)/Z : g \quad (X\backslash_{\times}Y)/Z : f \quad \Rightarrow \quad (X/W)/Z : \lambda z \lambda w. f z (g z w) \quad (<\mathbf{S}^2_{\times})$$

The Combinatory Projection Principle (5) of chapter 2 forbids substition rules like the following:

(4)
$$\begin{array}{llll} (X\backslash Y)/Z & Y/Z & \not\Rightarrow & X/Z \\ Y\backslash Z & (X\backslash Y)/Z & \not\Rightarrow & X\backslash Z \\ (X/Y)/Z & Y/Z & \not\Rightarrow & X\backslash Z \end{array}$$

10.1.1 Parasitic Gaps in Adjuncts

The backward crossing substitution rule (2d) allows the example (25) from chapter **??** and the related relative construction in figure 10.1a.

It should be noted that the **S**-combinatory rule interacts correctly with the entirely independent lexicalized apparatus of what is sometimes called "adjunct control" described in section **??** to make the agent of *filing* the same as that of *reading*. However, as with verbal control, this fact follows from the semantics of adverbial adjunction rather than from anything related to movement, anaphora, or meaning postulates external to the grammar.

Examples like (5) also require the backward crossing substitution rule (2d), and provide further evidence both for the mechanism for extracting nonperipheral arguments exemplified in the earlier example of heavy NP shift (9), and for the existence in the grammar of constituents like *throw in the trash*$_{VP/NP}$, derived by composing $VP\backslash VP$ with the verb by the backward crossing composition rule.

(5) articles which I will [throw in the trash]$_{VP/NP}$ [without reading]$_{(VP\backslash VP)/NP}$

The inclusion of second-level substitution rules like (3) allows derivations like figure 10.1b

It is easy to see that "stacked" parasitic gaps, as in the following multiply ambiguous examples, are accepted as well:

(6) a. articles which I will file without reading in order to evaluate
b. articles which I will file without reading in order to evaluate before burning

The theory moreover captures the fact that extraction obeys exactly the same Subjacency and ECP-related constraints within the adjunct as it does everywhere else.[2]

2. In paricular, we avoid the need in GPSG/HPSG to invoke a "Head Feature Convention" to

a. articles that I shall file without reading

$$
\begin{array}{llllll}
\text{articles} & \text{that} & \text{I shall} & \text{file} & \text{without} & \text{reading} \\
\hline
 & (N\backslash N)/(S/NP) & S/VP & VP/NP & (VP\backslash VP)/VP_{ing} & VP_{ing}/NP \\
 & \lambda p\lambda n\lambda x.p\,x \wedge n\,x & : \lambda p.shall\,(p\,me) & \lambda x\lambda y.file\,x\,y & \lambda p\lambda q\lambda y.without\,(p\,y)\,(q\,y) & \lambda x\lambda y.read\,x\,y
\end{array}
$$

$$\xrightarrow{>\mathbf{B}}\quad (VP\backslash VP)/NP : \lambda x\lambda q\lambda y.without\,(read\,x\,y)\,(q\,y)$$

$$\xrightarrow{<\mathbf{S}_{\times}}\quad VP/NP : \lambda x\lambda y.without\,(read\,x\,y)\,(file\,x\,y)$$

$$\xrightarrow{>\mathbf{B}}\quad S/NP : \lambda x.without\,(read\,x\,me)\,(file\,x\,me)$$

$$\xrightarrow{>}\quad N\backslash N : \lambda x.without\,(read\,x\,me)\,(file\,x\,me) \wedge n\,x$$

b. people that I shall persuade without meeting to vote for me

$$
\begin{array}{llllll}
\text{people} & \text{that} & \text{I shall} & \text{persuade} & \text{without meeting} & \text{to vote for me} \\
\hline
N & (N\backslash N)/(S/NP) & \xrightarrow{>\mathbf{B}} S/VP & (VP/VP_{to})/NP & (VP\backslash VP)/NP & \xrightarrow{<\mathbf{T}} VP\backslash(VP/VP_{to})
\end{array}
$$

$$\xrightarrow{<\mathbf{S}^2_{\times}}\quad (VP/VP_{to})/NP$$

$$\xrightarrow{<\mathbf{B}_{\times}}\quad VP/NP$$

$$\xrightarrow{>\mathbf{B}}\quad S/NP$$

$$\xrightarrow{>}\quad N\backslash N$$

$$\xrightarrow{>}\quad N$$

Figure 10.1:

The impossibility under the preceding account of multiple *wh*-queries of a parasitic gapping reading of the following example, noted by Nunes (2004:107) as presenting difficulties for standard minimalist approaches, is also predicted (cf. (12) in chapter 9):

(7)
$$\begin{array}{cccc}
\text{*Who} & \text{filed} & \text{which papers} & \text{without reading} \\
\hline
S_{whq}/(S\backslash NP) & (S\backslash NP)/NP & S_{whq}\backslash(S_{whq}/NP) & ((S\backslash NP)\backslash(S\backslash NP))/NP \\
\cline{2-4}
 & \multicolumn{3}{r}{***}
\end{array}$$

The Combinatory Projection Principle (5) of chapter 2, and in particular the fact that all combinatory rules apply to string-adjacent types, means that there is no possible derivation for the above sequence of categories. Any different choice of category for the *in situ* object *which papers* would overgenerate, as shown in section 9.1.

On the other hand, *in situ wh*-examples like the following are correctly predicted to have a parasitic reading (cf. example (12) of the previous chapter):

(8)
$$\begin{array}{cccc}
\text{Who} & \text{filed} & \text{without reading} & \text{which papers?} \\
\hline
S_{whq}/(S\backslash NP) & (S\backslash NP)/NP & ((S\backslash NP)\backslash(S\backslash NP))/NP & S_{whq}\backslash(S_{whq}/NP) \\
\cline{2-3}
 & \multicolumn{2}{c}{S/NP \quad {<}\mathbf{S}_{\times}} & \\
\cline{1-3}
\multicolumn{3}{c}{S_{whq}/NP \quad {>}\mathbf{B}} & \\
\cline{2-4}
 & \multicolumn{3}{c}{S \quad {<}}
\end{array}$$

The treatment in the previous chapter of stranding prepositions as particles means that a derivation involving the forward rule (2a) following example is blocked, because the prepositions *to* and *about* have the particle-like category $(VP/NP)\backslash_{\diamond\star}(VP/PP)$ of section 9.8, rather than PP/NP.

(9) ?People who I will talk to about

(That they do have the stranding category is evidenced by the possibility of *Who did you talk to Mary about* and *Who did you talk to about Mary?*.)

However, we have seen that in order to support heavy NP shift, as in (73) of the previous chapter, such prepositions must also have the following adjunct-heading category:

(10) about := $(VP\backslash VP)/NP : \lambda x \lambda q \lambda y.about\, x\,(p\,y)$

As a result, backward crossed substitution will potentially allow (9) under a derivation parallel to figure (10.1)b, as do many other accounts of parasitic

ensure that adjuncts are islands to non-parasitic extraction (Gazdar et al., 1985:164; Pollard and Sag, 1994:183).

extraction, including HPSG and sideward movement.

The status of (9) is unclear. Gazdar et al. note that some speakers accept it, Engdahl notes that certain related sentences are acceptable in Swedish. We will leave the question open here.

10.1.2 Subjects and Adjunct Parasitic Gaps

Engdahl (1983, (54)–(56)) points out that subjects do not in general support parasitic dependencies, offering examples similar to the following:

(11) a. *a man who [painted]$_{(S\backslash NP)/NP}$ [a picture of]$_{NP/NP}$
b. *a man who [remembered]$_{(S\backslash NP)/VP_{ing}}$ [talking to]$_{VP_{ing}/NP}$
c. *a man who [remembered]$_{(S\backslash NP)/S'}$ [that John talked to]$_{S'/NP}$

Chomsky (1982; 1986a:55) ascribes the unacceptability of such examples to an Anti-c-command Condition. However, as Koster (1986:346) points out, these and many other examples involving subjects are considerably worse than the earlier examples involving violations of this comparatively weak condition. Steedman 1987 noted that in CCG all three examples are excluded by the Combinatory Projection Principle (5) of chapter 2, and more specifically by the underlying universal Principles of Directional Type Consistency and Inheritance, without further stipulation or reference to structural properties.

There is in fact no possible combinatory rule, whether corresponding to **S** or any other combinator, that will permit (11a–c). All such putative rules would violate the Principle of Directional Inheritance, by equating $/NP$ with $\backslash NP$. Such examples are therefore universally excluded for any SVO language on grounds of incompatibility with the lexical definition of its clausal word-order.

The same holds for examples like the following (Nunes, 2004:109):

(12)
$$\begin{array}{ccc}
*\text{which man} & \text{called you} & \text{before you met} \\
\hline
S_{whq}/(S\backslash NP) & S\backslash NP & (S\backslash S)/NP \\
 & \multicolumn{2}{c}{\overline{\hphantom{xxxxxxxxxxxxxxxx}}^{\;***}}
\end{array}$$

Engdahl (1983) also points out that, by contrast with (11), extractable *embedded* subjects *can* take part in parasitic constructions. She gives the following example, which originates with Alan Prince:

(13) A person whom$_i$ Brutus will manage to imply$_i$ is no good while ostensibly praising$_i$

Example (13) has a nonsubject gap parasitic on an embedded subject gap.

Unlike the earlier cases, examples like (13) *are* allowed by the present theory, because the subject-extracted predicate *imply was no good* can be built by directly applying the special subject-extracting category $(VP/NP_{+wh})/(S\backslash NP)$ of the verb (see (39)) with the complement predicate category $S\backslash NP$. The resulting category VP/NP_{+wh} and the adjunct function *while ostensibly praising*, which has the category $(VP\backslash VP)/NP$ (built by the forward composition rule) are of the appropriate form and linear order for the familiar backward instance of the substitution rule to apply, thus:

(14)
$$\begin{array}{ccccc}
\text{who(m)} & \text{Brutus will} & \text{imply} & \text{was no good} & \text{while praising} \\
\hline
(N_{agr}\backslash N_{agr})/(S/NP_{agr}) & S/VP & (VP/NP_{+WH,agr})/(S\backslash NP_{agr}) & S\backslash NP_{3s} & (VP\backslash VP)/NP \\
\cline{3-4}
 & & \multicolumn{2}{c}{VP/NP_{+WH,3s} \quad >} & \\
\cline{3-5}
 & & \multicolumn{3}{c}{VP/NP_{+WH,3s} \quad <\mathbf{S}_{\times}} \\
\cline{2-4}
 & \multicolumn{3}{c}{S/NP_{+WH,3s} \quad >\mathbf{B}} & \\
\cline{1-3}
\multicolumn{3}{c}{N_{3s}\backslash N_{3s} \quad >} & &
\end{array}$$

This analysis provides additional evidence that English extractable subjects are lexically specified as *rightward* NP arguments of matrix bare-complement verbs like *imply*. If they were leftward arguments, like true subjects, no analysis at all would be permitted under the present theory, for the same reason that excluded (12), because it would require a combinatory rule violating the Combinatory Projection Principle (CPP). (Again, related examples involving the "adjunct effect" on subject extraction discussed in section 9.6, such as *a man whom Brutus will imply that in his opinion is no good, while ostensibly praising*, seem to show a similar effect and provide further support for the earlier analysis of such subjects as rightward $+WH$ arguments.)

Finally, it follows from this analysis that *rightward* movement out of this parasitic construction will be forbidden, like rightward movement of a simple embedded extractable subject, because right-node-raised subjects are lexical, and hence $-WH$.

(15) *Brutus implied was no good while ostensibly praising the man in the Brooks Brothers shirt.

It also follows that the related ill-formed sentence (16), in which the subject gap is not embedded, is (like other examples relating to nominative island constraints) prevented by the directionality of the argument of the predicate category $S\backslash NP$:

(16) *a man [who]$_{(N\backslash N)/(S\backslash NP)}$ [was absolutely useless]$_{S\backslash NP}$ [despite Brutus extravagantly praising]$_{(S\backslash S)/NP}$

The omnivorous VP-adjunct category appears here as $S\backslash S$ to make the point that *even if the adverbial were an S adjunct*, it still could not combine. The point here is that the impossibility of this combination is not language specific. The CPP means not only that there is no instance in English of substitution that will equate the arguments of a forward-combining and a backward-combining function, but also that no such instances are permitted by Universal Grammar, because the NP arguments have different directionality.

A similar asymmetry holds for parasitic subject gaps in phrases like *burn t without realizing t_p was valuable*. These would be allowed on the assumption that the embedded subject gap gives rise to $(VP\backslash VP)/NP_{+wh}$. But a *non*embedded subject gap in the adjunct is impossible, not only because it would require slash-crossing forward composition in English, but also because the CPP means that no instance of substitution that could combine the result is permitted for an SVO language by Universal Grammar:[3]

(17) (Who(m) did you) *[meet]$_{VP/NP}$ [before]$_{(VP\backslash VP)/S}$ [left?]$_{S\backslash NP}$

10.1.3 (Non)-parasitism of adjuncts including some PPs

Adjuncts, including adjunct *wh*-items, are not arguments, so they cannot undergo parasitic extraction via the **S**-combinatory rules.

(18) *How$_i$ did you mend$_i$ the car without fixing$_i$ the truck?

In particular, *wh*-items like *how* bear the category S_{whq}/S_{inv} so they cannot in general relate to embedded clauses:[4]

(19) *How$_i$ do you wonder whether he mended$_i$ the car?

Opinions seem to be divided on whether PPs can ever undergo parasitic extraction (Postal, 1993; Grover, 1995). Since some PPs are subcategorized for, the present theory predicts that the following will be equally acceptable:

3. Of course, relatives that in Minimalist terms involve "adjunct control" mediated by sideward movement or a PRO subject, like ... *who t_i was absolutely useless despite PRO$_i$ extravagantly praising Brutus*, are accepted by the present grammar. However, these examples do not involve rules related to **S** at all.

4. SP discusses the mechanism by which adjuncts extract exceptionally from bare complements of "bridging" verbs like *think* and *say*.

(20) a. To who(m) did you lend the car?

b.

$$\begin{array}{cccc}
\text{To who(m)} & \text{did you} & \text{lend the car} & \text{without giving the keys?}\\
\hline
S_{whq}/(S_{inv}/PP) & S_{inv}/VP & VP/PP & (VP\backslash VP)/PP \\
\cline{3-4}
 & & \multicolumn{2}{c}{VP/PP \quad {<}\mathbf{S}_{\times}} \\
\cline{2-4}
 & \multicolumn{3}{c}{S_{inv}/PP \quad {>}\mathbf{B}} \\
\cline{1-3}
\multicolumn{3}{c}{S_{whq} \quad {>}} &
\end{array}$$

Nunes (Nunes, 2001, 2004:98-103) has recently proposed a Minimalist analysis in which the *wh*-phrase in (1a) is initially generated *in situ* as an unattached adjunct *reading which articles*, and is then "sideward-moved" by copying to the object position of an (equally unattached) verb *file* to create a VP *file which articles*. These structures separately undergo further mergers to form structures respectively corresponding to the modifier *without reading which articles* and the vP *you file which articles*. The modifier then adjoins to the vP, and (after a number of further standard derivational steps) a further copy of *which article* is created, and (via considerations that need not concern us here) the two lower copies are finally deleted to yield the structure for (25a).[5]

It is interesting to note that sideward movement under this definitions is an instance of what was in *Syntactic Structures* called a "generalized" or "double-base" transformation, recently revived by Chomsky (1993; 1995b:188-9). As was pointed out in SS&I, the syntactic rules of Composition (6) and Substitution (2) are akin to generalized transformations, but differ in being conditioned on contiguous types, rather than on pairs of structural analyses. It is in part the type-dependence of Combinatory Rules, together with the Combinatory Projection Principle (5) of Chapter 2, that makes them considerably less expressive than structure-dependent generalized transformational rules, in a sense to be discussed in more detail in the appendix.

Sideward movement is in fact closely related to the **S**-combinatory rules, apart from being stated as a structure-dependent transformational operation over complete phasal structures, rather than as applying under the control of linearly adjacent constituent types.

Besides applying Sideward A-Movement to the analysis of parasitic extraction, Nunes (2001); Hornstein (2003); Stabler (2006), and Boeckx et al. (2010) analyse the binding of adjunct *subjects*, including those incidentally involved in parasitic gap constructions, which they refer to as "adjunct control". We saw in section **??** in connection with example (**??**) that the binding of adjunct

5. Stabler (2006:141) formalizes a version of Sideward A-Movement in a manner somewhat similar to the **S**-combinatory rules (2).

subjects to the matrix subject can be straightforwardly defined in the lexical logical forms of adjunct heads like *before*, rather than by the present equivalent of A-movement.

10.2 Discussion

It will be apparent at this point that

all "overt" movement, including the "remnant" and "sideward" varieties, are reduced to contiguous "external" Merge by the ability of the combinatory rules to assign a constituent type to the residue of movement, including the multiple gaps that arise from parasitic extraction and adjunct control. All "covert movement", including the "pied-piping" component of overt *wh*-movement, and *in-situ wh*-items, are also reduced to Merge in the same way, applying to the same variety of combinatory constituent.

When we come to the discussion of intonation prosody in a laterchapter 6, we shall see that the domain of the phonological phrase is also defined by the same notion of combinatory constituency. Our account of *in-situ wh* can therefore be seen as consistent with the insight of Richards (2010, 2016) that the domain of *in-situ wh* is the contiguous phonological phrase. The difference is that under the present account, the domain of *both in situ and* fronted *wh* is the phonological phrase, as defined by combinatory constituency.

The combinatory nature of the rules immediately ensures that the apparently non-adjacent dependencies that it allows are subject to the *c*-command condition, for the same reason that the adjacent ones are, namely that the rules apply to adjacent typed constituents, Thus, *all* dependencies are syntactically adjacent, or in the terms of dependency grammar, "projective".

Exercise : The above account assumes that leftward and rightward extraction are mediated by the same mechanism, and are therefore symmetrical. Is this true? (Hint: the pragmatic demands of rightward extraction are strong, so you will need to think about controling context, say by including preceding context-setting questions.

Chapter 11

Symmetry and Asymmetry in Left and Right Extraction

The possibility of conjunction offers one of the best criteria for the initial determination of phrase structure.
—*Syntactic Structure* Noam Chomsky, 1957:36

The reduction of both *wh*-extraction and right-node–raising of an element X to contiguous combination or merger depends in general on the possibility of making the residue of both into a unbounded constituent of type $S|X$ by identical processes of function composition, with or without coordination of like types. It follows that the present theory makes a broad prediction of symmetry for left- and right-extraction in English: whatever can undergo *wh*-extraction from the periphery of a typable constituent can potentially right-node–raise from it, and vice versa. subject to the same constraints. Conversely, if an element cannot right node raise, then it should not be able to left extract. A number of instances of this tendency were noted above, including that of extraction from English embedded *that*-complements, where embedded objects can both *wh*-extract and right-node raise, with identical residues of type S/NP, but where subjects of sentential complements can in general extract neither to the left nor to the right.[1]

However, there is a difference. English leftward-moving categories are special-purpose non-order-preserving exotypic categories, distinct from the order-preserving cased rightward movers, permitting exceptions to symmetry, all of which must under present assumptions be specified in the English lexicon, and as such be expected to vary across other languages. For example, we also saw in section 9.6 that English subjects of bare complements lacking a complementizer can wh-extract, but where there is no corresponding in situ argument category that would permit right-node raising or Heavy NP-shift:

1. The use of a single mechanism in leftward and rightward extraction was a feature of the first version of GPSG (Gazdar, 1981). This very attractive feature of the theory was almost immediately abandoned for theory-internal reasons that do not apply to the present theory under the Projective Dependency Principle (PDP) (1) of chapter 9.

(1) a. Who did they say won the election?
b. *They said won the election the heaviest candidate.

The following are some further cases where symmetry might be expected, but where exceptions to symmetry have been claimed.

11.1 Across-the-Board Extraction

Non-traditional unbounded constituents of the type S/NP that have most recently been encountered at length in chapter 9.4 as the type of the residue of relativization also showed up in an earlier chapter as the residue of rightward extraction, or right-node raising, as in (2a), whose derivation is shown in figure 11.1.

(2) Freeman admires, and Hardy says Willis detests, sincerity/#it.

There are two further points to note about the above example and the derivation in figure 11.1. First, the derivation crucially depends on the availability of composition merge and type-raising. Second, such right-node-raised objects must be grapho-phonologically "heavy" and "rhematic" or discourse-new.

For example, the pronoun alternative in (2), which is phonologically "light" and by definition discourse-given, is unacceptable in this construction. We can capture this fact by giving English pronouns only the "clitic" accusative category $(S\backslash NP)\backslash((S\backslash NP)/NP)$ and excluding $S/(S/NP)$ for them. Thus pronouns can only take part in traditional transitive derivations like (21) of chapter 4.1, and not in the nonstandard derivation (22) there.[2]

As in the GPSG account of Gazdar 1981, ã type-dependent account of extraction and coordination, as opposed to standard accounts using structure-dependent rules, makes the across-the-board condition (ATB) on extractions from coordinate structure a prediction for both lefteard and rightward extraction, without any additional stipulation of parallel structure conditions (Goodall, 1987) or syntactic multidominance (McCawley, 1982, 1987; Citko, 2017), as the following examples reveal:

2. This claim is borne out by the fact that pronouns generally fail to attract nuclear pitch-accents in the spoken language and cannot take part in the prosodic structure indicated in (2c), as predicted by the analysis of intonation structure in Steedman 2000a, 2014 and chapter 6.

$$
\begin{array}{cccc}
\text{[Freeman admires]} & \text{and} & \text{[Hardy says Willis detests]} & \text{sincerity} \\
\hline S/NP\ {>}\mathbf{B} & (\mathrm{T}\backslash_{\star}\mathrm{T})/_{\star}\mathrm{T} & S/NP\ {>}\mathbf{B} & S\backslash(S/NP) \\
: \lambda x.admires\,x\,freeman & : \lambda f \lambda g \lambda x.fx \sqcap gx & : \lambda x.says\,(detests\,x\,willis)hardy & : \lambda p.p\ sincerity' \\
\hline & \multicolumn{2}{c}{(S/NP)\backslash_{\star}(S/NP)\ {>}} & \\
& \multicolumn{2}{c}{: \lambda g.\lambda x.g\ x \wedge says\,(detests\,x\ willis)hardy} & \\
\hline \multicolumn{3}{c}{(S/NP) : \lambda x.admires\,x\ freeman' \wedge says\,(detests\,x\ willis)hardy\ {<}} & \\
\hline & \multicolumn{3}{c}{S : admires\,sincerity\,freeman' \wedge says\,(detests\,sincerity\,willis)hardy\ {<}}
\end{array}
$$

Figure 11.1: Unbounded right-node raising

(3) a. A saxophonist [that$_{(N\backslash_{\diamond\star}N)/_{\diamond\star}(S/NP)}$ [[Harry admires]$_{S/NP}$ and [Louise says she detests]$_{S/NP}$]$_{S/NP}$]$_{N\backslash_{\diamond\star}N}$

b. A saxophonist that$_{(N\backslash_{\diamond\star}N)/_{\diamond\star}(S/NP)}$ *[[Harry admires]$_{S/NP}$ and [Louise says she detests him]$_S$]]

c. A saxophonist that$_{(N\backslash_{\diamond\star}N)/_{\diamond\star}(S/NP)}$ *[[Harry admires him]$_S$ and [Louise says she detests]$_{S/NP}$]

(4) [Harry admires $_{S/NP}$ and [Louise says she detests]$_{S/NP}$]$_{S/NP}$ [the saxophonist]$_{S\backslash(S/NP)}$.

a. *[[Harry admires]$_{S/NP}$ and [Louise says she detests him]$_S$] [the saxophonist]$_{S\backslash(S/NP)}$.

b. A saxophonist that$_{(N\backslash_{\diamond\star}N)/_{\diamond\star}(S/NP)}$ *[[Harry admires him]$_S$ and [Louise says she detests]$_{S/NP}$] [the saxophonist]$_{S\backslash(S/NP)}$.

The theory also predicts the ill-formedness of the following violation of the "same-case condition" on the ATB exception to the Coordinate Structure Constraint, since the right conjunct is not of the same CCG type as the left conjunct:

(5) a. *A saxophonist that$_{(N\backslash_{\diamond\star}N)/_{\diamond\star}(S/NP)}$ [[Harry admires]$_{S/NP}$ and [detests bossa nova]$_{S\backslash NP}$]$_*$

b. *[[Harry admires]$_{S/NP}$ and [detests bossa nova]$_{S\backslash NP}$]$_*$ [the man in the Brooks Brothers shirt]$_*$

However, in the case of (6a) there is another derivation, in which *Harry admires* is analyzed as a bare relative under the mechanism set out in section 9.5, analogous to the unreduced relative in (6b)

(6) a. ?A saxophonist [[that detests bossa nova]$_{N\backslash N}$ and [Harry admires]$_{N\backslash N}$]$_{N\backslash N}$

b. ?A saxophonist [[that detests bossa nova]$_{N\backslash N}$ and [that Harry admires]$_{N\backslash N}$]$_{N\backslash N}$

c. *[[Detests bossa nova$_{S\backslash NP}$ and [Harry admires]$_{S/NP}$]$_*$ [the man in the Brooks Brothers shirt]$_*$

Since the fragment *Harry admires* can and must be analyzed as bare relative $N\backslash N$, for some speakers (6a) escapes the same-case condition via the "backdoor" of this alternative derivation. This analysis is not available for the right-node raised version (6c).

Finally, the fact that *embedded* subject extraction from bare complements of verbs like *think* can happily coordinate with object extraction confirms the analysis of embedded subject extraction in section (9.6) as arising from *rightward* subcategorization for NP_{+wh} of the matrix verb:

(7) a. A saxophonist that some critic [thought was a genius]$_{(S\backslash NP)/NP_{+wh}}$ and [praised]$_{(S\backslash NP)/NP}$.
b. A saxophonist that some critic [praised]$_{(S\backslash NP)/NP}$ and [thought was a genius]$_{(S\backslash NP)/NP_{+wh}}$.

Despite the overwhelming evidence that coordination is a rule operating over like types in the strictest sense, it has sometimes been suggested that the Coordinate Structure Constraint and the ATB exception are an illusion. The suggestion is most influentially made on the basis of some examples first noticed by Ross (1967), Goldsmith (1985), and Lakoff (1986) like the following:

(8) a. What did you [go]$_{VP}$ and [see]$_{VP/NP}$?
b. This is the stuff that people in the Caucasus [drink every day]$_{VP/NP}$ and [live to be a hundred]]$_{VP}$.

Such examples have been used by Kehler (2002) Asudeh and Crouch (2002) and (Zhang, 2010, 135-139) to argue against the reality of the ATB generalization.

Ross and Goldsmith themselves argued that extractions like (8a) involve a lexicalized multiword aspectual item “go and”, which in present terms must bear the category VP/VP (and tensed etc. variants). They also argued that cases like (8b) involve another, noncoordinate, subordinating lexical category for *and*, and as such do not constitute counterexamples to the CSC and ATB constraints. They note the presuppositional and volitional semantics of the sentences in question (and the absence of such overtones from true coordinates), as well as the fact that—as Postal 1998 points out—no *other* conjunctions support such extractions. Thus:

(9) a. What did you go to see?
b. This is the stuff that people in the Caucasus drink every day so that they live to be a hundred.

(10) a. *What did you go or see?
b. *This is the stuff that those guys in the Caucasus drink every day or play dominoes.

Nor are such leftward extractions mirrored by equivalent right-node raising, as in (11a), unlike the standrd across-the-board cases like (11b):

(11) a. *Those guys in the Caucasus drink every day and live to be a hundred a kind of fermented mare's milk.
b. Harry admires and Louise says she detests several very famous saxophonists.

The suggestion by these authors of additional asymmetric categories for "and" was tentatively adopted in *TS*was tentatively adopted in *TS*. However, the latter observation makes these sentences look akin to the phenomenon of so-called "subject gapping", seen in English examples like the following, in which the right conjunct is not adjacent to a left conjunct of the same type:

(12) Through the door bounded a huge dog and [bit me in the leg]$_{S\backslash MP}$.

The phenomenon is rather marginal in English, but it is very common in German and Dutch. Heycock and Kroch (1994) note the relation of subject-gapping to the apparent violation of the ATB constraint. Subject gapping is analysed in Steedman, 2017 as a form of stripping coordination.

11.2 Right-edge restrictions

: The following related asymmetry has led Wilder, 1999:(5a) to postulate a Right Edge Restriction (RER) on right-node–raising from which ATB *wh*-extraction is exempt (ibid.:(5b)):

(13) a. *I gave a present and congratulated all the winners.
b. The man who I gave a present and congratulated.

Related examples were the reason for GPSG abandoning Gazdar's 1981 claim that RNR was mediated by the same mechanism as *wh*-extraction, and to HPSG embracing a deletion/ellipsis analysis (Beavers and Sag, 2004). However, we saw for (8) and (9) that "gave a present" is not a constituent with the type of a transitive verb, as evidenced by the fact that, unlike all other double-complement verbs (with the possible exception of "promise" Bach, 1979, 1980), Heavy NP shift, which involves the same composition as medial RNR) is blocked for the ditransitive:

(14) a. *I gave a present all the winners.
b. I persuaded to take a bath all the winners

Instead, (13b) can be allowed, and (13a) excluded, by an antecedent-governed lexical category for ditransitive verbs, parallel to the embedded subject-extracting bare complement verb (**??**):[3]

(15) gave:= $((S\backslash NP)/NP_{ANT})/NP) : \lambda w \lambda x \lambda y.gave\,w\,x\,y$

The derivation of (13b) is then as follows:

(16)
$$\begin{array}{ccccccc}
\text{man} & \text{who} & \text{I} & \text{gave} & \text{a present} & \text{and} & \text{congratulated}\\
\overline{N} & \overline{(N\backslash N)/(S/NP_{ant})} & \overline{S/(S\backslash NP)} & \overline{((S\backslash NP)/NP_{ANT})/NP} & \overline{NP^{\uparrow}} & \overline{(X\backslash_{\star}X)/_{\star}X} & \overline{(S\backslash NP)/NP}\\
 & & & \multicolumn{2}{c}{\overline{\qquad (S\backslash NP)/NP_{ANT}) \qquad}{<}} & & \\
 & & & \multicolumn{4}{c}{\overline{\qquad\qquad (S\backslash NP)/NP_{ANT} \qquad\qquad}{<>}} \\
 & & \multicolumn{5}{c}{\overline{\qquad\qquad S/NP_{ANT} \qquad\qquad}{>\mathbf{B}}} \\
 & \multicolumn{6}{c}{\overline{\qquad\qquad N\backslash N \qquad\qquad}{>}} \\
\multicolumn{7}{c}{\overline{\qquad\qquad N \qquad\qquad}{<}}
\end{array}$$

(13a) is disallowed, because the right-node-raised NP is lexical, rather than *ANT*-marked.

The present analysis has the advantage that we do not need to stipulate exemptions from RER for (13b), and examples like the following, which RER would otherwise appear to exclude:

(17) a. I persuaded to take a bow and congratulated all the winners.
b. I sold to the library and Mary donated to the museum several very valuable books

Wilder's related example (18) is blocked under present assumptions, because the type of "Joss will donate to the library" and of "Mary will donate" have to be the same (that is, either S/NP or $(S/PP)/NP$) for coordination to apply and RNR to be possible.

(18) *[Joss will donate $\_{i}$ to the library and Mary will donate $\_{i}]_{S/NP}$ several old novels$_i$]$_S$] to the museum

In the former case, "to the museum" would have to distribute over "John will donate to the library, and Mary will donate, several old novels" (cf. Citko and Gračanin-Yuksek, 2020:36):

Wilder analyses (19a) as object RNR, conflicting with the present account:

(19) a. John should [fetch]$_i$ and [give the book$_i$ to Mary].
b. John should [fetch]$_{i,j}$ and [give]$_{i,j}$ the book$_i$ to Mary$_j$.

3. As one would expect, there are dialects of English that lack this special-case category and simply disallow (13b).

However, the analysis (b), in which "the book" and "to Mary" are right-node-raised out of both conjuncts is also available, and seems to correspond to the only available reading.[4]

11.3 Asymmetric islands

: There have similarly been many claims in the literature since Wexler and Culicover (1980:299) that left- and right- movement are not symmetrical with respect to island constraints (Beavers and Sag, 2004; Cann, Kempson, and Marten, 2005; Sabbagh, 2007; Citko and Gračanin-Yuksek, 2020. See Postal, 1998, Steedman, 2012:101-103, Bachrach and Katzir, 2009, and Kubota and Levine, 2020:325-327 for counter-arguments.)

Examples like the following are not entirely compelling (the judgments are Beavers and Sag's), especially when care is taken to make intonational prosody the same in (a) as in (b) (Steedman, 2000a, 2012:103):

(20) a. ??Those unflattering pictures of Qaddafi, Yo knows several men who buy, and Jan knows several men who sell.
b. Yo knows several men who buy, and Jan knows several men who sell, those unflattering pictures of Qaddafi.

It seems possible that the asymmetries between *wh-* and RNR–extraction claimed (with a degree of uncertainty) by Citko and Gračanin-Yuksek (2020:ch.3) for Slavic languages are similarly discourse-sensitive, rather than reflecting any fundamental difference in the nature of the long-range dependencies involved. Sabbagh, citing an anonymous reviewer, offers the following apparent violation of the adjunct island constraint:

(21) Politicians win when they defend, and lose when they attack, the right of a woman to an abortion.

Again, the judgment is his, but the example seems no better and no worse than the corresponding left-extraction "What right do politicians win when they defend and lose when they attack?". The same seems to be true of the "non-coordinate RNR" examples discussed by Hudson (1976); Postal (1994), and Phillips (2003). Bachrach and Katzir (2009) and Hirsch and Wagner (2015)

4. A similar analysis appears to apply to related examples discussed by Whitman (2009). See Kubota and Levine, 2020 for discussion and an alternative categorial analysis of coordination to the present one.

discuss related examples of ATB *wh*-extraction out of islands. The lack noted earlier of a clear distinction between strong and weak islands, and the dependence of the latter on pragmatic factors, make it hard to draw any conclusion concerning asymmetry in left/right extraction from these data.

Examples like (22a) require "focal" stress or accent on the stranded prepositions, as indicated by capitals, and that similar intonated conjoined fragments also seem to license wh-extraction, as in (22b) (Ambridge and Goldberg, 2008; Chaves, 2012):

(22) a. CHESTER likes the person who visited us FROM, and LESTER likes the person who gave us a ticket TO, the beautiful island of Capri.
b. A place that CHESTER likes the person who visited us FROM, and LESTER likes the person who gave us a ticket TO.

Beavers and Sag, 2004 also note, following Davis (1992), that in Hausa, an SVO language with object pro-drop, while ATB object *wh*-extraction is allowed, RNR is not (Davis, 15, 16). However, the availability of object pro-drop in Hausa means that phonological emptiness cannot be taken as evidence of movement per se. If Hausa "movement" is really left-dislocation with pro-drop then there may be asymmetries with respect to discourse characteristics of left- and right-dislocated elements, with the former being by definition discourse-old as required by pro, but the latter required to be new or contrastive, as in English RNR. (This suggestion seems consistent with Davis's own analysis of finite and non-finite verbs in Hausa (ibid:(22)).) Related considerations may explain the asymmetry noted by McCloskey (1986) for Irish prepositions, which engender obligatory pro-drop, and "strand" for right-node raising, but not for *wh*-extraction (see below, although *pro* can act as resumptive Legate 1999:(11)).

11.4 Ross's Argument/Adjunct Cluster Coordination Asymmetry

One of the clearest confirmations that type raising is a universal feature of natural language morphosyntax related to case comes from the phenomenon of so-called nonconstituent coordination:

(23) I showed Chester Harry, and Lester Louise.

The derivation is written in full including the semantics as in figure 11.2a, from which it will be evident that the argument clusters (and type-raised categories in general) bear logical forms reminiscent of the "VP-shell" constituents postulated for the ditransitive by Larson (1988, 1990). The construction was

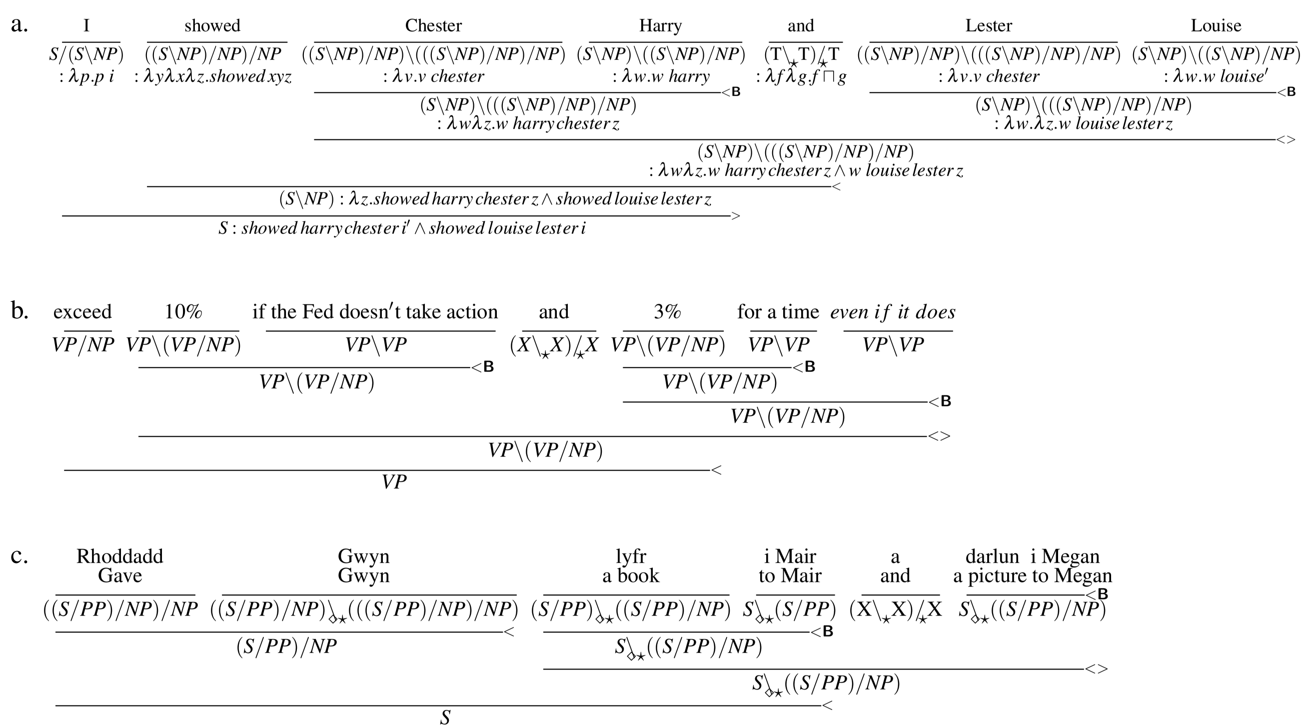

a.
I
showed
Chester
Harry
and
Lester
Louise
$S/(S\backslash NP)$: $\lambda p.p\ i$
$((S\backslash NP)/NP)/NP$: $\lambda y\lambda x\lambda z.showed\,xyz$
$((S\backslash NP)/NP)\backslash(((S\backslash NP)/NP)/NP)$: $\lambda v.v\ chester$
$(S\backslash NP)\backslash((S\backslash NP)/NP)$: $\lambda w.w\ harry$
$(T\backslash_\star T)/_\star T$: $\lambda f\lambda g.f \sqcap g$
$((S\backslash NP)/NP)\backslash(((S\backslash NP)/NP)/NP)$: $\lambda v.v\ chester$
$(S\backslash NP)\backslash((S\backslash NP)/NP)$: $\lambda w.w\ louise'$
$(S\backslash NP)\backslash(((S\backslash NP)/NP)/NP)$: $\lambda w\lambda z.w\ harry\,chester\,z$
$(S\backslash NP)\backslash(((S\backslash NP)/NP)/NP)$: $\lambda w.\lambda z.w\ louise\,lester\,z$
$(S\backslash NP)\backslash(((S\backslash NP)/NP)/NP)$: $\lambda w\lambda z.w\ harry\,chester\,z \wedge w\ louise\,lester\,z$
$(S\backslash NP)$: $\lambda z.showed\,harry\,chester\,z \wedge showed\,louise\,lester\,z$
S : $showed\,harry\,chester\,i' \wedge showed\,louise\,lester\,i$
b.
exceed
10%
if the Fed doesn't take action
and
3%
for a time
even if it does
VP/NP
$VP\backslash(VP/NP)$
$VP\backslash VP$
$(X\backslash_\star X)/_\star X$
$VP\backslash(VP/NP)$
$VP\backslash VP$
$VP\backslash VP$
$VP\backslash(VP/NP)$
$VP\backslash(VP/NP)$
$VP\backslash(VP/NP)$
$VP\backslash(VP/NP)$
VP
c.
Rhoddadd
Gave
Gwyn
Gwyn
lyfr
a book
i Mair
to Mair
a
and
darlun i Megan
a picture to Megan
$((S/PP)/NP)/NP$
$((S/PP)/NP)\backslash_{\diamond\star}(((S/PP)/NP)/NP)$
$(S/PP)\backslash_{\diamond\star}((S/PP)/NP)$
$S\backslash_{\diamond\star}(S/PP)$
$(X\backslash_\star X)/_\star X$
$S\backslash_{\diamond\star}((S/PP)/NP)$
$(S/PP)/NP$
$S\backslash_{\diamond\star}((S/PP)/NP)$
$S\backslash_{\diamond\star}((S/PP)/NP)$
S

Figure 11.2:

noted by Williams (1981) as posing serious problems for the related but differently type-driven GPSG account of coordination.

The semantics of type raising, composition, and coordination given earlier at (11) of chapter 2, (6) of chapter 9.4, and (2) of this chapter guarantees that the derivation delivers the same logical form as *I showed Chester Harry and I showed Lester Louise*, as shown. The analysis also has the important property of allowing a certain amount of attested nonparallelism. For example, the number of adjuncts included in coordinating clusters may differ, as in the following real-life example from the Penn Wall Street Journal treebank, in which the first conjunct includes one adjunct to the verb *exceed*, and the second, two:

(24) Inflation will exceed [10% if the Fed doesn't take action], and [3% for a time even if it does].

The derivation is parallel to (19) and figure 11.2a, apart from involvement of adjuncts rather than the raised object. The crucial part of the derivation is shown in figure 11.2b.

The fact that the finite and infinitival complements can also undergo cluster coordination in examples like the following shows that they too have categories raised over the matrix verbs, like other argument types such as NP:

(25) a. I told Harry that it was raining and Sally that it would snow.
b. I told Sally to take a bath and Harry to have a wash.
c. I saw Harry come and Sally go.

However, the fact that extraction is possible out of finite and infinitival complements means that they also retain the unraised type, unlike other argument types like NP:

(26) a. The house that I told Harry that I would buy.
b. The bath that I told Sally to take.
c. The horse that I saw Harry steal.
d. *The horse that I met a man who stole.

Ross (1970) pointed out that this construction illustrates a universal asymmetry across languages, namely that rightward arguments and enclitic adjunct clusters form as right conjuncts, while leftward arguments and proclitic adjuncts form as left conjuncts. For example, in Welsh, a VSO language, clusters are also right conjuncts, as in figure 11.2b. Borsley et al., 2007:52 note that this derivation consitutes a movement-free version of an across-the-board head movement or verb-fronting account of the construction in a theory like that of Roberts (2005) deriving Welsh VSO surface orders from underlying SVO, and is akin to the remnant-movement examples discussed in section 9.9.

Japanese, being a verb-final language, provides an example of leftward cluster-coordination, in which coordination itself requires a syntagmatic combinatory rule:[5]

(27)
$$\begin{array}{ccccc}
\text{Ken} - \text{ga} & \text{Naomi} - \text{o,} & \text{Erika} - \text{ga} & \text{Sara} - \text{o} & \text{tazuneta} \\
\text{Ken.NOM} & \text{Naomi.ACC,} & \text{Erika.NOM} & \text{Sara.ACC} & \text{visited} \\
\hline
S/(S\backslash NP_{nom}) & (S\backslash NP_{nom})/((S\backslash NP_{nom})/NP_{acc}) & S/(S\backslash NP_{nom}) & (S\backslash NP_{nom})/((S\backslash NP_{nom})/NP_{acc}) & (S\backslash NP_{nom})/NP_{acc} \\
\hline
\multicolumn{2}{c}{S/((S\backslash NP_{nom})/NP_{acc})\ {>}\mathbf{B}} & \multicolumn{2}{c}{S/((S\backslash NP_{nom})/NP_{acc})\ {>}\mathbf{B}} & \\
\hline
\multicolumn{4}{c}{S/((S\backslash NP_{nom})/NP_{acc})\ \&} & \\
\hline
\multicolumn{5}{c}{S\ >}
\end{array}$$

"Ken visited Naomi, and Erika Sara"

With more arguments, there are further derivational possibilities:

(28) Kyooju-ga gakusee-ni komonjo-o
[⟨Professor-NOM {student-DAT⟩ manuscript-ACC}]
miseta
showed$_{((S\backslash NP_{NOM})\backslash NP_{DAT})\backslash NP_{ACC}}$
'The professor showed the student the manuscript.'

The argument cluster *Kyooju-ga gakusee-ni komonjo-o* can form a constituent of type $S/(((S\backslash NP_{NOM})\backslash NP_{DAT})\backslash NP_{ACC})$ (indicated by [] brackets), which can coordinate with another preverbal argument cluster of the same type before applying to the ditransitive verb *miseta* on the right.[6]

Since there are three arguments in the cluster, there are two possible derivations of the cluster (respectively indicated by ⟨⟩ and { } brackets around binary clusters). The binary subunits themselves can coordinate within the cluster as a whole. All these variants are semantically correct, for the same reason figure 11.2 is. The implications for incremental parsing and interpretation of verb-final languages are discussed in appendix C.

The fact that the cross-linguistic availability of argument/adjunct-cluster co-

5. Scrambled argument orders like (ia) can apparently also also undergo cluster coordination in Japanese:
(i) a. Naomi-o Ken-ga, Sara-o Erika-ga tazuneta.
b. ?Naomi-o Ken-ga, Erika-ga Sara-o azuneta.
This observation might seem to call into question the account of scrambling in terms of crossed composition in chapter 5, and to suggest that Japanese verbs are simply lexically ambiguous as to argument order, as wa suggested by Steedman and Baldridge (2011). However, clusters that would under that assumption have different types can also coordinate, as in (ib). This suggests that something more is going on in these examples, as indeed is the case for the English gapped sentence offered as a translation in (27).

6. Whitelock (1991) discusses prosodic evidence for such argument-cluster constituents and a corresponding left-branching derivational analysis of the Japanese clause.

ordination is a prediction of CCG is one of its major explanatory achievements. The phenomenon was among those leading Pesetsky (1995:175) to postulate "cascade structure as an additional level of constituent structure in addition to standard "layered" syntactic structure. Under the present systems these are alternatives at a single non-representational level of derivation structure.

The fact that the inclusion of order-preserving type raising further predicts Ross's generalization and the non-existence of the unattested class of languages with the same verbal lexical types as Japanese, but allowing the English pattern of cluster coordination, or with the same lexical types as English, but allowing Japanese-style leftward cluster coordination like the following, is similarly important (see Steedman 1985; Dowty 1985/1988).

(29) *Chester Harry and I showed Lester Louise.

11.5 Preposition Stranding and Coordination

The stranding prepositional category (65) of chapter 9.4 yields derivations such as those in figure 11.3 under coordination:[7]

The related examples in figure 11.4 are problematic for any account that (explicitly or implicitly) treats the category of English prepositions as PP/NP or $PP^{\uparrow}/NP$—that is as specifier of argument PP.

In English, we have seen that proposition-stranding is possible under both leftward and rightward extraction. However, in reaction to the account of extraction in Gazdar (1981), which predicts a similar symmetry, McCloskey (1986) pointed out that Modern Irish, which never allows preposition stranding under relativization, does nevertheless allow it under Right Node Raising (McCloskey (1)):[8]

7. Right-node-raising of the kind exhibited in figure 11.3c is pragmatically demanding, and has consequences for the intonational prosody in which such sentences are uttered. It is a virtue in the theory that the nonstandard constituency that supports such analyses of coordinate structure, define exactly the notion of constituency that is required to support those intonation structures, as discussed in chapter 6.

8. McCloskey claims that preposition stranding is not possible with Heavy NP shift in Irish. However, we have seen in section 9.8 that the same holds for English preposition stranding ander the preent analysis.

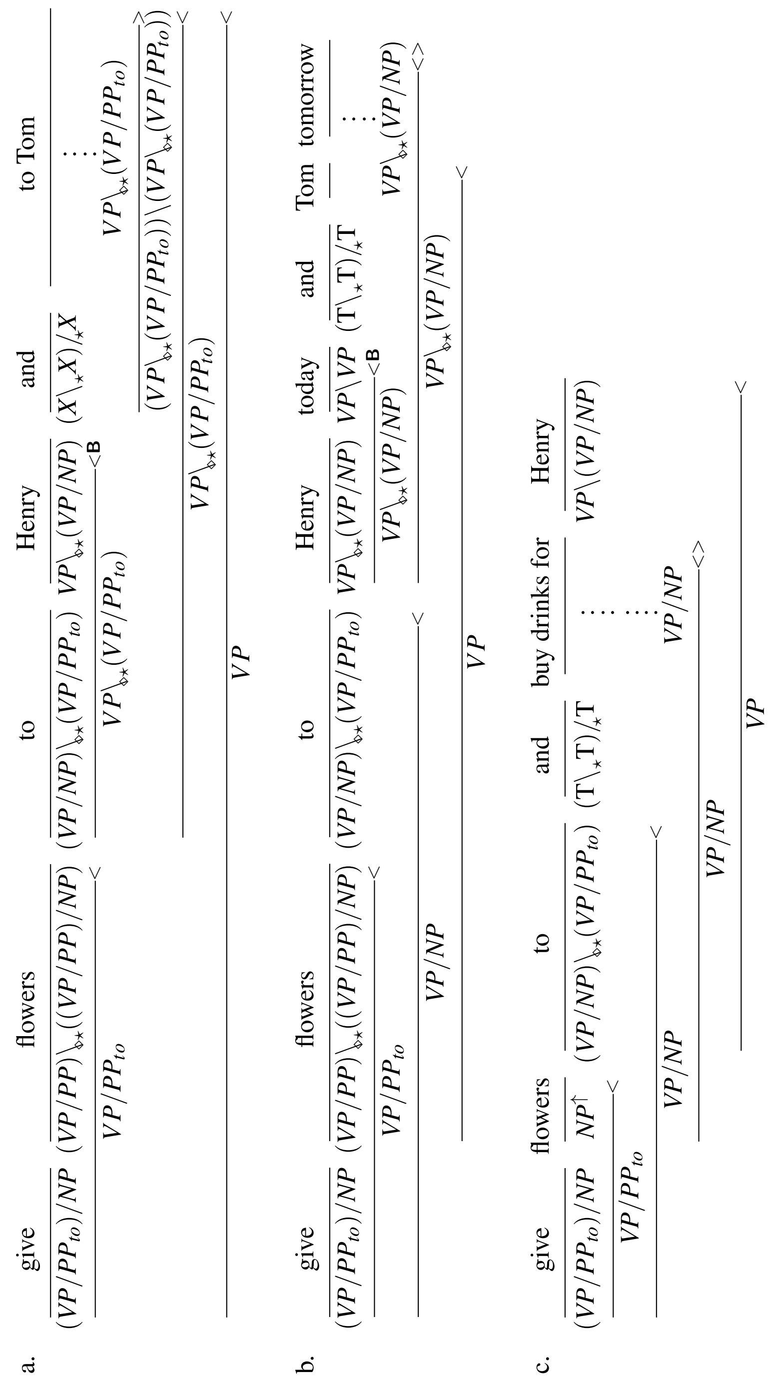

a. give flowers to Henry and to Tom
b. give flowers to Henry today and Tom tomorrow
c. give flowers to and buy drinks for Henry

Figure 11.3:

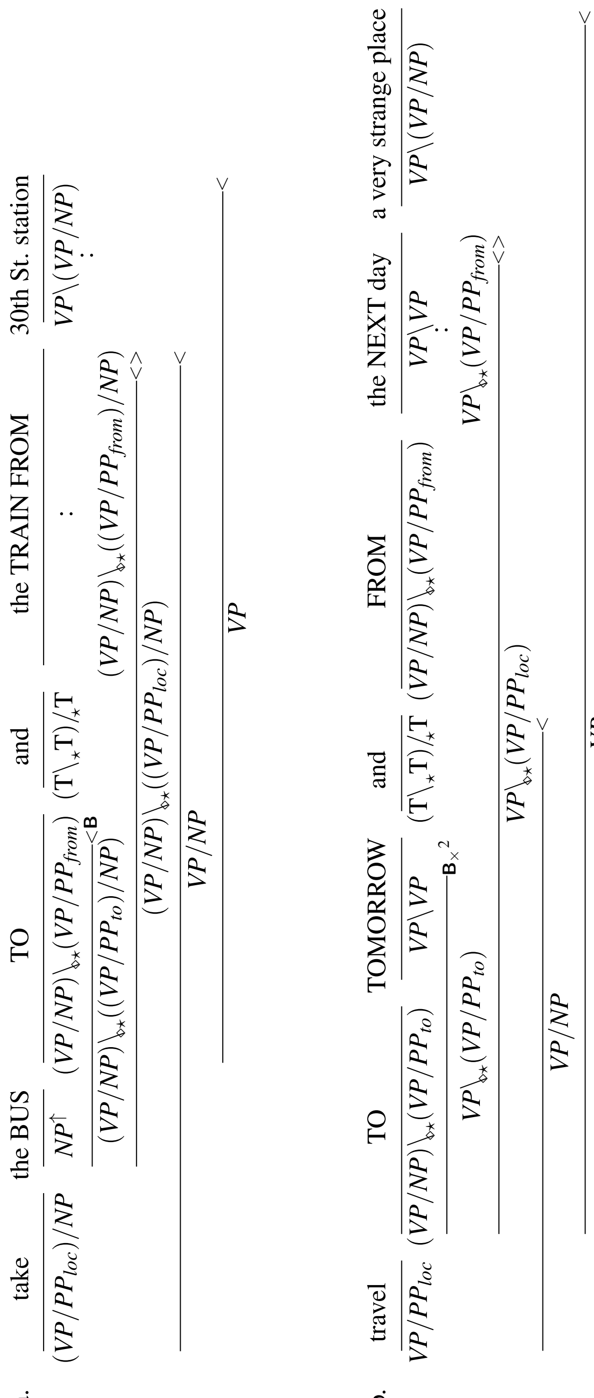

a.
take
the BUS
TO
and
the TRAIN FROM
30th St. station
$(VP/PP_{loc})/NP$
$NP^{\uparrow}$
$(VP/NP)\backslash_{\diamond\star}(VP/PP_{from})$
$(T\backslash_{\star}T)/_{\star}T$
$VP\backslash(VP/NP)$
$(VP/NP)\backslash_{\diamond\star}((VP/PP_{to})/NP)$
$(VP/NP)\backslash_{\diamond\star}((VP/PP_{from})/NP)$
$(VP/NP)\backslash_{\diamond\star}((VP/PP_{loc})/NP)$
VP/NP
VP
b.
travel
TO
TOMORROW
and
FROM
the NEXT day
a very strange place
VP/PP_{loc}
$(VP/NP)\backslash_{\diamond\star}(VP/PP_{to})$
$VP\backslash VP$
$(T\backslash_{\star}T)/_{\star}T$
$(VP/NP)\backslash_{\diamond\star}(VP/PP_{from})$
$VP\backslash VP$
$VP\backslash(VP/NP)$
$VP\backslash_{\diamond\star}(VP/PP_{to})$
$VP\backslash_{\diamond\star}(VP/PP_{from})$
$VP\backslash_{\diamond\star}(VP/PP_{loc})$
VP/NP
VP

Figure 11.4:

(30)

Níl	sé	in aghaidh	an	dlí	a thuilleadh	a bheith
is-not	it	against	the	law	anymore	be.INF
[ag éisteacht	le	nó	ag breathnú	ar]		
[listen.PROG	with	or	look.PROG	on]		
ráidió	agus	teilifís	an	Iarthair.		
radio	and	television	the	West.GEN		

"It is no longer against the law [to listen to or look at] Western radio and television"

This possibility can be captured by assuming that Irish does have the English particle-like category for strandable prepositions proposed in (65) of chapter 9.4, but that it is restricted to non-*wh*-government using the same feature that was used in that chapter to restrict English extractable complement subjects to *wh*-government, thus:

(31) le := $(VP/_{\diamond\star}NP_{-WH})\backslash_{\diamond\star}(VP/PP_{le})$

McCloskey further (n.4) claims that right-node-raising preposition stranding in Irish and a number of languages showing the same asymmetry is unlike English in only being possible if *both* conjuncts include a stranded proposition. This too can be captured if $VP/_{\diamond\star}NP_{-WH}$ marked for non-*wh*-government can only coordinate with a similarly restricted category.

11.6 Multiple and Correlative Coordination

The formal basis for multiple coordinations like the following have remained an irritatingly unsolved problem for all theories of coordination (Zwart, 2005; Lasnik, 2011; Chomsky, 2013):

(32) a. I met [Bob, (and) Carol, (and) Ted . . . , *(and) Alice].
b. I [marinated, (and) cooked, (and) . . . , *(and) ate] the mushrooms that I gathered beneath the gloomy ramparts of the Grange.
c. He is either a fool, (or) a madman, . . . , *(or)/*and a crook

The category in (2), $(T\backslash_{\star}T)/_{\star}T$, only accounts for case where the optional conjunctions are all present. However, in "monosyndetic" (infixing) coordination in English and many other languages only the final conjunction is obligatory. (In many languages, including Japanese and Thai, the final conjunction can also be omitted, but only when it is equivalent to "and".)[9]

9. In "bisyndetic" (postfixing) coordination of the kind found in some Athabaskan languages each conjunct seems obligated to carry a conjunction (Haspelmath, 2004:5)

One might be tempted to assume that all the intermediate conjuncts such as *Hardy* in (32a) bear the same category as as the final *and Willis*, either by virtue of the comma or the related phonological marking bearing the conjunction category, or (more likely, since the commas and intonational markers are frequently dropped) by virtue of a unary rule. (This is the approach taken by Maxwell and Manning 1996 within an LFG framework.) However, the same mechanism would allow the final conjunction to also be dispensed with, contrary to observation.

The disjunctive agreement between *either* and *or/*and* in (32c) shows that such "correlative" coordination or "conjunction doubling" (Dik, 1968; Progovac, 1998) has to mark the type of conjunction (Zhang, 2007). We will therefore replace the earlier conjunction types with categories like the following:[10]

(33) and $:= (\mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{+})/_{\star}\mathrm{T} : \sqcap$
or $:= (\mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{+})/_{\star}\mathrm{T} : \sqcup$
...
both $:= \mathrm{T}/_{\star}\mathrm{T}_{and} : \lambda x.x$
either $:= \mathrm{T}/_{\star}\mathrm{T}_{or} : \lambda x.x$
...

These categories combine with left conjuncts by the following special-purpose versions of the backward application rule:[11]

(34) *Backward Application: Conjunction Rules*

$$\mathrm{T}:a \quad \mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{+}:f \quad \Rightarrow_{<+} \quad \mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{\star}:\lambda x.f\,a\sqcap x$$
$$\mathrm{T}:a \quad \mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{+}:f \quad \Rightarrow_{<+} \quad \mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{\star}:\lambda x.f\,a\sqcup x$$
...
$$\mathrm{T}:a \quad \mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{*}:f \quad \Rightarrow_{<\star} \quad \mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{\star}:\lambda x.f\,a\sqcap x$$
$$\mathrm{T}:a \quad \mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{*}:f \quad \Rightarrow_{<\star} \quad \mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{\star}:\lambda x.f\,a\sqcup x$$
...
$$\mathrm{T}:a \quad \mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{*}:f \quad \Rightarrow_{<} \quad \mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{\star}:f\,a$$
$$\mathrm{T}:a \quad \mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{*}:f \quad \Rightarrow_{<} \quad \mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{\star}:f\,a$$
...
$$\mathrm{T}:a \quad \mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{*}:f \quad \Rightarrow_{<} \quad \mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{+}:f\,a$$
$$\mathrm{T}:a \quad \mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{*}:f \quad \Rightarrow_{<} \quad \mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{+}:f\,a$$
...

---

10. *neither... nor* can be treated similarly, but are omitted here because they involve the complication of negative polarity. The categories given here do not restrict *both... and* to binary conjunction, allowing *?both Bob, Carol, Ted, and Alice*. Further restricting them in this way is suggested as an exercise.

11. Languages that omit the final conjunction differ only in having an additional unary rule $\mathrm{T}: a \Rightarrow NP^{\uparrow}_{and}\backslash_{\star}NP^{\uparrow^{+}} : a\sqcap$.

The first two sets of rules ensure that there is at least one left conjunct, while the second two sets of rules terminate the iteration.

Thus, we allow derivations like the following:

(35)

$$
\begin{array}{cccccc}
\text{either} & \text{Bob,} & \text{Carol,} & \text{Ted,} & \text{or} & \text{Alice} \\
\overline{\mathrm{T}/_{\star}\mathrm{T}_{or}} & \overline{NP^{\uparrow}} & \overline{NP^{\uparrow}} & \overline{NP^{\uparrow}} & \overline{(\mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{+})/_{\star}\mathrm{T}} & \overline{NP^{\uparrow}}
\end{array}
$$

$$
\begin{array}{ll}
\underline{\qquad\qquad\qquad\qquad} > & \\
NP^{\uparrow}_{or}\backslash_{\star}NP^{\uparrow^{+}} & \\
\underline{\qquad\qquad\qquad\qquad} <+ & \\
(NP^{\uparrow}_{or}\backslash_{\star}NP^{\uparrow^{\star}}) & \\
\underline{\qquad\qquad\qquad\qquad} <* & \\
(NP^{\uparrow}_{or}\backslash_{\star}NP^{\uparrow^{\star}}) & \\
\underline{\qquad\qquad\qquad\qquad} < & \\
NP^{\uparrow}_{or} & \\
\underline{\qquad\qquad\qquad\qquad} > & \\
NP^{\uparrow}_{or} &
\end{array}
$$

The penultimate conjunction step of this derivation can iterate arbitrarily, and multiple conjunctions are allowed, so long as the final conjunction is included, to yield the following alternatives and contrasts:

(36) a. Bob, Carol, Ted, and Alice
b. Bob and Carol, Ted and Alice
c. Bob and Carol, and Ted and Alice
d. *Bob, Carol, Ted, Alice
e. *Bob, and Carol, Ted, Alice

As Winter (2001, 2007) points out, when intermediate conjunctions are present, there are multiple possibilities for distributivity across conjuncts, but when only the final conjunct is included, distributivity is fixed. This is predicted under the present account.

The fact that conjunctions and disjunction of arguments are type raised and marked for their conjunction type means that we immediately account for the fact that the categories given earlier for the *both* and *either* can "float" to higher positions in the sentence than the conjunction itself (Larson, 1985; Zhang, 2008):

(37) a. Kim married either Bob, Carol, Ted, or Alice.
b. Kim either married Bob, Carol, Ted, or Alice.

On the assumption that type-raised categories such as $itNP^{\uparrow}_{and}$ mark their result for the same conjunction type, we have derivations like the following for (37b)

(38)
$$\begin{array}{cccc}
\text{Kim} & \text{either} & \text{married} & \text{Bob, Carol, Ted, or Alice} \\
\overline{NP^{\uparrow}} & \overline{\mathrm{T}/_{\star}\mathrm{T}_{or}} & \overline{(S\backslash NP)/NP} & \overline{(S\backslash NP)_{or}\backslash((S\backslash NP)/NP)} \\
 & & \multicolumn{2}{c}{\overline{\hphantom{(S\backslash NP)/NP\quad(S\backslash NP)_{or}}}^{<}} \\
 & & \multicolumn{2}{c}{(S\backslash NP)_{or}} \\
 & \multicolumn{3}{c}{\overline{\hphantom{\mathrm{T}/_{\star}\mathrm{T}_{or}\quad(S\backslash NP)_{or}}}^{>}} \\
 & \multicolumn{3}{c}{S\backslash NP} \\
\multicolumn{4}{c}{\overline{\hphantom{NP^{\uparrow}\quad S\backslash NP\quad}}^{>}} \\
\multicolumn{4}{c}{S}
\end{array}$$

Since this is exactly the same mechanism used in other chapters for *in situ* *wh*-movement, clitic-climbing, and scope of quantifiers, negation, focus particles, and intonational phrasing, it is immediately predicted to be, like those constructions, both unbounded and sensitive to island constraints, as noted by Larson:

(39) a. Kim either [said that he would marry]$_{(S\backslash NP)/NP}$ [Bob, Carol, Ted, or Alice]$_{NP^{\uparrow}_{or}}$.

b. #Kim either [met a man who married]$_{\#}$ Bob, Carol, Ted, or Alice.

As in all those constructions, the similar intonation and information-structural requirements of correlative coordination are determined in the same way. However, correlatives like *either* are neither quantifiers (Dougherty, 1970; Stockwell, Schacter, and Partee, 1973), initial conjunctions, (Ross, 1967; Gazdar, 1981; Gazdar et al., 1985), nor focus particles (Hendriks, 2004; Johannessen, 2005; Zhang, 2008) (see Zhang for a historical and cross-linguistic review of these accounts).

From now on we will revert to the simpler category $(\mathrm{T}\backslash_{\star}\mathrm{T})/_{\star}\mathrm{T} : \sqcap$ as an abbreviation for full categories like $(\mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{+})/_{\star}\mathrm{T} : \sqcap$, and will regard the simple backward application rule as abreviating the conjunction rules (34), except in cases of actual multiple conjunction.

11.7 Gapping and Stripping

Steedman 1990 and *SP* show that it is also predicted that head-medial constructions like the following example of the English "gapping" construction universally pattern with the head-initial constructions:

(40) I like Ike and Chester, Adlai.

Unlike cluster coordination, a further discourse-based anaphoric process is required to recover the gap itself in example like (40). Such anaphoric processes are notoriously tolerant of minor feature mismatches such as agreement in the gapped material. Compare

(41) *I like Ike and Chester like Adlai.

The mechanism for gapping outlined in *SP* and extended in Steedman (2017) offers a second "gapping" derivational route for cluster coordinations like (23), according to which residues like *I showed* are recovered by this further process. Since as we have seen, gapping is notoriously less sensitive to syntactic detail than constituent coordination, this may explain the fact that examples like the following, in which the types of the cluster conjuncts require different diathesis alternates of the verb *show*, do not seem too bad, as Beavers and Sag (2004) point out:

(42) I showed [three boys a movie]$_{(S\backslash NP)\backslash(((S\backslash NP)/NP)/NP)}$, and [a video to two girls]$_{(S\backslash NP)\backslash(((S\backslash NP)/PP)/NP)}$.

If so, such examples do not controvert the generalization that coordination is essentially an operation over like types, contrary to their claim.

A further focus of doubts about the otherwise overwhelming generalization that coordination operates over like types arises from examples like the following:

(43) The temperature is [ninety]$_{NP}$ and [rising]$_{AP}$.

Cann et al. (2005, 216–219) discuss some more complex cases of apparent conjunction of unlike types, where there is a mismatch on a minor feature such as polarity or participial features of the VP under right-node raising.

TS:99–101 relates such examples to the phenomenon of right- and left-conjunct agreement in languages like Hebrew (Doron, 2000) and Welsh (Borsley et al., 2007), and suggests that conjunct dominance in agreement might better be analyzed as a low-level morphophonemic effect, like French *liaison*, the English *a/an* alternation in the indefinite article, or initial consonantal mutation in Celtic languages, including Welsh (cf. Benmamoun 2000 on agreement in Arabic as PF process).

For present purposes, we can ignore such minor details of exactly what counts as type compatibility for the purposes of coordination. The important generalization is that coordination is an operation over like types, however that is defined.

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

There is a natural tendency to see an analogy between across-the-board extraction under coordination and the "parasitic" extractions from adjuncts considered in the previous chapter, as Ross (1967) noticed, and as Williams (1990), Levine and Hukari (2006), Zhang (2010), and Chaves (2012) have proposed. Both involve a single extracted element with multiple dependencies, and therefore present a problem for the simplest movement accounts. It is inevitable that there should be some further parallels, because both of them involve an extraction from a right-adjunct from which extraction is normally disallowed.

(44) a. articles that$_i$ I [filed$_i$ [without [reading$_i$]]]
     b. articles that$_i$ I [filed$_i$ [and [forgot$_i$]]]

However, in other respects, the two constructions look very different. (Engdahl, 1983; Munn, 1993; Postal, 1993). Most important, as was seen in (25b) of chapter 9.4 the "true" extraction in (44a) is allowed without the "parasitic" gap in the adjunct, whereas *neither* extraction in ATB can occur without the other.[12]

Second, as Postal points out (among numerous other arguments), the types that can parasitically extract are much more restricted that those that can do so under the ATB condition, essentially to just NP and subcategorized PP.

Thus we have

(45) a. *How sick$_i$ did John look$_i$ without actually feeling$_i$?
     b. How sick$_i$ did John look$_i$ and actually feel$_i$?

(46) a. *How$_i$ did John mend$_i$ the car before fixing$_i$ the truck?
     b. How$_i$ did John mend$_i$ the car and fix$_i$ the truck?

The reason is that, as is well known, the way adjuncts extract is quite unlike that for arguments. In fact, such extractions are not in general unbounded, as we saw in section 9.7:

(47) a. How$_i$ did John mend$_i$ the car?
     b. *How$_i$ do you doubt that John mended$_i$ the car?

While Hornstein and Nunes (2002) offer an explanation in terms of the violation of the principle of Last Resort for why extraction of adjuncts is limited in this way, this principle then has to be "amnestied" by a Parallelism Constraint on coordination acting as a bare output constraint. That seems to mean that ATB remains an explicit constraint, contrary to Minimalist principles.

In this connection, it is worth recalling that the **B**- and **S**-combinatory rules of chapter 9 that permit extraction only apply to arguments Z. Coordination is mediated by a quite different mechanism limited to like-typable constitents, It

12. See section 11.1 on claimed exceptions

therefore allows the following facts to be captured without further stipulation:

(48) a. How$_i$ did John mend$_i$ the car and fix$_i$ the truck?
b. *How$_i$ do you doubt that John mended$_i$ the car and regret that he fixed$_i$ the truck?

Because we have lexicalized the bounded constructions like raising and control, they interact correctly with coordination. Thus in the following, the controled subject of the right-node-raised infinitival is separately bound by the ligical forms of the two conjuncts:

(49) a. Lester agreed and Chester refused to leave.
b. Chester seemed and Lester appeared to be drunk.

11.9 Discussion

Many of the supposed asymmetries between leftward and rightward outlined above depend upon the analysis of coordination. The predominant approaches within the minimalist program seem to be based either on the idea of sideward movement (Nunes, 2004) or the related idea of multidominance (McCawley, 1982; Goodall, 1983; Lin, 2002; Sarkar and Joshi, 1996; Wilder, 1999, 2008; Chen-Main, 2006; De Vries, 2009; Citko, 2011). More recently, there has been a return to the idea of movement and/or deletion under identity (Hirsch, 2017; Torr and Stabler, 2016; Schein, 2017), partly because of difficulties for multidominance arising from the interaction of quantifier scope and coordination, to which we turn in a later chapter. Zhang (2010) proposes yet another approach, based on pronominal anaphora.

However, all of these approaches require structural parallelism conditions on coordinate structures, and none of them seems to explain Ross's 1970 observations concerning the relation of the direction of gapping to canonical word-order, one of the strongest syntactic universal generalizations that is known.

In contrast, CCG defines parallelism in terms of simple identity of syntactic types, and a syntax that eliminates all operations equivalent to overt or covert "movement", "deletion", "copying", and all other varieties of action-at-a-distance, in favor of type-dependent combinatory syntactic and semantic merger over elements that are overt, adjacent, and directly interpretable, in which Ross's generalization is explained as a corollary of the Combinatory Projection Principle, which says that information specified in the lexicon, including directionality, must be projected by syntactic derivation, an assumption akin to the inclusiveness condition.

As a consequence, CCG differs from the GPSG account of coordination (Gazdar, 1980, 1981; Gazdar et al., 1985). Its use of directional types makes right node raising and right cluster coordination from head-initial verbs a prediction of the theory. GPSG requires an explict rightward displacement rule $\alpha \;\rightarrow\; \alpha/\beta \quad \beta$ to license rightward extraction (Gazdar, 1981:178-9).[13]

The accounts that CCG affords for *wh*-extraction and right node-raising dependencies depend on an identical mechanism of composition and case (type-raising), making the intervening material into a typable constituent. There is therefore a strong prediction of symmetry between the two constructions, such that whatever can undergo extraction can also undergo right node-raising, and vice versa.

Apparent exceptions to the symmetry of left- and right-extractions seem to reflect subtle interactions of the direction of extraposition with information-structural aspects of semantics, such as theme and rheme status, that are entirely independent of the syntactic mechanism of extraction, rather than revealing any shortcoming in the central assumption of CCG that rightward and leftward extraction are syntactically symmetric as a consequence of the Combinatory Projection Principle and the Projective Dependency Principle of chapter 2.

Exercise : The above account of coordination requires us to radically rethink our definition of the notion “constituent”. Defend (or attack) the CCG notion of constuency as combinatory typability against (or with) the traditional linguistic notion(s).

13. This rule is dropped in Gazdar et al., 1985, which does not include any analysis of right node raising or Heavy NP Shift.

Chapter 12

Combinatory Minimalism

There will be seen in it demonstrations of those kinds which do not produce as great a certitude as those of Geometry, and which even differ much therefrom, since whereas the Geometers prove their Propositions by fixed and incontestable Principles, here the Principles are verified by the conclusions to be drawn from them; the nature of these things not allowing of this being done otherwise
—*Treatise on Light* Christiaan Huygens, 1690 (tr. Silvanus P. Thompson)

The present account eliminates from the formal theory of grammar every form of movement, whether of the overt-, covert-, A-, Ā-, *wh*-. quantifier-, scrambling-, sideward-, remnant-, head-, topic-, or focus- variety. As a consequence, it necessarily also renders unnecessary many constraints and minimality conditions limiting movement.

This reduction is achieved by radical lexicalization—that is by specifying all the long-range dependencies that have been described in terms of movement via lexical entries for their functional governors (such as control verbs, quantifier determiners, and relative pronouns) with a syntactic type (such as $(VP/VP_{to})/NP$, $(S/(S\backslash NP))/N$, or $(N\backslash N)/(S/NP)$), and a logical form (such as $\lambda x\lambda p.persuade\,(p\,x)\,x$,, $\lambda n\lambda p.p\,(skolem\,n)$, or $\lambda p\lambda n\lambda x.n\,x \wedge p\,x$) that already expresses the (possibly unbounded) dependency statically, via λ-bound variables.[1]

These lexical entries express the entire combinatoric potential of the word-category pair, including subcategorization if it is a verb, and such semantic relations between its arguments as control—again, statically. In the terms of the competence theory, a subset of these lexical types corresponding to arguments are also *cased* or type-raised, either statically, via inflection as in Latin, or dynamically or "structurally" as in English, or by a "quirky" mixture of the two, as in Icelandic.

This means of course that many conditions on the grammar are recast as

1. A combinatory variable-free notation for logical form is possible, but is, as noted earlier, considerably less readable.

conditions on the lexicon, where they must be analysed as part of the definition of "possible lexical category". For example, subjacency must be explained in terms of the CCG type-system that is investigated in section 12.5 below, which universally excludes non-subjacent "super-raising" and "super-control" verbs from the theory of grammar.

However, the generalization in question can now be interpreted as applying to the type-system of logical form, and ultimately to the underlying conceptual representational "language of mind" that makes language and language acquisition possible, as proposed in chapter 2 and in appendices B and A.

The lexical entries of the radicalized lexicon are projected syntactically by a small set of combinatory rules of an even smaller fixed set of types confined to (besides application), composition and substitution, onto constituents including sentences, bearing derived categories, again consisting of a type and a logical form, the latter being assembled entirely compositionally by the derivation.

The inclusion of morpholexical type-raising, and of syntactic composition and substitution means that the inventory of constituent types onto which the lexicon is projected by combinatory derivation is extended greatly beyond any traditional account of constituency, so that, for example, "John says Mary found", "burn without reading" and "a policeman a flower" are all derivationally typable and interpretable constituents in their own right, rather than being of type S or VP containing the residues of movement or deletion.[2]

The following generalization then follows:

(1) *Coordination, relativization, quantifier-scope-taking, and intonation structure are all defined over the same generalized notion of syntactic constituency, which is entirely transparent to a surface compositional semantics. All of these constructions obey the constituent condition on rules directly, without requiring movement, deletion, or mediating structural representations.*

However, we also noticed in chapter 4.1 that, in order to achieve this generalization, we had to give up some traditional assumptions concerning constituency in surface syntactic derivation, and admit fragments like "I met", "I think I met", and "Mary a book last Saturday" as syntactically typable constituents and first-class citizens of the grammar, and to live with the idea that

2. This departure from more traditional notions of constituency is sometimes held against CCG as evidence of inconsistency as a theory of grammar (e.g. Osborne and Groß, 2016). It is actually evidence of the inconsistency of all of the traditional tests for constituency *except for coordination and intonation*.

even a simple transitive sentence has a surface derivation in which the object c-commands the rest of the derivation. This raises the question of how we can define processes which have been argued to depend on traditional constituency and structural relations of c-command, of which the two most important are anaphora and coreference on the one hand, and variable binding by operators such as Universal Quantifiers on the other. These topics are the subject of further chapters in Part III below on semantics and anaphora. However, the following remarks are in order at this point.

12.1 Binding and Coreference

First, neither phenomena can depend on surface constituency, since there are right node-raising derivations which force a possibly embedded argument to take derivational scope over the residue, without forcing either a violation of the anti-c-command condition on pronouns and antecedents or a wide-scope-only reading in the object, contrary to the facts:

(2) a. Lola$_{S_t/(S/NP)}$ [Anna thinks Manny likes]$_{S/NP}$.
b. [Anna$_i$ thinks Manny likes]$_{S/NP}$ HER$_{i\ S\backslash(S/NP)}$.

(3) a. Some saxophonist$_{S_t/(S/NP)}$, [every woman likes]$_{S/NP}$. $\forall\exists/\exists\forall$
b. [Every woman likes]$_{S/NP}$ some saxophonist$_{S\backslash(S/NP)}$. $\forall\exists/\exists\forall$

(It was partly for this reason that we adopted a lexicalized clitic analysis of reflexives and reciprocals in chapter 7.)

However, traditional notions of constituency and c-command still hold in CCG at the level of logical form, which is arguably the level of representation that those traditional intuitions concern in the first place. To the extent that they are grammatical phenomena at all, both anaphoric coreference and scope-taking must therefore be defined at the level of logical form, as they are under the program of Chomsky (1995b).

Nevertheless, it is also well-known from examples like the following, repeated from the introduction, that binding is not solely determined in any simple sense by structural command at any level (cf. Reinhart, 1983a; Higginbotham, 1983; Lebeaux, 1991, 2009; Jacobson, 1994; Hornstein, 1995:108-111; Romero, 1998, Fox, 1999, Büring, 2005:256-260, Takahashi and Hulsey, 2009, Barker, 2012, and Bruening, 2014):

(4) The person that Lola$_i$ works for $\_{j}$, she$_i$ likes$_j$.
Which person that Lola$_i$ works for $\_{j}$ does she$_i$ like$_j$?

The observation has led to proposals for "Late Merge", free type-lifting/continuation-passing, "Roll-up" movement, or definition in terms of "Almost C-command", higher-type $((e \rightarrow t) \rightarrow t)$ traces, and/or "hybrid" movement systems involving both copies and traces (Lechner, 1998, 2018; Keine and Poole, 2018).

Instead we will follow Bruening (2014) in assuming that coreference is not a relation between referring expressions represented by nodes in a logical form, but rather a relation between referring expressions and referents in a dynamically-changing contextual model. The reason that "Lola" and "she" can co-refer in (6) is that by the time "she" is processed, the referent of "Lola", namely an entity *lola*, has been added to the contextual model as a possible antecedent.

The examples themselves in (6) show that the process of updating the contextual model cannot wait until the end of the sentence. However, it cannot just be as soon as the referring expression "Lola" is complete, as that would allow examples like the following, contrary to Principle B:

(5) *Lola likes her.

It must rather be that the entity *lola* and any other referents that have been mentioned in the same binding domain gets added to the contextual model as soon as either (i) the domain itself is exited from, or; (ii) a subordinate new binding domain is entered,

So, for example, in addition to examples (6), the following are possible where $^{\uparrow lola}[$ and $]^{\uparrow lola}$ represent type i and type ii updates, and $^{\downarrow her}$ represents reference:

(6) a. Lola$_i$ thinks $^{\uparrow lola}[$she$_i^{\downarrow her}$ is a genius].

b. [That Lola$_i$ is a genius]$^{\uparrow lola}$ surprises her$_i^{\downarrow her}$

c. [Pictures of Lola$_i$]$^{\uparrow lola}$ amuse her$_i^{\downarrow her}$.

There is a little more to say about the examples in (31) of chapter 1, repeated here:

(7) a. [The person that she$^{\uparrow she}$ works for]$_j$, Lola likes$_j^{\uparrow lola}$.

b. [Which person that she$^{\uparrow she}$ works for]$_j$ does Lola like$_j^{\uparrow lola}$?

While sometimes misleadingly referred to as "backward anaphora", what is really going on in these examples is that when a pronoun without an available antecedent in the contextual model (here, "she"), is encountered, a proxy referent (here, *her*), with the sole property of being "given", or in the terms of chapter 6 "background", and the logical form of a subsequent definite (i.e. given) NP (here, "Lola") is then predicated of it, making it refer definitely (here, to *lola*). Note that "backward anaphora" to indefinites is degraded:

(8) a. #Which person that she_i works for does some $woman_i$ like?
b. Which person that a $woman_i$ works for does she_i like?

The mechanism for modifying the referential context in this way is located in the lecxical logical forms of the heads of clauses and NPs. For example:

(9) Lola $:= NP^{\uparrow} : \lambda p.p\,lola$
thinks $:= (S\backslash NP)/S : \lambda s\lambda y.thinks\,s\,y \wedge push\,y\,context$
she $:= NP^{\uparrow} : \lambda p.her = pop\,context \wedge p\,her$

(10)
$$
\begin{array}{cccc}
\text{Lola} & \text{thinks} & \text{she} & \text{is a genius}\\
\hline
NP^{\uparrow} & (S\backslash NP)/S & NP^{\uparrow} & S\backslash NP\\
:\lambda p.p\,lola & :\lambda s\lambda y.thinks\,s\,y\wedge push\,y\,context & :her=pop\,context\wedge\lambda p.p\,her & :\lambda y.genius\,y\\
\hline
\multicolumn{2}{c}{S/S:\lambda s.thinks\,s\,lola\wedge push\,lola\,context \quad {>}\mathbf{B}} & \multicolumn{2}{c}{her=lola\wedge genius\,lola \quad {>}}\\
\hline
\multicolumn{4}{c}{S: thinks\,(genius\,lola)\,lola \quad {>}}
\end{array}
$$

(Note that unlike most derivations in the book, this derivation depends on left-right incrementality.)

The notion of binding domain invoked above resembles the minimalist notion of "phase" (Chomsky, 2001). However, it is an exclusively semantic notion, rather than a syntacxtic one. It follows that no "Phase Impenetrability Condition" (PIC) is needed or can be defined.

12.2 Grammar without Action-at-a-Distance

The driving force behind CCG was from the very first to reduce transformational action-at-a-distance to operations over string-adjacent contiguous elements. This was enshrined in the earliest papers in the form of a Principle that can be stated as follows (cf. Ades and Steedman, 1982:533, *SP*:54).

(11) *The Principle of Adjacency*
Syntactic rules can only apply to pairs of string-adjacent typed categories.

Since combinators are by definition operators that apply to adjacent terms, and since we know from the work of Schönfinkel (1924), Curry and Feys (1958), and Smullyan (1985, 1994) that even quite small collections of unrestricted combinators have the full expressive power of the λ-calculus, the principle of Adjacency says little more than that a combinatory calculus is adequate to express natural language grammar including the language of logical form, and that action-at-a-distance is not a formal necessity, *contra* Chomsky (2005):10.

However, as we have seen, combinatory categorial grammar is much more restricted than the cominatory calculi that are equivalent to the λ-calculi. First, the categories are *typed*, and those types define linear precedence. Second, the syntactic combinatory rules that project precedence information are constrained by the Combinatory Projection Principle (5) of chapter 2, which is itself a corollary of the Principle of Adjacency (11) and the following two more specific principles (cf. *SP*:54) in which we continue to distinguish that argument functor of a combinatory rule whose range X is the range of the result of the rule as the "governing functor", and that argument functor whose range Y is that of one of the arguments of the principal functor as the dependent functor:

(12) *The Principle of Linear Consistency*
If the linear precedence specified for the argument of the governing functor corresponding to the result of the dependent functor is rightward (leftward), then the dependent functor must be adjacent to the right (left).

(13) *The Principle of Linear Inheritance*
The linear precedence specified for any argument of the input functors that appears in the result of a combinatory rule must be the same as that specified on the input functor type(s).

These three principles collectively constitute what in earlier chapters was referred to as the Combinatory Projection Principle (CPP). They constrain the expressive power of CCG, putting it in a class of grammars that is the least more expressive than context-free that is of linguistic interest. We shall see that, in the terms of the Chomsky hierarchy of natural families of language, it is far less expressive that the context-sensitive (Type 1) languages, and far smaller than Joshi's 1985 subclass of "mildly context-sensitive" languages, so it seems appropriate to call it "nearly context-free".

12.3 The Place of CCG in an Extended Language Hierarchy

Vijay-Shanker, Weir, and Joshi (1987), Weir (1988), Joshi et al. (1991), and Vijay-Shanker and Weir (1994) showed in a series of papers that, under the formalisms that were then current, four grammatical formalisms—TAG, CCG, Linear Indexed Grammars (LIG, Gazdar, 1988) and Head Grammars (HG, Pollard, 1984) were weakly equivalent—that is, that they could capture the same set of languages or stringsets. They also showed that the worst-case complexity of the decision problem for such grammars—that is, the problem of deciding whether a string was in the language or not—was polynomial, a property which has important consequences for efficient parsability. They did this by showing that such grammars were drawn from a "natural family of languages" (AFL), characterized by a class of phrase-structure rules, an automaton, and exemplars of languages that are not recognizable by any less powerful grammar or automaton.

That is not to say that these grammars are strongly equivalent. In particular, their derivation trees and their treatments of unbounded dependencies are quite different (although *SP* presents an intuitive demonstration that for every CCG there is a strongly equivalent LIG).

Vijay-Shanker and Weir went on to show that the natural generalization of TAG was to the class of linear context-free rewriting systems (LCFRS), which under some further restrictions are polynomially decidable (Satta, 1992). LCFRS are equivalent as a class to multiple context-free grammars (MCFG, Seki, Matsumura, Fujii, and Kasami, 1991).

12.3.1 The Extended Language Hierarchy

These observations require us to extend the Chomsky Language Hierarchy by interpolating a number of levels besides the original four types, as in figure 12.3.1, in which lowercase letters represent terminals, uppercase leters represent nonterminals, lowercase greek letters represent strings of terminals and nonterminals, and exponents on terminal symbols as in a^n represent sequences of n occurrences of that terminal.

As in the original hierarchy introduced in chapter ??, each level properly contains the level below, except that IG and LCFRS are mutually properly intersecting.

The fact that the original Chomsky hierarchy had four levels should not lead one to assume that those levels are in any computational sense equidistant. The type 0 languages are those that are recursively enumerable sets, which for present purposes means all sets with a formal definition. Such sets are explanatorily vacuous as a theory of natural language in the sense that they place no restrictions on possible languages in the sense of stringsets. Savitch (1987) shows that in this sense, type 1 (CS) grammars are not significantly less

Grammar Type	Automaton	Rule-types	Exemplar
Type 0: RE	Universal Turing Machine	$\alpha \rightarrow \beta$	PA-valid
Type 1: CSG	Linear Bounded Automaton (LBA)	$\phi A \psi \rightarrow \phi \alpha \psi$	$a^{n!}$
MCFG (LCFRS)	ith-order EPDA	$A_{[[(i),...]...]} \rightarrow \phi B_{[[(i),...]...]} \psi$	$\mathscr{P}(a^n b^n c^n)$
IG	Nested Stack Automaton (NSA)	$A_{[(i),...]} \rightarrow \phi B_{[(i),...]} \psi C_{[(i),...]} \xi$	a^{2^n}
LIG/CCG/TAG	Embedded PDA (EPDA)	$A_{[(i),...]} \rightarrow \phi B_{[(i),...]} \psi$	$a^n b^n c^n$
Type 2: CFG	Push-Down Automaton (PDA)	$A \rightarrow \alpha$	$a^n b^n$
Type 3: FSG	Finite-state Automaton (FSA)	$A \rightarrow \begin{cases} \text{a B} \\ \text{a} \end{cases}$	a^n

Figure 12.1: The extended language hierarchy

expressive than Type 0.[3]

The unnumbered overlapping LCFRS and IG classes are much less expressive than Type 1, the context-sensitive class. Nevertheless, they include many languages that seem to have highly unnatural properties, like Bach's MIX language $\mathscr{P}(a^n b^n c^n)$, consisting of all permutations on the same number n of occurrences of some number of terminals, and the "non-constant growth" language a^{2^n}.[4]

The unnumbered level comprising CCG, TAG, LIG and HG is much less expressive than LCFRS, and excludes such languages. We will refer to this natural family of languages as "near- context-free". It implies the extended version of the Chomsky and Schützenberger (1963) hierarchy in figure 12.3.1.

This result is surprising when one reflects that the natural generalization of TAG is to the LCFRS class, while the natural generalization of LIG is to IG, so that their weak equivalence seems almost accidental.[5]

3. The restriction implicit in the linear bounded automaton of the type 1 level is merely to a Turing machine with bounded memory—that is, to something that is more like a real computer.

4. There are recurring claims in the literature that the phenomenon of reduplication, which is quite widespread in natural languages, is productive and generates subsets of strings of the form a^{2^n}. An early example from Manaster-Ramer (1986) concerns reduplication in certain US English dialects, which is a marker of dismissive emphasis, as in the old joke about a mother who is told that her son has an Œdipus complex, and replies "Œdipus Schmœdipus, what does it matter if he's a good boy and loves his mother?". This construction might seem to open up the possibility of rejoinders such as "#Œdipus Schmœdipus SCHMŒDIPUS Schmœdipus, your son needs help." However, native speakers invariably reject such examples, saying that such markers can only apply once only to *bona fide* lexical items.

5. The definitions of CCG used in the present book is different from those in Steedman (1996) and used by Vijay-Shanker and Weir (1994) as the basis for the Joshi et al. (1991) proof of weak equiv-

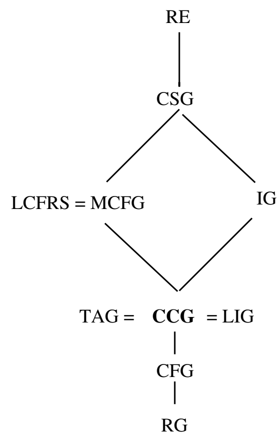


Figure 12.2: The Revised Extended Chomsky Hierarchy

12.3.2 Worst-Case Complexity of CCG Parsing

Vijay-Shanker and Weir (1993) also show that the TAG/CCG/LIG/HG near-context-free grammars are recognizable in the worst case in polynomial time On^6 for sentences of length n, and present full parsing algorithms with that bound, including one for CCG (Vijay-Shanker and Weir, 1990).

It is important to be clear that the real significance of this result is not the particular polynomial identified. Worst-case complexity tells us very little about average-case complexity, and the algorithm itself may not be practicable. The

---

alence to TAG/LIG/HG. The main difference is that in early version of CCG including *SP*, type restrictions on the variables $X, Y, Z, \ldots$ in combinatory rules were allowed. (For example, crossed composition could be restricted in English to cases where $Y = S'$, excluding crossed composition of nominal adjuncts $N\backslash N$). In the present version of the theory, such type-based restrictions on rules are disallowed. Instead, slash-types do similar work in limiting overgeneration. (For example, we saw that the type $N\backslash_{\diamond\star}N$ of English nominal adjuncts prevents the the crossed composition rule from applying to them to generate Germanic orders like *#a nice at the office man*.) Nevertheless, for the purposes of Their proof, the systems are eqivalent.

Kuhlmann et al. (2015) showed that under the definitions used by Steedman and Baldridge (2011), CCG would be strictly less expressive than under the old definition. In particular, as noted earlier, the Germanic crossed dependency construction modeled by the language $a^n b^n$ with only crossing dependencies could not be accepted without also accepting some additional word-orders with adjacent non-crossing dependencies. As we have noted in section 5.2, such mixed word orders are in fact generally characteristic of the Germanic constructions that allow such dependencies, like the Zurich version of German considered there. The Dutch construction that *only* allows the completely crossing depenendencies could not be handled syntactically by the version in Steedman and Baldridge (2011) and related papers. However, the present system of slash modalities, first proposed in FUNLG following Baldridge (2002), restores weak equivalence to TAG and LIG.

significance of polynomial complexity is, rather, to guarantee the applicability of simple and efficient generalizations of "divide-and-conquer" algorithms such as CKY (Cocke and Schwartz, 1970), of the kind used for compiling programming languages, including incremental algorithms, which in the average practical case may have acceptable costs in computational terms. We return to these questions in appendix C

12.4 Descriptive adequacy of CCG

The Combinatory Projection Principle (5) of chapter 2 limits the expressive power of the small set of combinatory rules to be "nearly context-free", either weakly equivalent to (Joshi et al., 1991), or, under the definitions used in the present work, properly contained in (Kuhlmann et al., 2015), the tree-adjoining languages. On either reckoning, this is the lowest linguistically interesting trans-context-free automata-theoretic natural family of languages that is known. Specifically, this class of grammar is very much less expressive than either indexed grammars (IG, Aho, 1968, 1969), or full linear context-free rewriting systems (LCFRS, Weir, 1988) or the equivalent multiple context-free grammars (MCFG, Seki et al., 1991). Stabler's minimalist grammars (MG, Stabler, 2011) are weakly equivalent to LCFRS/MCG (Michaelis, 2001).

Of course, low expressive power is of no interest unless the grammar is descriptively adequate—that is, capable of expressing the data cross-linguistically, supporting an adequate semantics. It is hard to prove descriptive adequacy in this strong sense, since there is no completely firm definition of exactly which phenomena must be covered syntactically or of their semantics. (We have argued for the exclusion of extraposition, VP-anaphora, and sluicing, but others may disagree). However, the theory outlined in this book and developed in different forms and at greater length in earlier publications, and has been applied to the syntactic and semantic analysis of coordination and unbounded dependency in a very wide range of languages (*SS&I*; Steedman 1985, 1990, 2000a; Steele 1990; Whitelock 1991; Hoffman 1995; Nishida 1996; Bozşahin 1998, 2002; Komagata 1999; Baldridge 1998, 2002, Trechsel 2000; Cha and Lee 2000; Park and Cho 2000; Çakıcı 2005, 2009; Hockenmaier 2006; Ponvert 2008; Ruangrajitpakorn, Trakultaweekoon, and Supnithi 2009; Boxwell and Brew 2010; Kubota 2010; Lee and Tonhauser 2010; Bekki 2010; Tse and Curran 2010; Steedman and Bozşahin 2016; Ambati 2016; Ambati, Deoskar, and Steedman 2016a.

Whether the near- context-free formalisms are fully descriptively adequate in the sense of capturing all linguistically attested phenomena is harder to demonstrate. In the case of CCG, we can make the strong prediction that if we have four elements of types $A|B$, $B|C$, $C|D$, and D , then of the 4!=24 possible orders, two cannot be recognized by CCG. Four such elements are the determiner, numerator, adjective and noun in English NPs such as "these five beautiful girls", a concrete example that has been investigated cross linguistica;lly by Greenberg (1963), Hawkins (1983), Cinque (2005), Abels and Neeleman (2009), Dryer (2009), and others). The two excluded orders are the following, corresponding in English to "*Five girls these beautiful" and its mirror image "*Beautiful these girls five":

(14) *a*. $B|C \quad D \quad A|B \quad C|D$
b. $C|D \quad A|B \quad D \quad B|C$

The prediction is that these word orders will not be allowed for any language on types of these forms. If a construction were found in any language that allowed them, then CCG, at least in its present form, based on the combinators **B**, **S**, and **T** would not be descriptively adequate, and would be falsified.

These particular orders are indeed absent from the 14 word-orders that are attested for these particular parts-of-speech (Cinque, 2005), although the fact that eight further orders that CCG could allow are also unattested means that this does not tell us very much. Because of the Zipfian distribution over such word-orders, the sample of attested languages is simply not large enough for a strong test of this kind, unless we turn to free word-order languages.[6]

For similar reasons, two out of the twenty-four possible permutations of the four elements of the English VP "will$_{A|B}$ have$_{B|C}$ eaten$_{C|D}$ beans$_{D}$", namely those corresponding to corresponding to "*have beans will eaten" and "*eaten will beans have", are excluded by universal grammar. If either order were attested, say in a language with an auxiliary system and freer word order than English, such as Dutch or German, then CCG in the present form would be falsified. (As we saw in chapter 2, this is not in fact the case for those languages.)

It is very difficult to test these predictions, because the availability of anaphoric linkage via pro-drop and phenomena like parenthesis and extraposition means that other analyses than those afforded by strictly syntactic projection are often implicit.

For example, Haug (2017) analyses the following Latin example (Caesar *De Bello Gallico* V.i.i) as an instance of backward adjunct control of the subject of

6. The two excluded orders are in fact among those allowed under Hawkins' (1983) version of the Greenberg's 20th universal. CHECK

the participial adjunct *discedens ab hibernis in Italiam* ("departing from winter quarters to Italy") by the subject *Cæsar* of the main clause *Cæsar ... imperat* ("Caesar ordered ..."). If that analysis is correct, then the categories are as follows, on the pattern (14a), and cannot combine:[7]

(15)

discedens ab hibernis	Cæsar	in Italiam	imperat
S/S	NP	$(S/S)\backslash(S/S)$	$S\backslash NP$
B/C	D	$A\backslash B$	$C\backslash D$

However, while the implicit subject of such participial adjuncts is frequently coreferential with the subject of the main clause, it can be otherwise. In particular, it can refer logophorically to the speaker or source of indirect discourse, as discusssed in section 14.3, as in the following absolute:

(16) Departing from winter quarters to Italy, the sun was shining.

De Bello Gallico is very much from Cæsar's point of view, being written as a self-justifying report intended to be read aloud by others (Mueller, 2012:xxiii-v).

Such dangling participials should probably be regarded, like parentheticals, as not "in construction" with the main verb. [8]

A more challenging counterexample would be an attestation in Latin of the following NP word orders as alternatives to *Hæ quinque puelæ pulchræ ambulant* ("These five beautiful girls are walking"):[9]

(17) a. *Quinque puellæ hæ pulchræ ambulant.
b. *Pulchræ hæ puellaæ quinque ambulant.

Both seem very bad to my schoolboy Latin ear, but I offer them as hostages to fortune.[10]

More generally, the prediction is that for any set of n categories

7. I assume that the PPs are adjuncts to *discedens*. type raising via case does not affect word order, so it is suppressed here.

8. In the full text, the main clause occurs in the middle of a blizzard of absolutes, extrapositions, and other parenthetical construction: *L. Domitio Ap. Claudio consulibus discedens ab hibernis Caesar in Italiam, ut quotannis facere consuerat, legatis imperat quos legionibus praefecerat, uti quam plurimas possent hieme naves aedificandas veteresque reficiendas curarent*: "with LD and AC consulted, departing from winter quarters for Italy, as he used to every year, Cæsar ordered the legates, who he put in charge of the legions, to build as many ships in the winter as they could, and to repair the old ones."

9. One needs to take care in considering such judgements that the words do carry the categories of determiner, numerator, adjective and noun—for example, that the adjective is not read instead as an extraposed NP modifier $NP|NP$ or a predicate $S|NP$, as opposed to $N|N$.

10. I am grateful to Rachel Hurley of Cardiff University for confirming (p.c.) that these two orders are indeed ungrammatical in Latin with the intended sense.

$\{A|B, B|C, \ldots, M|N, N\}$, of the $n!$ permutations that are possible, the number that cannot be recognized by CCG is the nth in a series called the Large Schröeder Numbers, of which the first few members are $\{0, 2, 30, 326, 3, 234, 31, 762, 321, 244, \ldots\}$. The Large Schröder Numbers represent the number of permutations of n that are non-separable, where separability is a property related to tree rotation. The Schröder numbers grow even faster than than the factorials, so that the proportion $n!/S(n)$of permutations that are non-seperable grows rapidly with n.

The observation of this property was first made in Williams, 2003:203-211 for his categorial calculus CAT. CAT has a standard directional categorial lexicon and rule of application, with a combinatory operation REASSOCIATE equivalent to composition, and an operation FLIP which reverses the directionality of a functor category which does the work of type raising, achieving the effect of higher-order fronting types.[11]

Williams (2003:229-234) applies a form of his calculus restricted to forward functors and forward composition to the analysis of Hungarian serial verb order discussed by Koopman and Szabolcsi (2000), which bears some similarity to the Dutch/German cross-serial verb complexes discussed in section **??**.

These two categorial accounts are therefore closely related. However, without the addition of morpho-lexical type-raising, or case, CAT will not express the coordinate structures discussed here in chapter 11.

If we renumber the original category set $A|B$, $B|C$, $C|D$, and D as X, 1, 2, 3, then (14b) also corresponds to the *1-3-X-2 constraint on movement observed by Svenonius (2007) for adjuncts, an observation which led Svenonius to complex stipulations of strong features and null functional heads to limit movement in Germanic roll-up derivations such as pied-piping. Such stipulations are unnecessary in CCG, and thereby explained.

12.5 Explanatory Adequacy of CCG

Since an explanatory theory should by definition have the smallest possible number of degrees of freedom, the fact that CCG is also of low, near-context-free expressive power, at the lowest linguistically interesting trans-context-free automata-theoretic level that is known, means that it also has some claim to the

11. Williams incorrectly claims (2003:209) that type-raising evades the constraints on movement that are corrolaries if FLIP. However, he has failed to notice that type-raising is a morpholexical operation in CCG (Steedman, 2000b), rather than a syntactic operation, and therefore has has no effect on the exclusion of non-separable permutations in CCG.

stronger level of *explanatory* adequacy.

Explanatory adequacy has been widely but confusingly associated in the linguistic literature since *Aspects* with the provision of an account of child language acquisition for the class of grammars involved. However, a model of language acquisition can be supplied for any theory of grammar, provided the recipient is prepared to stipulate sufficient innate apparatus to cut down the search space to a manageable size (Fodor, 1966).

The simplest and best account of acquisition is the one that minimizes the amount of innate machinery that must be stipulated, because it imposes less of a burden of explanation on the theory of evolution. That theory is the one that has the fewest degrees of freedom in the first place. Thus, a theory of acquisition requires, rather than constitutes, an explanatory theory of grammar.

To identify the degrees of freedom available in CCG, we must ask two more specific questions: "What is a possible lexical entry or category?"; and "What is a possible combinatory rule?".

12.6 On the Notion "Possible Lexical Category"

The set of lexical types that we have availed ourselves of in categories like (13) is also very restricted. It is a lot less than the set of all possible categories defined by the rule "if α and β are types, then α/β and $\alpha\backslash\beta$ are types" (Such restrictions correspond to what *Aspects* identified as "Substantive" Universals, which derive from the semantics.)

We therefore need to also compare the degrees of freedom implicit in the lexical type system of CCG with those of the Minimalist Program.

The definition of the CCG lexical type-system has a number of "levels," like the movement-based type-systems of Williams (2003), and others. These levels respectively correspond to: the language of the core predicate argument structures; the language of logical from which lexicalizes raising and control relations; the language of the morpho-lexicon, which lexicalizes the work of the minimalists "head movement" (Roberts, 2001), "scrambling" (Ross, 1967), and "sideward movement" (Nunes, 2001); the language of case or type-raising; and the language of *wh*-movement.

12.6.1 The Type-System of Predicate-Argument Structures

We start with the language of the predicate-argument structures of (1b) and (13), which is defined as follows:

(18) *The type-system of predicate-argument structure:*
1. types have bounded valency ≤ 4.
2. t and e are elementary types.
3. $e \rightarrow t$, $e \rightarrow e$, $t \rightarrow t$ and $t \rightarrow e$ are types.
4. If α is an elementary type and $e \rightarrow \beta$ is a type then $\alpha \rightarrow (e \rightarrow \beta)$ is a type.

Here e is the type of entities, standing in for a richer ontology of entities, distinguishing people, places, times, etc., and t is the type of propositions, also possibly standing in for a more diverse set of types. This simplified type system allows $in_{e\rightarrow e}$, $that_{t\rightarrow e}$, $think_{t\rightarrow(e\rightarrow t)}$ and $seemingly_{t\rightarrow t}$, for example.

The level of predicate-argument structure of the lexical type system corresponds to the level of thematic structure in Minimalism. To the extent that predicate-argument structure for any language is assumed to subsume expressions of a universal predicate-argument structure (UPAS) that supports language acquisition via semantic bootstrapping, although as with HPSG's ARG-ST and Simpler Syntax's "grammatical function tier" we do not assume any fixed repertoire of thematic roles (Dowty, 1991b).[12]

12.6.2 The Type-System of Logical Form

The second level of the lexical type-system corresponds to logical forms including relations such as raising and control. We have seen that the lexical logical forms constituted via the lambda-binders in (1b) are more diverse. They allow *properties* of type $e \rightarrow t$ as arguments of raising and control verbs, as well as arguments of types e and t. They also allow those arguments to combine in any order, regardless of lf-command relations, under the following principle:

(19) If a predicate p as defined in (18) has an argument of type t, then it can be realized in the predicate-argument structure of a corresponding lexical VP stem as *either* (i) an free-variable argument, also of type t, *or*; (ii) as the application of an argument free variable of type $e \rightarrow t$ to another free-variable argument of type e. The latter e-type argument can either be unique to this predication ("raising"), or it can be one of the original arguments of p ("control", "ECM").

For example, we can now have $seem_{(e\rightarrow t)\rightarrow(e\rightarrow t)}$, as well as $seemingly_{t\rightarrow t}$, and $quickly_{(e\rightarrow t)\rightarrow(e\rightarrow t)}$, as well as $quickly_{t\rightarrow t}$.

12. The present language-specific logical forms are in fact a proxy for UPAS. In particular, we do not assume that the predicate we write *promise* in the logical form of "promise" is actually atomic in UPAS.

This level of the lexical type-system corresponds to "A-movement" in the minimalist framework. The fact that such relations are subject to minimality conditions such as the Minimal Link Condition, together with exemptions from such conditions for subject control (Rosenbaum, 1967) follow immediately from the fact that these relations are lexicalized as between co-arguments.

12.6.3 The Type-System of the Morpho-Lexicon

To specify the stem syntactic types of the lexicon as they are usually thought of by linguists, we also need a language-independent mapping between the stem elementary semantic types and syntactic argument types (generalizing over minor features such as S', VP_{to}, etc.), where $\mapsto$ means "is the syntactic type corresponding to" :[13]

(20) $S \mapsto t$
$NP, PP \mapsto e$
$N, VP \mapsto (e \rightarrow t)$

(The mapping itself is non-essential. When written in full, it is one-to-one, so that we could use the original semantic types as syntactic types, at considerable cost to readability by mortals.)

Languages are then free to associate directional syntactic categories specifying the linear position of constituents corresponding to arguments of those types in all possible ways, with linking λ-binders to pass their values into universal predicate-argument structure (UPAS).

At this stage, the language-specific lexicon can further specify restrictions on lexical categories via slash-types, which limit the rules that can apply to them, and are projected by the rules onto the derivation under the combinatory projection principle (5), thereby imposing more or less rigid word-order on the language in question.

It is important at this point to note that $S\backslash NP$ and S/NP, the types of the tensed intransitive, are not yet syntactically typable, All verbs so far are functions into $VP_{e\rightarrow t}$, corresponding to the Minimalist vP. The stem types are then mapped onto further categories corresponding to Minimalist functional projections, in English either morphologically, as with tense, or in syntactic derivation, by auxiliary verbs, as with progressive and perfect aspect.

For example, English passive morphology and Dyirbal antipassive morphol-

13. N and VP don't actually correspond to the same semantic type $e \rightarrow t$. VP really maps to a more complex type that we might think of as $e \rightarrow (r \rightarrow t)$, where r is the Reichenbach/Davidsonian "event time" or "reference time". However, we use a simplified type system, for ease of reading.

ogy map accusative and absolutive transitive VPs onto intransitive VP. Thus for English

(21) -en := $VP_{pssv}\$\backslash\!\!\backslash VP\$: \lambda p\lambda x \ldots \lambda y.p \ldots x\, sk^{(x)}_{\lambda y.likely(p\,x \ldots y)}$

(where \$ schematizes over subcategorizations)

Similarly, past- and present- participial morphology turns English VP stems into participial forms. For example:

(22) -ing := $VP_{prog}\$\backslash\!\!\backslash VP\$: \lambda p\ \ldots\ \lambda v\lambda r.ongoing\,(p \ldots)$

(23) -en := $VP_{pppl}\$\backslash\!\!\backslash VP\$: \lambda p\ \ldots.consequent\,(p \ldots)$

Finally, either tense-morphology or a lexical rule maps infinitival verb stems $VP\$$ onto corresponding tensed forms of the same semantic type requiring a subject—for English, $S\backslash NP_{agr}\$$; for Welsh, $S\$/NP_{agr}$.

(24) -ed := $S\backslash NP_{agr}\$\backslash\!\!\backslash VP\$: \lambda p\ \ldots.past(p \ldots)$

(Thus, morphological operators may either increase or decrease valency.)

The adjunct types, are also defined at this stage, differing only from the various aspectual phrase governors in being endotypic, or type-preserving, rather than exotypic or type-changing. Thus we have the following categories for "seems" and "quickly" of type$(e \rightarrow t) \rightarrow (e \rightarrow t)$

(25) a. seems := $(S\backslash NP)/VP : \lambda p\lambda y.seemingly\,p\,x$
b. quickly := $(S\backslash NP)/(S\backslash NP) : \lambda p\lambda y.manner\,(p\,y)\,quick$

Such adjuncts may be similarly morpholexically derived,

(26) -ly := $(VP \mid VP)\backslash\!\!\backslash(N \mid N) : \lambda a\lambda p\lambda y.manner\,(p\,y)\,a$

This third level of the lexical type system corresponds to the minimalist levels of "head-movement" (Roberts, 2001), "scrambling" (Ross, 1967), and "sideward movement" (Nunes, 2001). The Head Movement Constraint (Travis, 1984) follows like other locality/minimality conditions from the fact that these relations are defined lexically.

A fourth stage of morpho-lexical derivation follows, in which case morphology (or "structurally" disambiguated underspecification) maps argument categories NP, VP, etc. and their heads or governing categories NP/N, VP/NP, etc. into order-preserving functors categories type-raised over all and only the lexical (tensed, nominal, participial, passive etc.) functor types that take them as arguments, such as $S/(S\backslash NP)$, $(S\backslash(S/NP))/N$.

In a final fifth stage of lexical derivation, the order-preserving cased categories that are specifically raised over $S\$$ are mapped to non-order-preserving lexical governors raised over tensed lexical types notably including S/NP, such as wh-words and *tough*-predicates, which differ from cased arguments in changing the type of their result to $N\backslash N$, $S\backslash NP$. In the case of topicalizers and *Wh*-question words, the changed category is a root construction S_{top} or S_{whq} for which no stem subcategorizes.

These levels feed each other sequentially in generating the lexicon. For example, the pied-piping *wh*-category of example (82) is repeated here:

(27) which $:= ((N\backslash N)/(S/NP))\backslash(((S\backslash(S/NP))/NP))$

This category is derived for the lexicon as follows. $e \to e$ is a level 1 type, so NP/NP is a level 3 syntactic type, so $(S\backslash(S/NP))/NP$ is a level 4 cased type, so $(S\backslash(S/NP))\backslash((S\backslash(S/NP))/NP)$ is a cased type, so $((N\backslash N)/(S/NP))\backslash((S\backslash(S/NP))/NP)$ is a level 5 *wh*-category.

All and only the Combinatory **B** and **S** Rules allowed under the CPP are then necessary to project the language specific lexical categories onto sentence derivations.

These various levels of lexicalization are reminiscent of Emonds's 1976 "structure-preserving", "local", and "root" transformations, and the related "levels" of movement phenomena of Williams, 2003 and Abels (2008), and the "tiers" of Culicover and Jackendoff, 2005. In particular, level 2 above corresponds to Abels' "A-mvt", level 3 to his "scrambling", and level 5 to his "*wh*-mvt" and "topicalization"

levels. (Level 4 of case assignment via type-raising is not recognized by Williams or Abels but may be related to Culicover and Jackendoff's "case tier".)[14]

The claim is that all and only the degrees of freedom this type system allows in the lexicon are observed in the languages of the world.

For example, we have noted that the type system allows the following categories for raising verbs:

14. It is interesting to speculate where this layered structure of levels stems from. Like all substantive universals, one plausible origin is the nature of the underlying cognitive representation. For example, Steedman (2002) suggests that the origin of type-raising/case lies in a represntation of objects in terms of the actions which they allow—in the terms of Gibson (1977), their "affordances"—a representation much of which which we share with other tool-using animals, notably chimpanzees.

(28) a. seems := $(S\backslash NP)/VP_{to} : \lambda p\lambda y.seemingly\,(p\,y)$
b. seems := $(S\backslash NP_{xpl})/S : \lambda p\lambda s.seemingly\,s$

This allows the observed English raising constructions

(29) a. John seems to be certain to leave.
b. It seems (that) John is certain to leave.

But we cannot write categories that would allow "super-raising", as in (30b), whose exclusion motivated the introduction of the Minimal Link Condition (MLC) (cf. Chomsky 1995b:82,(131b)). First, there are no wrapping or commuting combinatory rules in CCG that would allow *seems* to reach across *(that) it is certain* to directly control *to leave*. To allow (30b) we would instead need something like the lexical category (30a) for "seemingly", which throws away the optional complementizer and the expletive, and applies to the non-adjacent finite predicate *is certain to leave* with the intended interpretation:

(30) a. *seems := $(((S\backslash NP_{3s})/(S_{fin}\backslash NP_{3s}))/NP_{xpl})/(S'/S)$
$: \lambda q\lambda x\lambda p\lambda y.seemingly\,(p\,y)$
b. *John seems that it is certain to leave.

However, under the earlier definitions of the lexical type system and tense, tensed $S\backslash NP$ is not a possible argument for "seemingly" or any other verb, as opposed to adjuncts like "seemingly" (cf. Hornstein, 2009:168n25, who attributes this fact to the Principle of Greed, cf. Boeckx et al., 2010:ch.4):[15]

(31) a. *seems := $(S\backslash NP_{3s})/(S_{fin}\backslash NP) : \lambda p\lambda y.seemingly\,(p\,y)$
b. *John seems is certain to leave.
c. John seemingly is certain to leave.

In fact, we have seen that the only categories that can subcategorize for tensed predicates are type-raised arguments, adjuncts, and the "*wh*-movers", such as topicalized and relativized arguments (which are in effect themselves type-raised).

Nor, as in the related framework of Cormack and Smith (2004), is it possible to write "backward control" categories, of the kind required by the analyses of Monahan (2003) and Polinsky (2012)—see Boeckx et al., 2010:106-114 for discussion—since control is defined as a relation between coarguments of a single verb.

Nevertheless, the following lexical entries are both licensed, despite the fact that the second involves a controller y that is not the closest candidate x, in

15. Again the reason is ultimately semantic-conceptual.

apparent contradiction to the Minimal Link Condition (MLC) of Chomsky (1995b):

(32) a. persuades := $((S\backslash NP)/VP_{to})/NP : \lambda x\lambda p\lambda y.persuades\,(p\,x)\,x\,y$
b. promises := $((S\backslash NP)/VP_{to})/NP : \lambda x\lambda p\lambda y.promises\,(p\,y)\,x\,y$

The problem of how to limit movement so that it is bounded, excluding super-raising (31b) without at the same time excluding (32b), has been a problem since Rosenbaum, 1967 (cf. Hornstein, 2009:163-164), and was a major motivation for the PRO analysis of control, in which the binding of controlled elements is accomplished by extra-grammatical meaning postulates, rather than by movement. In the present framework, these constraints all follow from the fact that these constructions are lexicalized, rendering the MLC redundant.

Thus, the differing levels or domains of these authors' varieties of movement correspond to different levels of the morpholexicon: the stem which determines raising and control; the local operators such as tense, mood, aspect, and voice; the domain of case—that is, morpholexical order-preserving type-raising (including underspecification or "structurally" determined case), which determines scrambling, including so-called "long-range" scrambling; the domain of morpholexical type-changing higher-order categories such as relativizers, which determine the projection of lexical subcategorization onto unbounded dependencies by syntactic derivation. (Some of the latter, such as focalizing and topicalizing categories, limit the scope of the unbounded dependencies concerned to the root clause.)

However, none of these morpholexically-specified domains involves a distinct level of representation in the standard sense of the term. Each of them pairs an interpretation at the level of logical form—the only representational level that is countenanced in CCG—with a syntactic categorial type, which is the sole determinant of possible syntactic projections via the combinatory syntactic rules.

Up to this point, we have talked of the lexicon as if all forms related to a single stem are exhaustively listed. Given the degree of idiosyncrasy in the English lexicon, this is not unreasonable. Even in languages with highly productive morphology like Latin and Finnish, it may well be the most effficent way to run the processor. Nevertheless, to the extent that there are regularities across forms related to the same stem, it is useful to also have lexical rules rules expressing those regularities, so that once one form for a previously unseem stem is encountered, all the other forms can be predicted. However, whether such rules are used offline to compile those forms out in the lexicon in all their forms ("lexical redundancy rules"), or are used actively online ("lexical rules") is similarly a question of efficiency in implementation and emprical prediction for the psychologist. For purposes of the theory of grammar, all of these options are equivalent, so we will continue to ignore them here.

12.7 Envoi

Once an explanatory theory of grammar is achieved, the theory of linguistic competence is complete, and the purely theoretical side of the research program defined in *Syntactic Structures* is concluded.

However, an explanatory theory of grammar still does not constitute a complete theory of Language. To explain how a child acquires that grammar requires a theory of how they search the space of possibilities allowed by that theory, and of the evidence that guides this search. This process in turn presupposes an account of linguistic performance or use. Finally one must ask how evolution could come up with that resource in what seems to have been a very short space of time indeed—at most a few million years.

These further questions of the space of possible grammars allowed by CCG and the nature of the processing mechanism, their use in a model of language acquisition, and their possible origins in biological evolution. all concern performance mechanims, rather than the competence grammar that is the focus of this book. They have accordingly been relegated to a series of brief appendices, which the purely linguistically inclined reader may prefer to skip.

Exercise : The system of lexical types outlined above just applies to the content words, and omits function words like conjunctions, negation, and the quantifier determiners. How could they be brought into the system?

Part III

Semantics and Anaphora—See 2018 md. for changes

In most cases where we use the "indefinite" article we have really something very definite in our mind, and "indefinite" in the grammatical sense practically means 'nothing but "what shall not (not yet) be named".
—Otto Jespersen, *The Philosophy of Grammar*, 1924:113-4

Chapter 13

Quantifier Scope

QR and minimalism fit together awkwardly.
—Hornstein, 1999a:45

It is often assumed that the ambiguity of sentences like (1) is to be captured by assigning two alternative first-order logical forms which differ in the scopes assigned to the quantifiers, as in (2a,b):

(1) Every boy admires some saxophonist.

(2) a. $\forall x.boy\,x \rightarrow \exists y.saxophonist\,y \wedge admires\,yx$
b. $\exists y.saxophonist\,y \wedge \forall x.boy\,x \rightarrow admires\,yx$

The first reading seems to have all the elements of meaning roughly in the position that that they occur in in the original sentence, but the second has the elements of the existential saxophonist object above those of the universal boy subject.

Similarly, a universal object can take scope over a commanding existential subject in the strong sense of "distributing over" it, as in (3) and (4):

(3) Some boy admires every saxophonist.

(4) a. $\exists x.[boy\,x \wedge \forall y.[saxophonist\,y \rightarrow admires\,x\,y]]$
b. $\forall y.[saxophonist\,y \rightarrow \exists x.[boy\,x \wedge admires\,y\,x]]$

This observation has on occasion been explained in terms of "quantifier movement" (May, 1977, 1985) or essentially equivalent computational operations of "quantifying in" (Montague, 1973) or "storage" (Cooper, 1983) at the level of Logical Form. Since the string position of the quantifier does not change, such movement is often referred to as "covert". However, such accounts present a problem for theories of grammar like CCG that try to do away with movement or the equivalent in syntax. Having eliminated movement from the syntax, to have to allow it at the level of Logical Form would be a backward step, given the strong assumptions of transparency between syntax and semantics behind the theory. Instead we need to find a way to deliver all the readings for ambiguous sentences using nothing but the derivational combinatorics of surface grammar like (1) and (3).

13.1 The Universals

In chapter 2, we noted in connection with example (**??**) that the typed-raised categories that had been associated in that chapter with case were also compatible with a standard generalized quantifier semantics for universals like *every dog*. Accordingly, we gave the quantifier determiner *every* the category given in (**??**) which we can now schematize over the entire set of English cased raised types as follows:

(5) every, each := $NP^{\uparrow}_{3SG}/_{\diamond\star}N_{3SG} : \lambda p\lambda q\lambda \ldots \forall x[px \rightarrow qx\ldots]$

The category makes the universal determiners semantically transparent functions from nouns (the "restrictor") to type-raised nounphrases, schematizing over them using the $NP^{\uparrow}$ abbreviation, where the schematized types are simply the syntactic types corresponding to a (polymorphic) generalized quantifier, and . . . stands for any further arguments of the predicate q.[1]

Thus, (5) can be instantiated by the following categories, among others:

(6) a. every, each := $(S/(S\backslash NP_{3SG}))/_{\diamond\star}N_{3SG} : \lambda p\lambda q.\forall x[px \rightarrow qx]$
b. every, each := $(S\backslash(S/NP_{3SG}))/_{\diamond\star}N_{3SG} : \lambda p\lambda q.\forall x[px \rightarrow qx]$
c. every, each := $((S\backslash NP)\backslash((S\backslash NP)/NP_{3SG}))/_{\diamond\star}N_{3SG}$
$: \lambda p\lambda q.\lambda y.\forall x[px \rightarrow qxy]$
d. every, each := $(((S\backslash NP)/NP)\backslash(((S\backslash NP)/NP)/NP_{3SG}))/_{\diamond\star}N_{3SG}$
$: \lambda p\lambda q.\lambda y\lambda z.\forall x[px \rightarrow qxyz]$
e. etc.

In what follows, for example in derivations like (15) and (16) below, the interpretations will on occasion be spelled out as the relevant specific instance. Usually, where the instance is obvious, the syntactic type of raised NPs will be abbreviated as usual as $NP^{\uparrow}$ to save space and avoid visual clutter in derivations.

There is a strong relation between the semantic interpretation of the univer-

1. The use of polymorphic generalized quantifiers avoids the complication of further raising the type of transitive verbs over monomorphic GQs, as in Montague 1973. The $\diamond$ modality is required on all English determiners to prevent crossed composition into them analogous to that permitted for verbs (Hepple 1990; Baldridge 2002).

sal determiners schematized in (5) and its instances (6) and the minimalist' notion of "covert quantifier movement to specifier of vP" (Steedman 2005, 2006). Both give the universal quantifier scope over the entire clause. The difference is that categorial grammars of all kinds achieve this effect statically, via lexicalization.

As in Montague Grammar, such categories distribute correctly over conjunction and disjunction. Thus:

(7) a. Every man walks and talks.
b. $\forall x[man\,x \rightarrow (walks\,x \wedge talks\,x)]$

(8) a. Every man walks or talks.
b. $\forall x[man\,x \rightarrow (walks\,x \vee talks\,x)]$

13.2 Eliminating the Existentials using Generalized Skolem Terms

It is well-known that the non-universals, including plurals, behave quite differently from the universals in terms of scope alternation (*TS*, chapter 3). Most importantly, while they give the appearance of taking wide scope in the weak sense of not being distributed over by or dependent on a commanding universal, they do not do so in the stronger sense of being able themselves to distribute over over an existential that they command as in reading(4b) for (3).

For example, although *at least six hundred languages* can distribute over *some linguist* in (9a), to yield a (plausible) reading in which there is a possibly different linguist for each of the languages, and there is an (implausible) reading in which at least one Stakhanovite linguist has studied all the languages, in (9b) the first reading is not available, despite its greater plausibility:

(9) a. At least six hundred languages have been studied by some linguist.
b. Some linguist has studied at least six hundred languages.

Whereas we have followed Montague in treating *each* and *every* as generalized quantifier determiners, we follow *TS* in treating the non-universals quite differently, as individual-denoting *generalized Skolem terms*.

Standard Skolem terms are widely used in automated theorem proving to eliminate existential quantifiers and their bound variables. They are in general function terms whose *domain* is defined in terms of all variables that are bound by a universal quantifer in whose scope the existential falls. In the case where there are no such universals, the Skolem term is a Skolem constant. Constant generalized Skolem terms give the effect of wide scope existentials, because constants “have scope everywhere”.

All existentially bound variables can be replaced by such Skolem terms, and the existential quantifier itself may then be eliminated.

Generalized Skolem terms further associate a *restrictor* with the Skolem term, which may be an arbitrary property defined as a λ-term. They also carry a *cardinality condition* capturing the counting properties of determiners like least six hundred, and an *index* distinguishing similar Skolem terms arising from distinct noun phrases in minimal pairs like (10).[2]

(10) Some man walks and some man talks $\nvDash$ Some man walks and talks

Thus a generalized Skolem term can be written as follows, where $\mathscr{E}$ is the (possibly empty) environment of universally bound variables, p is the restrictor, c is the cardinality condition, and n is the index:

(11) $sk_{n,p,c}^{(\mathscr{E})}$

In the case of simple indefinites, we can usually suppress c, and n. In the case of generalized Skolem constants, we also suppress the empty environment $()$.

We can now rewrite the two first-order logical forms for (1), *Every boy admires some saxophonist*, that we saw in (2) as follows:

(12) a. $\forall y[boy\, y \Rightarrow admires\, sk_{saxophonist}^{(y)})$

b. $\forall y[boy\, y \Rightarrow admires\, sk_{saxophonist})$

The advantages of this representation will become clear later. For the moment we just note that the two formulæ are structurally identical, differing only in the dependence or non-dependence of the Skolem term. This fact will allow us to eliminate the covert Existential Quantifier Raising that seemed to be required by the traditional first order formulæ (2).

To derive logical forms like those in (12) in a CCG grammar, we need to lexicalize generalized Skolem terms. In the lexicon, generalized Skolem terms are necessarily *unspecified* as to their bound variables, if any.[3].

We do this via the determiner categories, similarly to the universal quantifier determiner (5). Thus, we have:

(13) a/an := $NP^{\uparrow}_{3S}/N_{3S} : \lambda n \lambda p \ldots p\,(skolem\, n)$

2. The fuller account of generalized Skolem terms in *TS* further associates a *polarity marker with them, reflecting the scope of negation operators. We pass over this complication here.*

3. This is a different sense of underspecification from the scope underspecification proposed by Woods (1978), Kempson and Cormack (1981) and much subsequent work in DRT.

Thus, the underspecified translation of *a saxophonist* in (1) can be written $\lambda p.p\ldots(skolem\,saxophonist)\ldots$, (suppressing n and c).

Specification of an underspecified Skolem term of the form $skolem_n\,(p;c)$ involving a property p and a (possibly vacuous) cardinality condition c is defined as an "anytime" operation that can occur *at any point in a grammatical derivation*, to yield a generalized Skolem term obtained as follows.

First, the *environment* of an unspecified Skolem term can be defined informally as follows:[4]

(14) *The environment* $\mathscr{E}$ of an unspecified Skolem term $\mathscr{T}$ is a tuple comprising all variables bound by a universal quantifier or other operator in whose structural scope $\mathscr{T}$ has been brought *at the time of specification, by the derivation so far*.

We can now derive the logical forms in (12) We show the unorthodox left-branching derivation (readers may wish to assure themselves that the same two readings can be obtained from the latter, more standard, right-branching derivation).

(15)

$$
\begin{array}{ccc}
\text{Every farmer} & \text{owns} & \text{a donkey}\\
\hline
NP^{\uparrow}_{3S} & (S\backslash NP_{3S})/NP & NP^{\uparrow}_{3S}\\
:\lambda p.\forall y[farmer\,y \rightarrow p\,y] & :\lambda x.\lambda y.own\,x\,y & :\lambda q.q(skolem\,donkey)\\
\hline
\multicolumn{2}{c}{S/NP : \lambda x.\forall y[farmer\,y \rightarrow own\,x\,y]\quad >\mathbf{B}} & \\
\hline
\multicolumn{3}{c}{S : \forall y[farmer\,y \rightarrow own\,(skolem\,donkey)\,y]\quad <}\\
\cdots\cdots & \cdots\cdots & \cdots\cdots\\
\multicolumn{3}{c}{S : \forall y[farmer\,y \rightarrow own\,sk^{(y)}_{donkey}\,y]}
\end{array}
$$

(16)

$$
\begin{array}{ccc}
\text{Every farmer} & \text{owns} & \text{a donkey}\\
\hline
NP^{\uparrow}_{3S} & (S\backslash NP_{3S})/NP & NP^{\uparrow}_{3S}\\
:\lambda p.\forall y[farmer\,y \rightarrow p\,y] & :\lambda x.\lambda y.own\,x\,y & :\lambda q.q(skolem\,donkey)\\
\hline
\multicolumn{2}{c}{S/NP : \lambda x.\forall y[farmer\,y \rightarrow own\,x\,y]\quad >\mathbf{B}} & :\lambda q.q(sk_{donkey})\\
\hline
\multicolumn{3}{c}{S : \forall y[farmer\,y \rightarrow own\,(sk_{donkey})\,y]\quad <}
\end{array}
$$

In both cases, it is the generalized quantifier determiner categories schematized at (5) that give the universal quantifier-scope over the main predicate q. They therefore have the effect of a restricted form of "covert movement" of the quantifier itself to the specifier position of some functional projection such as vP under the assumptions of the minimalist program (Johnson, 2000). How-

4. A more formal definition is provided in *TS*. The idea is a estricted form of the "Nested Cooper Storage" of Keller (1988).

ever, in present terms, such "movement" does not require action at a distance but is specified at the level of lexical logical form, via λ-abstraction over q in (5). Syntactic derivation merely projects the scope relation defined for the determiner in the lexicon, and the restrictions on scope to be discussed below follow as predictions from the syntactic combinatorics.

The effect of the existential is quite different. Existentials are strictly non-quantificational. In (15), it is important that the logical form for nonuniversal determiners like (13) packs the restrictor inside the generalized Skolem term, rather than predicating it separately as in a standard existential generalized quantifier. In (16), the Skolem term indefinite is a constant, rather than a function term in the bound variable y in its environment.

The present account is to be contrasted on this point with the superficially similar categorial accounts of Bernardi (2002) and Casadio (2004), in which the alternate scopings are distinguished derivationally.

The lack in the present theory of any independent operations of quantifier movement of the kind proposed by May or reanalysis of the kind proposed by Huang and Aoun and Li imposes strong restrictions on scope ambiguities of universals with respect to intensional verbs. For example, under the analysis of raising and expletive *it* in section **??**, the following sentence is correctly predicted to lack any meaning paraphrasable by 'It seems that every/each woman is approaching':

(17) Every/Each woman seems to be approaching.

Similar examples have motivated minimalist claims that quantifiers raise to the specifier of vP (Johnson, 2000).[5]

13.3 Coordination and Quantifer Scope Alternation

TS and earlier publications explore the extensive further implications of the generalized Skolem term account of non-universals, including their interaction with negation.

For example, the following well-known inequalities are predicted, because the Skolem terms corresponding to *a man* and *a fish* necessarily bear the same indices in both conjuncts in the reduced cases and are necessarily distinct in the unreduced cases.

5. Beghelli and Stowell (1997) claim that some of the existentials that are here treated as underspecified Skolem terms similarly to indefinites, such as *at least/exactly five books*, cannot take wide scope in the same way with respect to commanding universals, as derivations analogous to (16) would allow. I do not share their intuitions.

(18) a. A man walks and a man talks $\neq$ A man walks and talks
b. A man cooked a fish and a man ate a fish $\neq$ A man cooked and ate a fish.

Similarly, the fact that the following example from Geach (1970) has only *two* scoped readings, rather than the four that would seem to be allowed by optional covert raising of the two existential quantifiers assumed by a conjunction reduction analysis is also predicted by the fact that Skolem specification must either happen before reduction with the conjoined S/NP ("wide scope" reading) or after ("narrow scope" reading.

(19) Every boy admnires, and every girl detests. some saxophonist.

Thus, the "narrow-scope saxophonist" reading of this sentence results from the type-raised object category (13) applying *before* Skolem specification to *Every boy admires and every girl detests* of type S/NP (whose derivation is parallel to that in (3) of chapter 11), as in (20):

(20)
$$
\begin{array}{cc}
\text{Every boy admires and every girl detests} & \text{some saxophonist} \\
\hline
S/NP & S\backslash(S/NP) \\
: \lambda x.\forall y[boy\,y \to admires\,xy] \wedge \forall z[girl\,z \to detests\,xz] & : \lambda q.q(skolem\,sax) \\
\hline
\multicolumn{2}{c}{S : \forall y[boy\,y \to admires\,(skolem\,sax)y] \wedge \forall z[girl\,z \to detests\,(skolem\,sax)z]} \quad < \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\
\multicolumn{2}{c}{S : \forall y[boy\,y \to admires\,sk_{sax}^{(y)}y] \wedge \forall z[girl\,z \to detests\,sk_{sax}^{(z)}z]}
\end{array}
$$

Since Skolem specification happens *after* the syntactic combination and semantic reduction, both become generalized Skolem terms dependent on the respective quantifiers of the two conjuncts. Each term therefore denotes a potentially different individual, dependent via the Skolem terms $sk_{sax}^{(y)}$ and $sk_{sax}^{(z)}$ on the boys and girls that are quantified over, yielding the narrow-scope reading.

The "wide-scope saxophonist" reading arises from the same categories and the same derivation, when Skolem term specification occurs *before* the combination of *Every boy admires and every girl detests* and the object, when the latter is not in the scope of any operator. Under these circumstances, specification yields a Skolem constant, as in the following derivation, repeated from *SP* in current notation:

(21)
$$
\begin{array}{cc}
\text{Every boy admires and every girl detests} & \text{some saxophonist} \\
\hline
S/NP & S\backslash(S/NP) \\
: \lambda x.\forall y[boy\,y \to admires\,xy] \wedge \forall z[girl\,z \to detests\,xz] & : \lambda q.q(skolem\,sax) \\
 & \cdots\cdots\cdots \\
 & : \lambda q.q(sk_{sax}) \\
\hline
\multicolumn{2}{c}{S : \forall y[boy\,y \to admires\,sk_{sax}y] \wedge \forall z[girl\,z \to detests\,sk_{sax}z]} \quad <
\end{array}
$$

These categories do not yield a mixed reading in which the boys all admire the same wide-scope saxophonist but the girls each detest a different narrow-scope one, as one might expect on a quantifier movement account. Nor, despite the anytime nature of Skolem term specification, do they yield one in which the girls all detest one wide-scope saxophonist, and the boys all admire another different wide-scope saxophonist. Both facts are necessary consequences of the combinatorics of CCG derivation, and require no further stipulation of parallelism conditions. Further cases are discussed in *TS*.

On the assumption that intensional verbs like *seek* include scope-defining intensional operators, a similar absence of mixed *de-dicto* and *de re* readings is predicted for sentences like the following:

(22) Keats wants to date, and Chapman wants to marry, a Norwegian.

13.4 Embedded Subjects and Intermediate Scope

According to the presnt theory, the possibility of unbounded extraction under relativization (23a) and right-node-raising (23b) mean that sequences like *Some woman said that she attended* are typable constituents:

(23) a. A play that some woman said that she attended
b. Some woman said that she attended, and some man said that he missed, the play that I wrote.
c. Some woman said that she attended each play. $(\exists\forall/\forall\exists)$

The prediction that (23c) has an (albeit dispreferred) inverse scope reading, in which the universal distributes over the indefinite, necessarily follows.

However, because English is an SVO language, and the type-raised nominative universal quantifier NP category (6a) is a forward combining category $S/(S\backslash NP)$, it is not possible for an embedded *subject* to invert scope over the matrix subject in the same way:[6]

(24) Some critic wrote that every play was terrible. $(\exists\forall/\#\forall\exists)$

However, under the analysis of infinitival complementation presented in chapter 7, according to which *every play* in (25) is syntactically an argument of the matrix verb, it is correctly predicted to invert scope over the subject *Some critic*:

(25) Some critic considered every play to be terrible. $(\exists\forall/\forall\exists)$

6. The mechanism from the last chapter allowing subject extraction from bare complements of verbs like *say* does not allow scope inversion, since the universal is a lexical NP.

13.5 Distributivity

The possibility of downward distribution of the nonspecific and counting existentials, illustrated in (26), cannot arise from generalized quantifier semantics, since they cannot in general invert scope, as is evident from (27).

(26) a. Some/few/at most two/threeboys ate a pizza.
b. Some/few/at most two/three farmers who own a donkey feed it.

(27) a. A boy ate some/few/at most two/three pizzas.
b. A farmer who owns some/few/at most two/three donkeys feeds it.

This section follows *TS* in arguing that such downward distributivity arises from the verb. In particular, besides having the normal translation (28a), many verbs with plural agreement like *read* have a "distributivizing" category like (28b).[7]

(28) a. read := $(S\backslash NP_{agr})/NP : \lambda x \lambda y.read\,xy$
b. read := $(S\backslash NP_{3pl})/NP : \lambda x \lambda y.\forall w[w \in y \rightarrow read\,xw]$

The above logical form assumes that plurals like *Three boys* translate as set individuals that we can quantify over directly, rather than plural individuals of the kind proposed by Link (1983). In other words, plural generalized Skolem terms are set-valued.

Categories like (28b) are assumed to arise by the application of the following lexical rule to standard (noncollective) verbs, where as usual $(S\backslash NP)/\$$ denotes any member of the set of categories including $S\backslash NP$ and any rightward function into $(S\backslash NP)/\$$ (cf. (50)):

(29) $(S_{-coll}\backslash NP)/\$: \lambda \ldots \lambda y.p \ldots y$
$\Rightarrow_{LEX} (S\backslash NP)/\$: \lambda \ldots \lambda y.\forall w[w \in y \rightarrow p \ldots w]$

In English, this rule is not morphologically realized, but we must expect other languages to mark the distinction, morphologically or otherwise.

The possible occurrence of verbs in English like intransitive *gather* and *meet*, which have only the collective meaning and require set individuals as subject, is also predicted, along with that of the following asymmetry, first pointed out by Vendler (1967) and discussed by Beghelli and Stowell (1997)

7. In invoking a "subordinated" use of universal quantification, this proposal resembles the treatment of distributive nonquantifiers in Roberts 1991; Kamp and Reyle 1993, 326–328; and Farkas 1997.

and Farkas (1997):[8]

(30) a. All/Most (of the)/No participants gathered in the library.
b. #Every/Each/No participant gathered in the library.

In support of this analysis, we might note that "floating quantifier" *each* in the following examples appears to be an adjunct selecting for distributive VPs, despite the lack of an explicit morphological marker:

(31) a. The boys each ate a pizza.
b. The boys ate a pizza each.
c. #The boys each gathered in the library.

Under this account, subjects in examples like (26), besides having a collective reading arising from a set-individual subject undertaking a single act of reading a given book, can optionally distribute over the function that applies to them at the level of logical form, such as $read\,(skolem\,book)$, to yield not only standard forms like (32a), but also (32b):[9]

(32) a. $read\,sk_{book}sk_{boy'\ ;\ \lambda s.|s|=3}$

b. $\forall z[z \in sk_{boy'\ ;\ \lambda s.|s|=3} \rightarrow read\,sk^{(z)}_{book}z]$

Thus, the subject can distribute over more oblique arguments, as in (26) and (33):[10]

(33)

$$\begin{array}{cccc}
\text{Three} & \text{boys} & \text{read} & \text{a book} \\
\hline
NP^{\uparrow}_{3pl}/_{\diamond\star}N_{3pl} & N_{3pl} & (S\backslash NP_{3pl})/NP & NP^{\uparrow} \\
:\lambda n\lambda p.p(skolem\,n\ ;\ \lambda s.|s|=3)) & :boy & :\lambda x\lambda y.\forall z[z\in y\rightarrow read\,xz] & :\lambda p.p(skolem\,book)
\end{array}$$

$$\underline{\qquad\qquad\qquad\qquad}{>}\qquad\qquad\underline{\qquad\qquad\qquad\qquad}{<}$$

$$NP^{\uparrow}_{3pl} : \lambda p.p(skolem\,boy'\ ;\ \lambda s.|s|=3)) \qquad S\backslash NP_{3pl} : \lambda y : \forall z[z\in y\rightarrow read\,(skolem\,book)z]$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$NP^{\uparrow}_{3pl} : \lambda p.p(sk_{boy'\ ;\ \lambda s.|s|=3}))$$

$$\underline{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}{>}$$

$$S : \forall z[z\in sk_{boy'\ ;\ \lambda s.|s|=3}\rightarrow read\,(skolem\,book)z]$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$S : \forall z[z\in sk_{boy'\ ;\ \lambda s.|s|=3}\rightarrow read\,sk^{(z)}_{book}z]$$

8. Similarly, the English floating "A-type" quantifier *each* seems to disambiguate verbs and verbphrases and/or the nounphrases raised over them as the distributive version:

(i) a. Three boys each read a book.
b. Three boys read a book each.
c. #Every boy read three books each.
d. #Three boys each gathered in the library.

9. The connective ";" in the Skolem term is needed because cardinality is a property that applies separately to the *maximal* set of boys reading books that has been identified as the referent of the generalized Skolem term, as in the model theory of *TS*.

10. The relevant subject and object type-raised categories are as usual abbreviated as $NP^{\uparrow}$ to save space and reduce typographical clutter.

Since Skolem specification is a free operation, it can apply early in derivations like the above, to give a third reading, in which a plural subject distributes over a Skolem constant object, so that there are distinct acts of different boys reading the same book.

Further consequences of, and cross-linguistic evidence for, this analysis of downward-distributivity of plural subjects is given in *TS*.

13.6 Other uses of Skolem indefinites

While underspecified generalized Skolem terms can be dynamically specified as dependent or independent individuals, in order to capture apparent scope alternation in indefinites, such terms can also be fully specified lexically to capture cases where scope does *not* alternate.

13.6.1 Intransitivization

It is natural to think that the intransitive sentence (34a) must have the same logical form as the transitive (34b) (Bresnan, 1978:*passim*):

(34) a. Endicott ate.
b. Endicott ate something.

The assumed common logical form could be written in standard first-order logic as the following formula:

(35) $\exists x[ate\, x\, endicott]$

Fodor and Fodor (1980) noticed that such an assumption failed to explain the available interpretations when the subjects in (34) are universally quantified, as in (36):

(36) a. Everyone ate. $\forall\exists/\#\exists\forall$
b. Everyone ate something. $\forall\exists/\exists\forall$

As with (1), (36b) is ambiguous between *two* scoped interpretations, the first of which ($\forall\exists$) is that *Everyone is such that there is something that they ate* and the second of which ($\exists\forall$) is that *There is something such that everyone ate it*. However, as Fodor and Fodor pointed out, (36a) has only the former "narrow-scope indefinite" $\forall\exists$ interpretation: the latter "wide-scope indefinite" $\exists\forall$ reading is not available for (36a).

The Fodors:767 note a further class of intransitive that also fail to show wide scope null intransitive readings, namely certain intrinsically reflexive actions such as *shaving, washing*, etc.:

(37) a. Everyone shaved. $\forall\exists/\#\exists\forall$

b. Everyone shaved someone. $\forall\exists/\exists\forall$

This class of intransitives, which are sometimes referred to as "middle voice", seem more restricted semantically than the *ate* class, in that (37a) seems to entail not merely that everyone shaved someone different, but that it was themself in each case. It is not entirely clear whether these verbs constitute a separate class of reflexive intransitives, or whether all the null indefinite objects of *ate, shaved* etc are indefinites meaning *what the subject normally eats, shaves* etc., which in the case of the "natural" relexives just happens to be themself.

There is a further related class of intransitives that are reciprocal, rather than reflexive,

(38) a. Everyone kissed. $\forall\exists/\#\exists\forall$

b. Everyone kissed someone. $\forall\exists/\exists\forall$

Fodor and Fodor distinguish the null arguments implicit in the intransitives and short passives considered above and those implicit in intransitives and passives like the following:

(39) a. (Endicott dropped the ball.) Everybody noticed.

b. (Endicott dropped the ball.) Everybody was astonished.

Such intransitives require a contextually "given" non-dependent referent, paraphraseable with a pronominal or otherwise definite argument, presumably represented at the level of logical form by a syntactically and phonologically null anaphoric element, as in "prodrop" constructions of the kind discussed by Ruppenhofer and Michaelis (2010).

The Fodors' own explanation for the non-movement of the existential in such intransitives can be seen as a extreme version of the lexicalist position, according to which the existential simply isn't there in the logical form in the first place. The intuition that if *everyone ate* then *they all ate possibly different things* stemmed from an entailment arising from a meaning postulate (Carnap, 1952, 1956) to the effect that anyone who eats, eats something. This meaning postulate applies externally to the grammatical derivation of logical form, in the inferential or model-theoretic component. Thus, the logical forms of intransitive sentence like (34a) and (36a) are the following, according to Fodor and Fodor:

(40) a. $ate\,endicott$

b. $\forall y[person\,y \Rightarrow ate\,y]$

In the case of the intransitive, the meaning postulate in question could in the present notation be written as follows using a free variable y (1980:767):

(41) $eat\,y \vDash \exists x[eat\,x\,y]$

This rule can be interpreted model-theoretically (as Carnap intended), when (via the conjunctive interpretation of $\forall x.Px$) it means that all models or worlds in which someone ate are also models in which there was something that person ate. It can also be presented proof-theoretically (as it is by Fodor and Fodor), via the application of the rules of universal instantiation and generalization. These interpretations are essentially equivalent: in either case, if one hears (34a) or (36a), one can infer (34b) or (36b) as entailments.

However, it is somewhat hard to believe that to understand that "Everyone ate" entails that "Everyone is such that there is something that they ate" we have to do several steps of inference, including universal instantiation, generalization, and introduction, and the enumeration of potentially infinite and/or unknown extensions. The temptation to believe that we access such entailments directly, via the literal meaning of the intransitive alone, persists, despite its so far unexplained restriction to the narrow-scope–indefinite reading.

The present paper proposes a different account of the absence of a wide scope reading for (36a). The claim will be that (36a), "Everyone ate", does in fact have approximately the same logical form as the narrow scope indefinite reading of (36b), "Everyone ate something", just as Bresnan and Dowty (1981) claimed. However, the other wide-scope indefinite reading that *is* available for (36b) cannot be derived from it.

If we can specify non-null indefinites as bound Skolem terms during derivation, we can similarly specify them in the lexical entries for intransitivized objects and passive subjects, as follows:[11]

(42) ate $:= S\backslash NP : \lambda y.ate\,sk^{(y)}_{\lambda x.high\,(prior\,(eat\,x\,y))}\,y$

The property in the null indefinite object Skolem term $sk^{(y)}_{\lambda x.high\,(prior\,(eat\,x,y))}$ in the above intransitive category for *ate* is subtly different from the dependent skolem term $sk^{(z)}_{thing}$ in the narrow-scope reading of the transitive (36b). It identifies the null indefinite as dependent on the verb as well as its subject y, meaning "something that y might be expected to eat".[12]

11. We assume that these categories are produced from the stem via morphology and/or lexical rules whose details we pass over here.

12. It is an advantage of the Skolem representation that the Skolem terms representing null indef-

The derivation of (36a) is the following:

(43)
$$\dfrac{\dfrac{\text{everyone}}{NP^{\rightarrow} : \lambda p.\forall z[person\, z \Rightarrow p\, z]} \quad \dfrac{\text{ate}}{S\backslash NP : \lambda y.ate\, sk^{(y)}_{\lambda x.high\,(prior\,(eat\,x\,y))}\, y}}{S : \forall z[person\, z \Rightarrow ate\, sk^{(z)}_{\lambda x.high\,(prior\,(eat\,x\,z))}\, z]}>$$

In the case of the reflexive intransitive (37a), *Everybody shaved*, we can assign intransitive *shaved* a parallel category to *ate* in (42), giving rise to a derivation for (37a) parallel to (43), and to regard the reflexive anaphor property *self* defined in the next chapter as the pragmatic value of $sk^{(z)}_{\lambda x.high\,(prior\,(shave\,x\,z))}$, “whoever z can be expected to shave”.

(44) shaved := $S\backslash NP : \lambda y.shaved\, sk^{(y)}_{\lambda x.high\,(prior\,(shave\,x\,z))}\, y$

This category will project a logical form onto (37a) that will be in most cases be true in the same models as the sentence *Everyone shaved himself*, via the following derivation.

(45)
$$\dfrac{\dfrac{\text{Everyone}}{NP^{\rightarrow} : \lambda p.\forall z[person\, z \Rightarrow p\, z]} \quad \dfrac{\text{shaved}}{S\backslash NP : \lambda y.shaved\, sk^{(y)}_{\lambda x.high\,(prior\,(shave\,x\,z))}\, y}}{S : \forall z[person\, z \Rightarrow shaved\, sk^{(z)}_{\lambda x.high\,(prior\,(shave\,x\,z))}\, z]}>$$

However, in situations where *everyone* ranges over a set of barbers, the skolem term in (45) could range over a set of customers, rather than themselves.

If we can build dependent Skolem *functions* into lexical logical forms, forcing narrow scope readings for null indefinites, the theory clearly allows us equally to incorporate *non*-dependent Skolem *constants*, forcing wide scope readings.

The examples in (39) seem to constitute a suitable case for treatment. Since we noted that the implicit null element is given or definite, it seems appropriate to include such an element as a Skolem constant, whose property such as $\lambda x.high\,(prior\,(notice\,x\,y))$ whose property will pick out a contextually-available noticeable event as a referent.

(46) noticed := $S\backslash NP : \lambda y.noticed\, sk_{\lambda x.high\,(prior\,(notice\,x\,y))}\, y$

Thus the examples in (39) yield the following logical forms, with the constant

inites can be made dependent in this way not only on a particular argument of the intransitivized predicate, but also on the content of the predicate itself.

sk_{pro} anaphoric to the given event of *Endicott dropping the ball*:

(47) $\forall y[noticed\, sk_{\lambda x.high\,(prior\,(notice\,x\,y))}\, y]$

The transitive verb *gave* briefly considered in the discussion of example (9) in chapter 9 seems to be a related case where the Skolem constant demands a single contextually available recipient as referent. So we have:

(48) gave := $(S\backslash NP)/NP : \lambda x \lambda y.give\, sk_{recipient}\, x\, y$

Thus, we capture the fact that a sentence like the following means everyone gave possibly different donations to the same recipient:

(49) Everyone gave money.

13.6.2 Null Indefinites in the Short Passive

Fodor and Fodor (1980) also noted that if we treat short passives as having the same logical form as the corresponding active, with an null existential generalized quantifier subject like (50b), then we will similarly have to explain the absence of a wide scope subject reading for (50a), according to which there would be a single thing such that it saw every person:

(50) a. Everyone was seen. $\forall\exists/\#\exists\forall$
b. Something saw everyone. $\forall\exists/\exists\forall$

We therefore reformulate the passivized verbs in (50) of section 7.3 as follows, replacing the inert placeholder term *something* by a Skolem term appropriate to the verb concerned dependent on the passive subject:

(51) a. see+en := $VP_{pass} : \lambda x.see\, x\, sk^{(x)}_{\lambda y.likely\,(see\,x\,y)}$
b. persuade+en := VP_{pass}/VP_{to}
$: \lambda p \lambda x.persuade\,(p\,x)\, x\, sk^{(x)}_{\lambda y.likely\,(persuade\,(p\,x)\,x\,y)})$

As with other voices (reflexive and reciprocal) to be discussed in chapter 14, Skolem terms such as $sk^{(x)}_{\lambda y.likely\,(see\,x\,y)}$ meaning "whatever has a high prior probability of seeing x" are introduced by the verbal governor of the construction, in this case via *-en* morphology, as in the morphological derivations in figure 13.1a,b (cf. (18) and (19), chapter 7:

The categories that result represent the underlying agent as a generalized Skolem term dependent upon the lf-commanded x, an assumption that we shall have to make sure is consistent with the binding theory to be developed in a.

$$\dfrac{\dfrac{\text{see}}{VP_{inf,tel}/NP : \lambda x \lambda y.accomplishment\,(see\,x\,y)} \quad \dfrac{+\text{en}}{VP_{pass}\$\backslash\!\!\backslash VP_{inf,tel}\$/NP : \lambda p \lambda y \ldots .p \ldots y\,sk^{(y)}_{\lambda x.likely\,(p \ldots y\,x)}}}{VP_{pass} : \lambda y.accomplishment(see\,y\,sk^{(y)}_{\lambda x.likely\,(see \ldots y\,x)}}{<}LEX$$

b.

$$\dfrac{\dfrac{\text{persuade}}{(VP_{inf,tel})/VP_{to}/NP : \lambda x \lambda p \lambda y.accomplishment\,(persuade\,(p\,x)\,x\,y)} \quad \dfrac{+\text{en}}{VP_{pass}\$\backslash\!\!\backslash VP_{inf,tel}\$/NP : \lambda p \lambda y \ldots .p \ldots y\,sk^{(y)}_{\lambda x.likely\,(p \ldots y\,x)}}}{VP_{pass}/VP_{to} : \lambda p \lambda y.accomplishment(persuade\,(p\,y)\,y\,sk^{(y)}_{\lambda x.likely\,(p \ldots y\,x)})}{<}LEX$$

c.

$$\dfrac{\dfrac{\text{Everyone}}{NP^{\rightarrow} : \lambda p.\forall x[person\,x \Rightarrow p\,x]} \quad \dfrac{\dfrac{\text{was}}{(S\backslash NP)/VP_{pass} : \lambda p \lambda y.past\,(p\,y)} \quad \dfrac{\text{seen}}{VP_{pass} : \lambda y.\exists x[likely\,(accomplishment\,(see\,y\,x)) \wedge accomplishment\,(see\,y\,x)]}}{S\backslash NP : \lambda y.past\,(\exists x[likely\,(accomplishment\,(see\,y\,x)) \wedge accomplishment\,(see\,y\,x)])}{>}}{S : \forall y[person\,y \Rightarrow past\,(\exists x[likely\,(accomplishment\,(see\,y\,x)) \wedge accomplishment\,(see\,y\,x)])]}{>}$$

Figure 13.1:

section (14).

Example (50) is then derived as in figure 13.1c via the category (51a).

Evidence for the lexicalized dependent status of intransitivized objects and subjects of short passives can be found in their notorious inaccessibility to anaphora:

(52) a. Endicott/everyone ate. #It was delicious.
b. Endicott/everyone was seen. #It was John.

The above is essentially the analysis of Dowty (1981), with lexicalized generalized Skolem terms instread of existential quantifiers, consistent with the non-quantificational account of account of alternating scope of non-universals elsewhere in the grammar.

13.6.3 *Tough*-movement and arbitrary control

If we can handle the null indefinite subject in short passives using a Skolem function, then we can also treat the subject in cases of so called "arbitrary" (i.e. contextual reference-based) control with an anaphoric Skolem constant. Thus we can rewrite *tough*-movement predicates like *easy* in (78) of chapter 9 as follows, replacing the earlier placeholder *something* of section 9.10 with a skolem anaphor distinguished by the restrictor *given* to some contextually-given set of referents:

(53) a. easy := $(S_{ADJ} \backslash NP_{XPL})/VP_{to} : \lambda p.easy\,(p\,sk_{given})$
b. easy := $(S_{ADJ} \backslash NP_{agr})/(VP_{to}/NP_{-wh}) : \lambda p \lambda y.easy\,(p\,y\,sk_{given})$

13.7 Generics

It is tempting to believe that generics like those in (54) are simply Skolem constants identical to specifics like those in (55), and that the difference resides solely in the "individual level" (stative) *verbs* in the former and "stage level" (eventive) ones in the latter (Carlson, 1977b; Carlson and Pelletier, 1995).

(54) a. Dogs bark.
b. A dog is a man's best friend.
c. The dog is a carnivore.

(55) a. Dogs are barking.
b. A dog bit a man.
c. The dog ate my homework.

The fact that that universals do not distribute over generics in sentences like the following is consistent with their status as Skolem constants:

(56) a. Everybody likes cats.
b. Cats like everybody

In contrast to Carlson, we assume that there is a distinct specific indefinite plural *cats* with an unspecified logical form $\lambda p.p\,(skolem\,cats)$ that is selected by stage/eventive verbs in examples like the following and can be bound or unbound by universals:

(57) a. Everybody fed cats.
b. Cats attacked everybody.

The difference is supported by the fact that the related generics and specific indefinites are differentially lexicalized in French as "les chats" and "des chats".

13.8 *Same* and *different*

Both indefinite and (less obviously) definite determiners give rise to underspecied skolem terms that may be bound by scoping determiners, differing only in the prosuppositon or otherwise of uniqueness or bridging-inferential givenness in reference (*SP*:51-52):

(58) a. Each man kills the/a thing he loves.
b. Every house on the block has the/a bathroom at the bottom of the/a garden

There is a temptation to view nounphrases like *the same book* and *a different book* in the following sentences as disambiguated with respect to distribution—that is, to be Skolem constant and functional books, respectively.

(59) a. Every boy/three boys read the same book.
b. Every boy/three boys read a different book.

However, this temptation should probably be resisted. As Carlson (1987) has pointed out, the reference of terms like *the same book* and *a different book* is to some extent external to sentence grammar, as is apparent from the following:

(60) a. I bought and sold different books $\neq$ I bought different books and sold different books.
b. I liked and you disliked the same book. $\neq$ I liked the same book and you disliked the same book

We will therefore tentatively assume that *same* and *different* are noun modifiers that add an (in)equality to a contextually given entity with the same property to the denotation.

(61) same $:= N/N : \lambda n\lambda x.nx \wedge z = sk_{\lambda y.given\,y \wedge n\,y}$

(62) different $:= N/N : \lambda n\lambda x.nx \wedge z \neq sk_{\lambda y.given\,y \wedge n\,y}$

13.9 Discussion

The most salient minimalist approaches to quantifier scope alternation seem to be those of Hornstein and Kyle Johnson. Hornstein (1999a) accounts for the scope alternation in (3) in terms of the copy-theoretic version of A-movement to agrO and agrS positions, accounting for scope alternation in terms of which copy of each quantifier is deleted (when the moved copy of the universal and the *in situ* version of the existenial are retained, the inverting ∀∃ reading results. For the other three choices, it is the *in situ* ∃∀ reading that follows). Hornstein also considers an analysis that eliminates agrS and agrO but requires A-*lowering*. Johnson (2000) presents a related account in terms of scrambling rather than A-movement in which quantifiers move to spec of vP. Kratzer (1998), Schlenker (2006), and Szabolcsi (2010) discuss a different use of Skolem terms involving existential quantification over them.

Unfortunately, there are seriously open questions about the actual data on the basis of which these thories could be compared. For example, the choice by Hornstein and Johnson of A-movement and scrambling as the mediating mechanism immediately predicts that scope-taking should be bounded, whereas the mechanism proposed here, like those of Williams (1986) and Reinhart (2006), predicts that the domain of scope-taking for universals should be related to that of unbounded relativization. However, the literature on this point has remained hopelessly divided since Lakoff (1970b)—see *TS*:30-31 for discussion. Because of the apparently inescapable intrusion of common-sense knowledge of the way the world actually works into people's judgements of scoped readings, it seems likely to remain so.

The narrower problem of the scope of null indefinites does not seem to have received much attention within the movement theory since Fodor and Fodor's paper, although Merchant (2013):89 cites them as having shown that intransitivized objects and other indefinite null arguments are "inaccessible" to covert quantifier movement. The question remains: if they are existentials like other indefinites, then *why* don't they move, and how in formal terms can we *prevent* movement from "accessing" them? The lexicalist hypothesis of Dowty and others answers this question.

The present work does so within a framework that also eliminates all quantifier movement, in the following sense.

In the case of the universals, the effect that covert quantifier movement was devised to bring about, namely that of giving a quantified argument scope over its matrix predicate, is accomplished syntactically in categories like those in (6) by morpholexical type-raising over the (potentially unbounded) matrix type, and semantically by a corresponding variable *q* over the matrix logical form.

In the case of the non-universal generalized Skolem terms, such as indefinites, definites, counting quantifers, and other "conservative" quantifiers, the effect of scope alternation is accomplished at the level of logical form, via Skolem specification in terms of binding to all variables bound by operators such as universal quantifiers whose scope the unspecified term falls in *at the point in the derivation where specification occurs*. When there are no such operators, for example, when the specification proceeds any derivation, the Skolem term is a constant, and gives the effect of wide scope, since constants behave as if they "have scope everywhere".

The advantage of the present approach is that constraints on possible scope alternations, particularly those involving English subjects and intermediate scope, and the fact that the plurals, unlike universals, do not invert scope in the strong sense of distributing over commanding indefinites, are emergent from the pure combinatorics of syntactic variation. They are discussed at greater length than is appropriate for the present purpose in *TS*.

Exercise : Derive both scoped readings for (3), *Some boy admires every saxophonist*, using generalized Skolem terms in place of existential quantifiers..

The above account assumes that many English verbs have a separate distributive lexical entry, despite the lack of an explicit morphological marker. Defend (or attack) this assumption on the basis of evidence from another language or languages.

Chapter 14
Anaphora

I found that the reported judgements are often very subtle and highly controversial. I sincerely believe now that much more systematic primary work on establishing a firm data base needs to be done.
—*Preface* to *Binding Theory*, Büring, 2005

Anaphora as reflected in the binding theory of Chomsky 1981 is one of the most fraught areas of grammatical analysis, involving syntax, semantics, and extragrammatical discourse processes (Kuno, 1987; Büring, 2005). The present chapter is confined to a demonstration that some of the more clearly syntax-based cases of anaphor binding can be captured compositionally in the present version of CCG. We have already dealt with the most restricted case of reflexive anaphora in chapter **??**secn:bounded.

14.1 Pronominal Anaphora

Pronominal anaphora is not subject to the kind of syntactic constraints that govern syntactic long-range dependencies, such as island constraints and nesting vs. crossing asymmetries:

(1) a. Every man$_i$ saw a woman who knew him$_i$
b. *Who$_i$ did every man see a woman who knew$_i$?

(2) a. Every man$_j$ thought that every woman$_j$ said that she$_j$ knew him$_j$.
b. Every man$_i$ thought that every woman$_j$ said that he$_i$ knew her$_j$.

While such pronominal dependencies have been investigated in combinatory categorial terms by Szabolcsi (1989), Jacobson (1999) and Barker and Shan (2014), doing so requires the introduction of further operators that threaten to increase expressive power and conflict with the present account of long-range dependencies, and we will not follow their lead here.

Instead, we assume that pronouns are underspecified Skolem terms, like other noun phrases, as in the following category, in which *him* stands for a more specific conjunction of properties such as givenness, as well as masculine gender:

(3) him := $(S\backslash NP_{agr})\backslash((S\backslash NP_{agr})/NP) : \lambda p.p\,(skolem\,him)$

Like other definite referring expressions such as *the boy*, pronouns obtain their referents from a dynamically changing contextual model of the entities under discussion. As in DRT (Kamp and Reyle, 1993), this contextual model includes quantifier bound variables, which are added to the model when the processor enters their scope, and are removed when it exits.[1]

14.2 Donkey pronouns

It has long been noticed that it is impossible to give a translation of sentences like the following using standard first order terms.

(4) Every farmer who owns a donkey feeds it.

Thus, none of the following captures the fact that there are different donkeys, each owned by possibly different farmers who each feed them:

(5) a. $\forall y[farmery \exists x[donkey\,x \wedge own\,x\,y] \Rightarrow \wedge feed\,x\,y]$
b. $\forall y[farmery \Rightarrow \exists x[donkey\,x \wedge own\,x\,y \wedge feed\,x\,y]]$
c. $\forall y[existsx[farmery \wedge donkey, x \wedge own\,x\,y \Rightarrow feed\,x\,y]]$
d. $\exists x[\forall y[farmery \Rightarrow own\,x\,y \wedge feed\,x\,y]]$
e. $\forall y[\forall x[farmer\,y \wedge donkey\,x \wedge own\,x\,y \Rightarrow feed\,x\,y]]$

(Formula (a) corresponds most closely to the form of the sentence but it has the *y* corresponding to the pronoun outside the scope of the existential that binds it. (b) fixes that problem but falsely entails that every farmer owns a donkey. (c) has the same problem as b, while (d) means that there is a donkey that is owned and fed by every farmer. (e), meaning that every farmer feeds every donkey he or she owns comes closest to capturing the meaning of (4), but it is hard to see how it can be built surface compositionally from the sentence structures, and Heim (1990) pointed out that for other quantifiers like *most*, the fact that it essentially quantifies over farmer-donkey pairs means that it seems to hold in models that it should not. For example, in a model in which there are three farmers, one of whom owns many donkeys all of whom he or she feeds, and the other two own one donkey each which they do not feed, it does not seem true to claim the following:

1. A different but related mechanism is discussed in *TS*.

(6) Most farmers who own a donkey feed it.

Much subsequent literature has attempted to square this circle.

The possibility of representing indefinites and bound pronouns as Skolem terms offers a very simple and entirely surface compositional solution. The donkey in (4) is represented by a Skolem term dependent on the universal quantifier, and so is the pronoun, derived by specification of the term in the analogus category (3). Since they are dependent on the same quantifier and in its scope, they can corefer.

(7) $\forall x[farmer\,x \wedge own\,sk_{donkey}^{(x)} \Rightarrow feed\,sk_{it}^{(x)}]$

The derivation is shown in figure 14.1a, from *TS*.

A similar problem is posed by a variant of the donkey sentence first noticed by Postal and Ross, 2009:(66), wrongly claimed by Bozşahin, 2012:91-92 to be unanalyzable under the Skolem term account:

(8) Every farmer's donkey loves him.

The logical form we want is the following, paraphraseable as "Every donkey of a farmer loves him":

(9) $\forall x[donkey\,x \wedge own\,x\,sk_{farmer}^{(x)} \Rightarrow love\,sk_{him}^{(x)}\,x$

To obtain this logical form surface compositionally, we need the following categories for possessives, which solve a more general problem for natural language semantics noted by Pulman (2013)

(10) 's $:= (N\backslash N)/N : \lambda n \lambda m \lambda x.n\,x \wedge own\,x\,(skolem\,m)$
of $:= (N\backslash N)/NP : \lambda m \lambda n \lambda x.n\,x \wedge own\,x\,m$

The derivation of (8) is then as as shown in figure 14.1b.

See *TS* for a more extensive account of donkey anaphora.

14.3 Logophoric Anaphora

Some languages exhibit a distinct third "logophoric" variety of pronominal, referring to a non-clausemate higher argument, usually a subject, denoting a protagonist or "point of view". (For example, Dutch and many other Germanic languages recognize a distinction between pronouns like *hem* ("him"), reflexive bound anaphors like *zichzelf* ("him/herself"), and an unbounded logophor *zelf* (Reinhart and Reuland, 1991; Pollard and Sag, 1992). Of course there is nothing to stop a language having the same proform be ambiguous between

Every farmer who owns a donkey feeds it

$$\text{Every} := (S/(S\backslash NP_{3SG}))/_{\diamond\star}N_{3SG} : \lambda n\lambda p.\forall x[nx \rightarrow px]$$

$$\text{farmer} := N_{3SG} : farmer$$

$$\text{who} := (N_{agr}\backslash_{\diamond\star}N_{agr})/_{\diamond\star}(S\backslash NP_{agr}) : \lambda q\lambda n\lambda y.ny \wedge qy$$

$$\text{owns} := (S\backslash NP_{3SG})/NP : \lambda x\lambda y.own\,xy$$

$$\text{a donkey} := (S\backslash NP)\backslash((S\backslash NP)/NP) : \lambda p.p(skolem\,donkey)$$

$$\text{feeds it} := S\backslash NP_{3SG} : feeds\,(skolem\,it)$$

$$\overline{S\backslash NP_{3SG} : \lambda y.own\,(skolem\,donkey)y}^{<}$$

$$\overline{N_{3SG}\backslash N_{3SG} : \lambda n\lambda y.ny \wedge own\,(skolem\,donkey)y}^{>}$$

$$\overline{N_{3SG} : \lambda y.farmer\,y \wedge own\,(skolem\,donkey)y}^{<}$$

$$\overline{S/(S\backslash NP_{3SG}) : \lambda p.\forall x[farmer\,x \wedge own\,(skolem\,donkey)x \rightarrow px]}^{>}$$

$$\cdots\cdots\cdots\cdots$$

$$S/(S\backslash NP_{3SG}) : \lambda p.\forall x[farmer\,x \wedge own\,(sk^{(x)}_{69:donkey})x \rightarrow px]$$

$$\overline{S : \forall x[farmer\,x \wedge own\,sk^{(x)}_{donkey}x \rightarrow feeds\,(skolem\,it)x]}^{>}$$

$$\cdots\cdots\cdots\cdots$$

$$S : \forall x[farmer\,x \wedge own\,sk^{(x)}_{donkey}x \rightarrow feeds\,(sk^{(x)}_{it})x]$$

b. Every farmer 's donkey loves him

$$\text{Every} := (S/(S\backslash NP))/N : \lambda q\lambda p\lambda x.\forall x[qx \Rightarrow px]$$

$$\text{farmer} := N : farmer$$

$$\text{'s} := (N\backslash N)/N : \lambda n\lambda m\lambda x.nx \wedge own\,x\,(skolem\,m)$$

$$\text{donkey} := N : donkey$$

$$\text{loves} := (S\backslash NP)/NP_{pro} : \lambda a\lambda y.loves\,a\,y$$

$$\text{him} := NP_{pro} : \lambda p.p\,(skolem\,him)$$

$$\overline{N\backslash N : \lambda m\lambda x.donkey\,x \wedge own\,x\,(skolem\,m)}^{>}$$

$$\overline{S\backslash NP : \lambda y.loves\,(skolem\,him)\,y}^{<}$$

$$\overline{N : \lambda x.donkey\,x \wedge own\,x\,(skolem\,farmer)}^{<}$$

$$\overline{S/(S\backslash NP) : \lambda p.\forall x[donkey\,x \wedge own\,x\,(skolem\,farmer) \Rightarrow p\,x]}^{>}$$

$$\overline{S : \forall x[donkey\,x \wedge own\,x\,(skolem\,farmer) \Rightarrow loves\,(skolem\,him)\,x]}^{>}$$

$$\cdots\cdots\cdots\cdots$$

$$S : \forall x[donkey\,x \wedge own\,x\,sk^{(x)}_{farmer} \Rightarrow loves\,sk^{(x)}_{him}\,x]$$

Figure 14.1:

pronoun and anaphor. This is arguably the case for English, which allows "logophoric" homophones of reflexives such as *himself* in examples like the following (Pollard and Sag 1992):

(11) [Near him$_i$/himself$_i$], Frank$_i$ saw a snake.

The Japanese proform *zibun* may be a parallel case. Such unbounded homophones of reflexives could also be specialized to particular verbal inflections.[2]

It is important to realize that it is the logical form in English lexically cliticized verb categories like (60) that ensures the reflexive binder is always local, in conformity to Condition A. We assume that it is the dynamics of the contextual model supporting definite reference that ensures that the pronominal binder is never local, in conformity to Condition B, and never c-commanded, in conformity to Condition C. Specifically, we assume that the pronouns accessible in a given verbal domain are assigned possible referents *before that domain is entered*. Thus, no co-argument nor argument of a commanded domain can be among those preassigned referents.

Thus, we follow Jacobson (2004, 2007), Reinhart (1983a,b, 2006) and Büring, 2005:122-4 in assuming that all binding conditions can be eliminated from the grammar.

Steedman and Bozşahin (2016) argue that all languages including ergative ones uniformly exhibit a binding theory defined in terms of local command relations at the level of logical form, at which actors/agents command other arguments (Bach and Partee 1980; Manning 1996), as in the predicate-argument structures introduced in chapter 2.

14.4 Null Anaphora

We assume that pro-drop is lexicalized with categories in which the dropped argument is represented by a generalized Skolem constant sk_{pro}, whose restrictor *pro* identifies it as a contextually accessible discourse entity, and which is otherwise indistinguishable from the result of combining a clitic pronoun with an agreeing verb category. For example, Welsh *gwelest*, "(thou) sawest" has the following transitive subject-pro-dropped category:

(12) gwelest := $S/NP : \lambda x.saw\,x\,sk_{2s,pro}$

2. Icelandic appears to offer a case in point, where such an unbounded homophone of a reflexive is restricted to subjunctive mood (Thráinsson, 1990:290).

14.5 Discussion

As in the case of available scoped quantifier readings, only more so, there are real uncertainties as to the data that are to be accounted for (Büring, 2005). The extent to which bound variable pronominal anaphora in particular should be handled by purely grammatical mechanisms, or whether it is better viewed as a form of discourse anaphora, remains unclear. In English, it is also frequently hard to decide whether reflexive pronouns are true reflexives, or logophoric pronouns. The present chapter makes no stronger claim than that some of the simpler forms of anaphora can be accomodated within the present theory. Steedman and Bozşahin (2016) discuss the cross-linguistic applicability of a related account.

Chapter 15
Conclusions

Well, less is more, Lucrezia: I am judged.
—*Andrea del Sarto* Robert Browning, 1855

Part IV

Appendices: CCG in Application

There seems to be little reason to question the traditional view that investigation of performance will proceed only so far as understanding of underlying competence permits.
—Noam Chomsky, *Aspects of the Theory of Syntax* (1965:ch.1,§2)

Appendix A
The Evolutionary Emergence of Language

you cannot learn a language whose terms express semantic properties not expressed by the terms of some language that you are already able to use.—Jerry Fodor, *The Language of Thought* (1975:61)

The present appendix argues that the faculty of language comes essentially for free in evolutionary terms, grace of a capacity shared with some evolutionarily quite distantly related animals for deliberatively planning action in the world. The reason humans have language of a kind that animals do not, is because of a minor qualitative difference in the nature of human plans, rather than anything unique to language.

A.1 Views of Language Evolution

This book began with the commonplace observation that, while human languages appear to be very diverse in form, human children can nevertheless acquire any of them in roughly the same amount of time in interaction with their speakers, No other animal, not even those closest to ourselves in evolutionary terms, appears to use or be able to acquire any communicative system of comparable productivity. Since the divergence of the human line from that of the chimpanzees occurred only around 6M years ago, its evolution seems to have been very rapid indeed, despite its singular nature.

Two kinds of explanation for the rapid evolutionary development of language seem to be on offer.[1]

The first kind of explanation is based on the idea that the tree-like structures that are characteristic of human language, and which have been claimed to distinguish it from the kind of symbol systems that animals actually *can* learn, predate the existence of language itself in evolutionary terms on a much larger timescale. The evolution of language itself is then presumably seen as a com-

1. Narratives of a third kind, appealing to universal grammar and asserting the singular nature of language and its divergence from all other forms of animal cognition and communication in terms of a language instinct or organ, via notions like emergence and saltation, and the biolinguistic nature of the enterprise (Di Sciullo and Boeckx 2011), seem to be essentially methodological, combining a useful restatement of the problem with advocacy of an antireductionist methodology relying primarily on linguistic data. In terms of actual explanations, its adherents generally seem to fall into one of the two following camps.

paratively simple attachment of word-like concepts to a prexisting structural component.

Jespersen, following Rousseau (1781) and Darwin 1871, seems to have believed something like this when he fancied the first utterances of speech to have been “like something between the nightly love-lyrics of puss upon the tiles and the melodious love-songs of the nightingale” (1922:434).

Much recent work on structure and learning of song in birds seems to have the related aim of showing that systems of communication including recursive or treelike structure can evolve independently of human language, and therefore offer a possible precursor to it (e.g. Jarvis 2004; Hilliard and White 2009; Fitch 2014).

The other kind of explanation is based on the assumption that there is a preexisting non-linguistic, non-communicative, but nevertheless symbolic homomorphic conceptual representation, onto which articulated communicative language can be hung via its semantics. Since individual languages differ considerably not only in their word order, but in the way they carve nature at the joints semantically, as partly revealed by their morpholexicons, that universal conceptual representation must be independent of linear order, and of a degree of ramification or elaboration which will support all of those different world views. (In other words, different grammatical devices across languages may correspond to different but overlapping substructures of this conceptual representation.) Early exponents of this view were Pinker (1979) and in very different terms, Wexler and Culicover (1980).

One version of the idea is implicit in the suggestion of Hauser, Chomsky, and Fitch (2002):1578 that structural recursion may have evolved for non-linguistic purposes of navigation and “number quantification,” and might possibly be shared by other animals, a position that is not unlike the present one.

These two styles of explanation at first sight seem very different. The first emphasizes the primacy of structure in evolution and acquisition. It seems to depend on the assumption that unsupervised learning of grammars from strings alone is possible.

The latter assumption is questionable. It may be possible to learn finite inventories of candidate phones and morpheme boundaries from mere exposure to native speech (although even that is not trivial, and depends on prelinguistic auditory structure Barlow, 1961). However, learning natural language grammars from exposure to nothing but the strings of the language remains an open problem in computational linguistics on which very little progress has been made, despite considerable effort over many years. (In particular, unsupervised-learned parsers lag far behind the performance of supervised parsers trained on treebanks, to such an extent that the best unsupervised parsers can be beaten by supervision trained on as few as fifty syntactically annotated sentences, or a very small number of syntactic rules (Naseem, Chen, Barzilay, and Johnson, 2010).)

The second style of explanation emphasises the primacy of cognition. It has the advantage that we know that it works. If we assume that something homomorphic to the semantics is available to the child, then we are talking about *supervised* learning, analogous to the induction of parsers from treebanks, which is the way that all currently successful wide coverage parsers are induced (Charniak 1997; Collins 1997). (The term "supervised" does not of course imply the presence of a human supervisor or instructor, but refers rather to the availability to the child of information concerning the correct analysis.)

The child's task of learning a grammar from strings and meaning representations in a homomorphic "language of the mind" is of course a little harder than inducing a parser from the Penn treebank of newspaper text annotated with syntactic trees. The child has to also discover the alignment of the structures of the meaning representation with the words of the target language. Doing all this in a single incremental pass through the data involves building and continually updating a parsing model for all possible languages that are consistent with the data observed so far, rather than just a modelfor string-aligned trees, as outlined in the preceding apppendix to the present one, following Kwiatkowski, Goldwater, Zettlemoyer, and Steedman, 2012 and Abend et al., 2017.[2]

These two views are not as different as they might seem at first glance. As Jespersen realized, the view that structure precedes language necessarily assumes a mechanism for correctly combining the words that are supposedly attached to preexisting structures, projecting them in a way that supports compositional semantic interpretation. Jespersen's proposal was that these birdsong-like structures were initially attached to complex meanings as analyzed holophrastic labels, and were later decomposed into the phrasal and word-level units of the target language. This process is of course only possible if the meaning representations themselves have homomorphic structure, as assumed under the other view, that meaning precedes language specific syntax,

2. I refer to the homomorphic conceptual substrate as a "language of mind" because Fodor defines his "language of thought" as the language-specific logical form of the language, essentially isomorphic to its syntax and lexicon. The language of thought in Fodor's sense must be assembled from a more elaborated and ramified homomorphicly related universal language of mind during child language acquisition.

and it is in fact in broad outline the way that the variational Bayes model of the preceding appendix works in practice.[3]

As another example of the interdependence of the structure-first and meaning-first views, it is interesting to consider George Miller's 1967 report on *Project Grammarama*. Miller makes the following remark in the conclusion to his chapter, concerning the nature of the information necessary for successful language acquisition:

> "When skilled behavior can be analyzed into independent responses, either overt, or covert, that can be reinforced individually, and assembled without significant interaction, the principles of learning derived from conditioning experiments may be applicable. Independent components, however, are not characteristic of rule-guided human behavior, and the systemic aspects cannot be avoided. Under those conditions, therefore, it seems reasonable to assume that *the feedback must convey information at least as complex systemically as the rules to be learned*." (Emphasis added.)

A similar point is made by Fodor (1966, 1975) in the epigraph to this appendix: you can't learn a language unless you already know an equally expressive language (1975:64).

A.2 Plans and the Structure of Mind

If you can't learn a language unless you already know an equally expressive language, where does that preexisting language come from? If we are to escape infinite regression, we know it cannot by definition be learned.

However, a language of mind can be evolved. Learning has to be done with the bounded resources of individual finite machines. Evolution has virtually unbounded resources, with numbers of processes limited only by the physical resources of the planet, and processing time limited only by the lifetime of the sun. It essentially works by trying every possible variation on every viable variation so far.[4]

3. Jespersen also believed that children could do this for themselves, provided there was more than one of them and they could stay alive, a position for which he offered anectdotal evidence from the greater diversity of Native American languages in the benign Californian region in comparison to the hostile North American Arctic, and from certain rare cases of language development in isolated twins.

4. It is a little more complex than this. Even genomes must be structured programs, and even evolution needs the occasional mass-extinction to escape overfitting.

The only plausible source for the this language of mind is as a result not merely of recent primate evolution, but rather of half a billion years of chordate evolution, resulting in a symbolic language of mind, grounded in physical existence in the world, most of which we must share to varying degrees with our animal cousins.

This observation raises the question of why, in that case, even our closest animal relatives (who must share almost all of this precursor) show no signs of being able to learn anything like natural language as a formal system. The answer is of course that there must have been *some* evolutionary advance, but it must be quite minor to have arisen so quickly. Most of what language specifically depends on must have arisen previously for quite different evolutionary reasons, probably multiple times, and be simple enough to arise simultaneously in multiple individuals, since that is the way that evolution works.

From this point of view, it seems rather unlikely that the crucial advance could have been anything as complicated and singular as the sudden introduction *de novo* of devices specific to language, such as trees, recursion, or the related "merge" operation of the Chomskian minimalist program, as proposed most recently by Berwick and Chomsky (2016). Such mechanisms must be more ancient, evolved over a very long period for a more cognitively general purpose which we share with some animals, such as deliberative planning of action in the world (Miller, Galanter, and Pribram 1960; Steedman 2002). The widespread failure on the part of psychologists to show that animals can learn externalized recursive concepts should not be taken as evidence against this claim, for the simple reason that it is extremely difficult to prove that even *human language* is recursive on the basis of finite stringsets alone. It is only our intuitions about the underlying *semantics* that makes its recursivity obvious.

Far from being something unique to it, the specific advance that supports human language must be something that evolution can come up with *easily*. And if it is easy, then it is likely to be something familiar, something that evolution *already has* come up with repeatedly, which if added to a mix of other traits from our immediate relatives, including recursive processes evolved over a much longer period, will provide a conceptual base to which human language is homomorphic.

One plausible candidate is cooperative planning (Tomasello 1999, 2009), in particular the service of cooperative raising of offspring (Burkart, Hrdy, and Van Schaik 2009; Hrdy 2009), both of which have arisen a number of times in mammalian evolution, though apparently not among apes other than ourselves and at least some of our hominid ancestors. Dunbar (1992) shows that size of the front-brain or neocortex in contemporary primates is highly correlated with the size of the social group with which a typical individual interacts, and suggests on this basis that the main selective force driving evolutionary increase in primate intelligence arises from social interaction—see Maynard-Smith and Szathmáry, 1995:276-8 for discussion. If so, then this component of the language of mind developed in primates over millions of years, rather than hundreds of thousands

We will not rehearse these arguments here, except to note that the kind of cooperation they invoke is specifically deliberative planning involving calculation about other minds, rather than mere collective action that reinforcement learning or evolutionary selection has selected as working better when there are multiple agents, such as hunting in packs or flocking behavior.

Deliberative planning—in particular, planning involving tools—is an ability that we do share with other species, notably chimpanzees, but also elephants and some of the more recently evolved birds. Chimpanzees really can solve the so-called "monkey and bananas problem" using such tools as crates in stacks (Köhler 1925). This kind of planned action is quite unlike the undirected reactive behavior that results when Skinnerian shaping with reinforcement is used in attempts to get pigeons to solve monkey-and-bananas problems.[5]

We can think of planning as the mental construction of a sequence of actions that will bring about a desired goal state. Such planning involves:

1. Retrieving appropriate actions from memory (such as piling boxes on top of one another, and climbing on them),
2. Sequencing them in a way that has a reasonable chance of bringing about a desired state or goal (such as having bananas).
3. Remembering plans that succeed, for re-use on future occasions.

Köhler showed that, in apes at least, such search seems to be *reactive* to the presence of the tool, and to be *forward-chaining*, working forward from the tool to the goal, rather than backward-chaining (working from goal to tool).

The first observation implies that actions are accessed via perception of the objects that mediate them—in other words that actions are represented in memory *associatively*, as properties of objects—in Gibson's 1966 terms, as *affordances* of objects.

The second observation suggests that in a cruel and nondeterministic world it is better to identify reasonably highly valued states that you have a reasonable

5. See YouTube for examples.

chance of getting to than to optimize complete plans.

The problem of planning can therefore be viewed as the problem of *Search* in a "Kripke model" or directed lattics of states of the world for a sequence of actions or affordances α; β, etc. leading to a goal state: The problem of

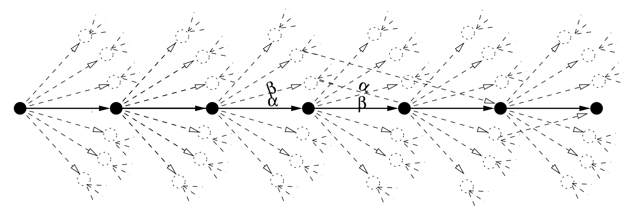


Figure A.1: S4/Kripke Model of Causal Accessibility Relation

searching a lattics, in which different action-sequences may lead to the same state, is isomorphic to the problem of the algorithmic component we saw in an earlier appendicx on parsing. The only difference is that in parsing, the actions α; β, etc. on the arcs correspond to the application of rules of the grammar. (These rules might be at any level of the language hierarchy of figure 12.3.1 of chapter **??**, from finite-state to context-free and beyond. Thus, this is a very general mechanism

Of course, to plan successfully in the real world, one also needs a model of utility and probability of success for α, β, and the ability to replan in real time when things go wrong.

In parsing, we saw in the earlier appendix that one similarly needs an oracle or parsing model, to assign a value or a probability

Causal actions/affordances in plans can be represented in search-efficient logics of change, such as STRIPS operators (Fikes and Nilsson, 1971), or the related operators of MICROPLANNER (Sussman and Winograd, 1970) or their modern descendant PDDL (McDermott, Ghallab, Howe, Knoblock, Ram, Veloso, Weld, and Wilkins, 1998). Such operators specify preconditions on their execution and the state-changes that they effect. For example, the action of *exiting* via a door has the precondition that you are *in* and that the door is *open*, and has the effects that you are no longer in, but *out*.

More formally, such operators are *functions* from (partial) state representations to state representations. Planning is then the *composition* of the affordances of the objects in the situation (including the agent) to form a new function that results in the goal state. That is to say that planning involves the operation we have seen in syntax as the combinator **B**. For example, if the door is not open, but *shut*, then your plan for getting *out* is the composition of *pushing* the door (which makes it change from shut to *open* and *exiting*.

This implies that the objects themselves should not be thought of as inert arguments of actions, but as *functions* from their affordances—the actions that can apply to them–to the results of those affordances.

For example, one affordance of a door when it is *shut* is *pushing*, which if you apply it to the door, results in a state where the door is *open*. That is to say that the planner requires objects to be *type-raised*, the operation that we have seen in syntax as the morpholexical case operator corresponding to the combinator $\mathbf{T}$ or $\lambda x \lambda p.px$.

Thus, the basic operations of deliberative planning already include the two major families of operators we have seen in language in projective morphosyntax, and offer a precursor found in animals to the language faculty. So we are still left asking, why don't the animals have productive language?

There are some significant differences observed by Köhler between the kind of planning that apes and humans are capable of that it will be useful to bear in mind in considering this question. While apes can form quite complicated plans involving multiple tools (such as towers of many crates), their doing so depends on the perceptible availability of the tools (such as the crates). Plans which require going to another room to *fetch* crates, even if they have in the recent past been observed there, are much harder for the animal to attain than for humans. The apes also show deficiencies in comparison with humans in their abilities to make plans involving other minds (although they clearly have the concept of other-awareness).

There is an interesting connection between these two kinds of planning difficulties in apes. If I want to visit you, and I don't know where you live, then a very bad plan is to start knocking on doors until the person who answers is you. A better plan involves the use of a variable we might call *address*, making a plan to set *address* to your address (say by getting out my phone and looking it up on my address book) and then going to *address*, wherever that is, and knocking on the door there. Or if your address isn't on my phone, then I could use my phone to call one of your friends to ask them for it. The first alternative involves obtaining a non-present tool (my phone). The second involves the use of a non-present other mind.

Thus, tools and other people seem rather similar in terms of human planning. It is an open question in evolutionary and developmental terms whether the use of non-present tools precedes and is a model for the use of agents and other minds, or whether it is the other way round, with evolved cooperation modeling tool use and manufacture.

There are some interesting limitations in the use of tools by apes that make it seem possible that the latter is at lest in part the case in humans. Apes seem good at making plans like $place\,box_1; climb\text{-}on\,box_1$ ("place that box and climb on it"), but specifically bad at finding plans that call for "execution-time variables" whose value is not known at the planning stage, but which will be established by executing the plan. An example of such a variable is x in $find\,x \wedge affords(climb\text{-}on\,x)\,x; place\,x; climb\text{-}on\,x$ ("find something that affords climbing, place it, and climb on it"), and $find\,x \wedge affords(ask\,x\,y)\,x; call\,y$ ("find someone to ask the number, then call it").

The distinctive character of the latter kind of plan is that they are *functions over (possibly multiple) tools and/or agents such as people*, whose result is a plan of the first kind. with the variables instantiated by values, such as specific persons or boxes. The number of variables in plans like "persuade someone to find someone to authorize someone to send me the account number", and hence the valency of the corresponding function, seems to be essentially unbounded.

Such plans are semantically recursive, and show up in recursive structures in natural languages in sentences like the following, which were discussed in chapters **??**, 2, 9, and **??** (indices associate verbs with their nominal arguments:

(1) a. that we_1 let_1 the children_2 help_2 Hans_3 paint_3 the house_3.
b. daß wir_1 die Kinder_2 dem Hans_3 das Haus_3 streichen_3 helfen_2 lassen_1.
c. das mer_1 d'chind_2 em Hans_3 es huus_3 lönd_1 hälfe_2 aastriiche_3.

In English (1a) this construction is tail-recursive, but in German (b) (simplifying somewhat) it is center-embedding, while in Zurich German (c) it is center-embedding with crossing dependencies. In the latter case, it is known that such languages are non-context-free (Shieber, 1985).

The language of fully instantiated plans is itself context-free, and we noted in chapter 2 that we must seek the reason for the narrowly trans-context-free character of natural grammars to other factors, such as the infomation-structural need to keep subject and tense constituent-initial.

The extension to LI languages requires the extension of the syntactic operators defining the search space for the parser beyond mere application and composition of the kind that are needed for even planning in a finite state language of plans, to include at least second-order composition $\mathbf{B}^2$, rules (**??**) in chapter 9.

We may therefore conjecture that the following progression, spanning a couple of hundred million years of mammalian evolution, provided the necessary substrate for the essentially instantaneous subsequent development of human language and the other cognitive faculties we have mentioned.[6]

1. Reactive planning using Piagetian circular reactions or "while loops", of the kind seen in pigeons and rats, defines a corresponding finite-state language of plans. Such planning may or may not involve search, the composition of actions or state-change functions (Tolman, 1948; Hull, 1943).
2. Deliberative planning with actions including use of available tools, of the kind found in apes and other animals, requires recursive search to some drastically limited depth through a state-space for a sequence of actions that results in a goal-state. Since different sequences of actions may lead to the same state, the state-space can be thought of as a lattice, with the states as nodes, and actions as directed arcs between states. Such search involves composition of state change function, and type-lifting of tools to yield the actions they afford, and is computationally homomorphic to the search problem involved in parsing human language, although the ape language of plans is still finite-state.
3. The evolution of human cooperative behavior under such pressures as neotenous child-rearing and size of social group requires plans involving other agents, and recursive concepts such as obligation, and information structural concepts such as common-ground, given and new. Such agent-based planning also supports tool manufacture.
4. The semantics of information structure requires alignment of referential items with the temporal thread of discourse, which may require semantic dependencies to cross, taking the syntactic component of natural grammars to the level of the linear indexed languages, expressible in CCG.

A.3 Conclusion

The only working model that we have for the acquisition of natural language grammar by children requires them to have access to a structural representation that is homomorphic to the representational level of their first language(s). The process of language acquisition then corresponds to the attachment of that language to this prexisting content-free armature, a process that may result in a

6. I do not intend to suggest that parts of this progression may not have occurred in non-mammals. Observation of tool use in species of Crow and Parrot Emory and Clayton (2004); Huber and Gajdon (2006) suggest that partial parallel evolution may have occurred inbirds.

grammar that is slighlty more expressive than context free. The question of the evolution of human language is then the question of how that homomorphic representation language arose.

This appendix has argued that it is there because it constitutes a preexisting symbolic language, using the same inventory of combinatory operator types, but having its origins in sensory-motor planning, of a kind that we share much but not all of with chimpanzees. This language of mind constitutes the substrate to the semantics of human languages, and is one of the inputs to child language acquisition. The reason other animals don't show evidence of language arises from a qualitative difference in their plans. Human plans can deal with situations that are not that of the phenomenal sensory motor present, and with actions upon other minds that change knowledge. These show up formally in the planning language as execution-time variables, and seem to have their origin in an evolved propensity for cooperation with other minds, of the kind hypothesized by Tomasello, possibly arising from pressures of human child-rearing, of the kind discussed by Hrdy.

The question of when this evolutionary progression occured is of course one for which we can have no firm evidence (Lewontin, 1998). A considerable section of the literature speculates that it was very recent, around 100,000 years ago in the upper neaolthic period, which is the earliest that we have evidence for other kinds of symbolic activity, such as painting and music.

However, unless one believes that painting and music are inevitable and immediate concomitants of language, there is no evidence for such recency in the emergence of language itself. It seems likely in fact that for paintings and musical instruments that will last a hundred thousand years for us to find, one might need very substantial economic resources, such as suffient surpluses of media like oil, eggs, and bone marrow, and suffient spare labor capacity to collect and prepare suitable pigments and instruments, and to allow members of the group to spend their time on such non-essential activities as painting and musical performance. It might have taken a very long time for human society to progress to the necessary level of material security. And there is another possibility.

We have noted that the kind of planning that is necessary for cooperation with other human agents is identical to the kind of human planning with tools that are not immediately available. The evidence for the use of manufactured stone tools is much older than the evidence for painting and music. Shaped stone tools are found continuously in conjunction with skeletal remains dating from the transition between australopithicines and the first hominids. two to three million years ago. That fact in itself does not tell us whether the early forms of *Homo* had language. However, there is evidence from reconstruction of the vocal tract of the successive forms of hominid that there was a continuous lengthening of the tract and lowering of the larynx towards the extreme form found in modern humans (Lieberman, Laitman, Reidenberg, and Gannon, 1992; Lieberman and McCarthy, 2007). While there are other possible evolutionary explanations for such a change (Fitch and Reby, 2001; Bolhuis, Tattersall, Chomsky, and Berwick, 2014), it is suffiently maladaptive in humans for every other function *except* vocal language (functions such as swallowing food without choking) that it seems likely that it arose from selection for the advantages of primarily vocal form of language over original primarily gestural forms, as Deacon has argued (1997:354-365).[7]

It therefore seems possible that there was a single evolutionary transition, contemporary with the earliest emergence of the hominid line, possibly based on a change to cooperative rearing of infants, under an evolutionary presssure that remains obscure, but of a kind that has quite frequently occurred in other primates (Hrdy, 2007) and mammals such as elephants. This change in hominids required a kind of planning that immediately supported both language and tool-manufacture, and initated the slow ascent of human culture to the point where it could support civilization, including enduring art and music.

7. This is not to argue that the original gestural form was in any way less than full human language, any more than one would argue that for modern sign languages.

Appendix B
Child Language Acquisition

I hafta pee-pee just to pass the time away.
—"Eve" (*A First Language* Roger Brown, 1973:15)

It is generally agreed that, in learning such basic aspects of language-specific grammar as which words of the language are the verbs and which the nouns, and in what linear spatio-temporal order(s) the two may occur, children must have access to something more than mere strings of words constituting a subset of the legal sentences of the languages.

This agreement is based in part on observation of the extreme rapidity with which language acquisition proceeds, and the absence of negative data in the input to the child. While it is theoretically possible, using probabilistic models and unsupervised machine learning, to approximate grammars of linguistically relevant classes to any desired degree of accuracy (Horning, 1969), the computational costs of such learning for realistic grammars are prohibitive, and there has been little success so far among computational linguists in achieving practical unsupervised induction of natural language grammars from positive data alone. This is the so-called "logical" problem of language acquisition (Baker, 1979, *passim*).

There is much less agreement concerning the actual nature of the "something more" that the child brings to the task. It is sometimes referred to as "Universal Grammar", and as such is sometimes talked about in exclusively syntactic terms, as in the "parameter-setting" account of language acquisition of Chomsky (1981), Hyams (1986), and much subsequent work. According to this account, exposure to "trigger sentences" (made up of words whose meanings and parts of speech have already been learned) "flips switches" (Chomsky, 1986b:146; Pinker, 1994) corresponding to syntactic parameters such as head-finality and *pro*-drop until the "universal grammar engine" uniquely specifies the language, in a process that has been likened to a game of Twenty-Questions (Yang 2006:Ch.7).

Such accounts raise as many questions as they answer about the mechanism by which such learning could proceed. In particular, the specific inventory of parameters that this universal machine embodies, the way in which the very large search spaces engendered by even quite small sets of binary independent parameters can be effectively explored (Clark and Roberts 1993; Fodor and Sakas 2005), and the aspects of the data that "trigger" their setting (Gibson and Wexler 1994; Fodor 1998) remain rather unclear (see Newmeyer, 2005:ch.3 for a review, and cf. Haspelmath, 2008).

Nevertheless, there is something deeply appealing in the idea that the process of language learning proceeds by entertaining all grammatical possibilities, and eliminating alternatives, because that is pretty much what the child's developmental behavior looks like. In particular, Crain and Thornton (1998) and their students have shown (using ingeniously forced elicitations of child utterance) that learning is characterized by great initial variation in productions for any given construction, apparently covering alternatives characteristic of many other languages, followed by abrupt transitions to stable adherence to the correct form for the target language.

Accordingly, a number of researchers have attempted to model language acquisition above the level of the lexicon as an *unsupervised* learning problem, in which the candidate grammars are enumerated, and searched in an orderly fashion for the one that covers the data. Since the number of candidate grammars is exponential in the number of parameters, and the assumption is that the child must update its choice of grammar incrementally, solely on the basis of a new sentence which is not covered by the existing grammar, without negative information about overgeneration, most of these approaches work by choosing a new grammar on the basis of similarity to the old, under a Subset Principle (Berwick, 1985), Fodor's 2009 statement of which can be paraphrased as follows:

(1) When there is a choice to be made between grammars that are all compatible with the available input sample, and the languages licensed by some are proper subsets of the languages licensed by the others, do not adopt any superset grammar.

Gibson and Wexler 1994 embody a kind of subset principle in a procedure which, when the currently hypothesised grammar fails on a new input, switches to a grammar that differs from the previous one on as few parameter values as possible. However, they note that this procedure can easily become trapped in false maxima (cf. Fodor and Sakas, 2005).

Fodor, Sakas, and Hoskey (2007) propose a version of the Subset Principle embodied in a lattice which compiles out the language inclusion relationships between grammars expressed as vectors of parameters. Thus the subset principle is adhered to if one parses with one of the least inclusive grammars. If such a least inclusive grammar fails to parse a new input sentence, then it is removed from the lattice.

The evolutionary approaches of Clark and Roberts (1993) and Yang (2002, 2004) avoid explicit invocation of the subset principle by learning over all possible parameter settings, incrementally eliminating or penalizing any grammar which fails to provide a parse for each successive string of words (whose parts of speech are known), and maintaining or rewarding grammars which succeed. Yang's model is based on Reinforcement Learning using classical Mathematical Learning Theory (Bush and Mosteller, 1955).

Thornton and Tesan (2007) argue that the changes they observe in childrens elicited productions are too abrupt and switch-like to support Yang's model. However, the model presented below shows that other kinds of probabilistic model are capable of approximating catastrophic, switch-like behavior.

It is clear from these papers that the unsupervised learning problem that they are attempting, in which the task is to discriminate between huge numbers of entire grammars on the basis of highly ambiguous single sentence data-points, is very hard indeed.

Perhaps the problem the child solves is a much easier one. Since many putative parameters (such as the pro-drop parameter) are specific to one category or construction of the grammar (such as the verb or clause), perhaps it might be better to learn at the level of the rules or lexical heads over which such parameters express generalizations than at the level of entire grammars. Perhaps also the childs learning problem is more like *supervised* grammar induction from a treebank, with language-independent logical forms taking the place of language-specific parse trees in corpora like the Penn treebank.[1]

This appendix, following work by Siskind (1996), Villavicencio (2002), Butt and King (2004); Butt (2006), Zettlemoyer and Collins (2005) and Bannard, Lieven, and Tomasello (2009), shows that a very simple statistical model and learning algorithm makes the notion of parameter-setting epiphenomenal. The only notion of trigger that it requires is the notion "reasonably short sentence in the target language, paired with a reasonably small number of reasonably simple contextually accessible candidate meanings, usually including the

1. The term "(un)supervised" is used here in the narrow machine-learning sense of learning from correctly labeled examples from whatever source. Of course, we do not mean the child's learning involves human instruction.

intended one". Unlike these earlier models, it does not require word-learning to precede syntactic-learning. The only locus of language-specific grammar is the lexicon for the language. The only locus of universal grammar is a universal mapping from semantic types to the possible corresponding lexical syntactic types of the kind described in chapter 2, together with a universal machine for merging or projecting lexical types and their meaning representations onto grammatical derivations. The former element is the locus of Chomsky's (1965) "substantive" universals concerning linguistic categories, while the latter is that of "formal" universals concerning syntactic projection.

B.1 Semantic Bootstrapping

The most plausible source for substantive universals is a universal *semantics*, broadly construed, in the form of structured meaning representations, closely related to the conceptual representations that enable the child's cognitive understanding of the world, to which the child already has access before language acquisition begins (Chomsky 1965:27-30; Chomsky 1995b:54-55), and to fragments of which syntactic types are rather directly attached, drastically limiting the search space (Hume (1738); Schlesinger (1971); Brown (1973); Bowerman (1973a); Clark (1973); Pinker (1979); Wexler and Culicover (1980); Hyams (1986); *passim*).

To say this much is not very helpful in psychological or linguistic terms, since (as Chomsky often points out) linguists don't know very much about the semantics. One of the problems that they face is that human semantics is greatly, perhaps even mainly, concerned with highly dynamic, mostly interpersonal and intensional content, of a kind that is deeply embedded in physical and interpersonal interaction with the world, and distinctly under-represented in formal linguistic theories of semantics (Tomasello, 1999).

However, the child doesn't *need* to articulate such a semantics. They just need to label chunks of it with linguistic categories, so our theories need to represent it somehow. As a temporary stopgap, we'll use terms of the lambda calculus, and defer the question of what a more psychologically realistic human semantics, of the kind called for by Tomasello, might actually look like till section B.9.

This approach makes the child's problem resemble that of treebank grammar induction for wide coverage parsing (Collins 1997; Charniak 2000; Hockenmaier and Steedman 2002b), where sentences hand-annotated with syntactic trees are used to derive a grammar and a statistical parsing model. However, the child's task is a little harder. First, they have to induce the grammar from strings paired with *unordered logical forms*, rather than language-specific ordered derivation trees. That is, they have to work out *which word(s) go with which element(s) of logical form*, as well as the directionality of the syntactic categories (which are otherwise universally determined by the semantic types of the latter). Second, while they do not seem to have to deal with a greater amount of error than is found in the Penn WSJ treebank (MacWhinney 2004), they may need to deal with *situations which support a number of logical forms*. Third, they need to be able to recover from temporary *wrong lexical assignments*. Fourth, they need to accomodate *lexical ambiguity*.

B.2 The Proposal

If we continue to take CCG as our theory of grammar, then the task that the child faces is simply to learn the language-specific categorial lexicon on the basis of exposure to (probably situationally ambiguous, possibly somewhat noisy) sentence-meaning pairs, given the universal combinatory projection principle of chapter 2, and the mapping from the same chapter from semantic types to the set of all universally available lexical syntactic types. To do the learning soundly and efficiently, the primary desideratum for such a system is that *probabilities learned in earlier stages of learning should influence the probabilities assigned to categories for unseen items encountered at a later stage*. Thus if every word or rule except one in a novel sentence has frequently been seen with the category required for its analysis, leaving only one possibility for the unseen word, that information should immediately be reflected in a high probability for that category for the unseen word.

B.2.1 The First Words

We begin with a running example, attested at first hand by a colleague.[2]

The child Emma, who had yet to learn her first word of such a grammar was taken for a walk by accompanying adults. She encountered a dog, in which she showed great interest. Later, she encountered some *more dogs*, and exhibited wild excitement, at which point the adult, observing the child's reaction, said "MORE DOGGIES!."

For the next couple of days, Emma used some approximation to the word "doggies" to mean *more*. The adult audience encouraged this progress, and offered no negative information whatsoever. Nevertheless, the child soon

2. Cathy Urwin, p.c., c.1979, University of Warwick.

switched to using "more" to mean *more*. and as far as we know never again used "doggies" with that meaning.

Emma's first utterances appear to reflect a overgenerating grammar fragment, in violation of the subset principle. yet she easily recovers. How is this possible?

We can assume that the child has already learned some phonological regularities of the language, and in particular is in a position to consider the possibility that the utterance consists of more than one word and the location of candidate word-boundaries (Saffran et al. 1996; Mattys, Jusczyk, Luce, and Morgan 1999; Mattys and Jusczyk 2001). Because we treat word boundaries probabilistically, the model we are developing will posutulate multi-word lexical entries, and in the absence of any evidence for the lexical independence of the components, will assign them a high probability. This actually happens with sequences like English *have+to*, often reduced as *hafta*, in both the model and real children, as the epigraph to the present chapter.

We will further assume that the first thing the child does is to take the string-category-logical form triple $\langle \textit{More doggies} := NP : \textit{more}\,\textit{doggies}\rangle$, and apply the rules of the grammar to it in the generative direction, to retrieve the final step of every derivation of that category possible under universal combinatory grammar. (This involves knowing the mapping between (known) semantic types and (unknown) legal CCG syntactic categories defined in chapter 12.5), This procedure, which Kwiatkowski (2012) shows is polynomial in complexity, is then recursively applied to all of the pairs of possible daughters, terminating when either the whole substring is treated as a lexical entry or the logical form is an atom (in which case, the latter step is forced). This results in a "shared forest" of derivations that efficiently represents the set of all derivations licensed by UG, resembling the "chart" of standard parsing algorithms like CKY, and forming a basis for calculating frequencies of events for the statistical parsing model, according to the algorithm given below.[3]

The only two combinatory rules that can have a non-function category as

3. This is a departure from the related approach of Zettlemoyer and Collins (2005), who assume that the child considers a larger set of lexical candidates onstructed by taking the cross-product of every non-empty substring ϕ of the utterance "More doggies!" with every connected typed subterm λ_τ of type τ the logical form *more doggies*, together with all syntactic categories σ_i that universal grammar allows for the semantic type τ of each such subterm. For the example to hand, this set would include certain potential categories, such as more:=$NP\backslash N$: *more* for which there is in fact no evidence from the sentence "more doggies!", and which the present algorithm will not generate. These spurious candidates are subsequently excluded by running a CCG parser over all possible sequences of lexical categories compatible with the string and detecting the fact that there is no derivation that they take part in.

their result are the rules of function application, here presented with phonological concatenation explixit:

(2) *Forward Application:*
$$\langle \phi_l, X/Y, F\rangle \quad \langle \phi_r, Y, A\rangle \Rightarrow \langle \phi_l \phi_r, X, FA\rangle$$
Backward Application:
$$\langle \phi_l, Y, A\rangle \quad \langle \phi_r, X/Y, F\rangle \Rightarrow \langle \phi_l \phi_r, X, FA\rangle$$

Since we know the value of all elements of the result, then if we know the universal syntactic types that correspond to the semantic types of F and A, we know all possible values of all elements of the left hand side. The utterance to hand, "More doggies", generates just three derivations, as follows:[4]

(3) a.
$$\dfrac{\dfrac{\textsc{More}}{\text{more} := NP/N : more_{((e,t),e)}} \quad \dfrac{\textsc{doggies}}{\text{doggies} := N : dogs_{(e,t)}}\ !}{\text{more doggies} := NP : more\,dogs_e}{>}$$

b.
$$\dfrac{\dfrac{\textsc{More}}{\text{more} := N : dogs_{(e,t)}} \quad \dfrac{\textsc{doggies}}{\text{doggies} := NP\backslash N : more_{((e,t),e)}}\ !}{\text{more doggies} := NP : more\,dogs_e}{<}$$

c.
$$\dfrac{\textsc{More doggies}\ !}{\text{more doggies} := NP : more\,dogs_e}$$

The following set of candidate lexical entries can be read off the three derivations in (3):

(4) The child's lexical candidates:

more:=	$\mathbf{NP/N : more_{((e,t),e)}}$
	$N : dogs_{(e,t)}$
doggies:=	$NP\backslash N : more_{((e,t),e)}$
	$\mathbf{N : dogs_{(e,t)}}$
more doggies:=	$NP : (more\,dogs')_e$

Since this is the only data-point that the child has processed so far, the correct hypotheses that "more" is a determiner meaning *more* and "doggies" is a noun meaning *dogs* are assigned the same probability mass as the incorrect hypotheses where the categories are reversed.[5]

4. The example is simplified for exposition. The assumption that the child immediately considers the hypothesis that more is a determiner is particularly far-fetched, and will be reviewed later.
5. The hypothesis that "more doggies" is a multiword item meaning *more dogs* attracts less proba-

Given such a lexicon, the child is at this stage as likely to choose the word "dog" to express the meaning *more* as to choose the word "more", as observed by Cathy Urwin.

The set (3) of derivations also defines a partial generative parsing model that can be lexicalized with head-word dependency information, of the kind discussed in the preceding appendix, whose details we pass over here.

Such a model, including the lexicon, can be learned using an incremental version of the expectation-maximization (EM) algorithm (Neal and Hinton, 1999). For each sentence meaning pair EM algorithm operates in two steps. The first step is *estimation* on the basis of the existing model (the prior) of the probabilities of all possible analyses of the sentence-meaning pair(s) to hand, such as the set (3) for the single pair ("more doggies",*more dogs*). The second step is *maximization* of the model by updating the probabilities of rule instances and lexical entries (adding new ones as needed) in proportion to their observed frequency in the set of derivations to hand weighted by the prior probability of the particular derivation that they occur in.

The idea behind this algorithm is that observing an event repeatedly in derivations with high overall prior probability should lead us to assign more probability mass to that event in the model.

For example, Let us suppose that the second utterance the child hears is "More cookies", paired with the meaning *more cookies*. There are again three derivations and five lexical entries parallel to (3). However, the correct derivation—that is, the one parallel to (3a)—will now be assigned a higher prior probability than the incorrect derivation parallel to (3b), because the lexical pairing of "more" with the category NP/N : *more* has been seen already in a possible derivation for the sentence "more doggies", whereas the pairings of "more" with N : *cookies* and "cookies" with $NP\backslash N$ have not. Because the probability of that derivation is higher, the update in step 2 of the algorithm to the probability of the lexical entry for "more" meaning *more* will be greater than the probability assigned to the new spurious lexical entry for "cookies" with that meaning, while the probability that it means *cookies* is greater in proportion.

At this stage, the model is already more likely to generate "more" from the meaning *more* than either "doggies" or "cookies".

For the same reason, although the instances $NP/N \quad N \Rightarrow NP$ and $N \quad NP\backslash N \Rightarrow NP$ of the application rules (2) have both been seen before,

bility mass because the prior probabilities over wordforms and logical forms are set to favor shorter and simpler entries.

the probability associated with the derivation including the former is higher than that of the one including the latter, so the probability associated with the former is updated by more than the latter.

The added probability mass associated with events of type $NP/N \quad N \Rightarrow NP$ means that for a third sentence of the form "more X", the model will assign an even higher probability to the derivation parallel to (3a), in a "rich get richer" process that will soon make the learner able to identify new nouns in one trial.

It should be noted that, due to the incremental nature of the algorithm, the probability associated with the spurious lexical entry associating the word "more" with the meaning *dogs* is unchanged. In that sense the incremental model only approximates a sound batch model. It is only further encounters with sentences including the word "doggies" that will weaken this spurious lexical entry in relation to the correct one.

This model generalizes immediately to the more realistic situation where the child has to deal with more than one candidate for the intended meaning of the utterance. While the irrelevant distractors will generate spurious lexical entries, as learning proceeds, the associated derivatrions from the words in the sentence will be low likelihood, and very little further evedence in their support will be forthcoming in future. Their probability will diminish, and they can if necessary be pruned on that basis

B.3 Syntactic Bootstrapping

Many languages, perhaps all, allow a number of lexical alternations of transitives, as in the case of English "chase/flee" where the same physical situation seems to support more than one logical form. Fisher, Hall, Rakowitz, and Gleitman (1994) raise the question of how do children faced with (artificial) examples like the following avoid the error of making an OVS lexical entry for "flee" with the meaning *chase*?

(5) Kitties flee doggies!

It is important that syntactically unmarked alternations of this kind comparitively rare among verbs, so that we are justified in assuming that the child will have encountered plenty of non-alternating transitive concepts like *seeing* see before encountering examples like (5).

This means that the learned prior probability of the instantiated rules for combining transitive SVO verbs with their object and subject will be exponentially greater than the priors on the rules for OVS verbs by the time they eventually do encounter (5), as well as the priors that "kitties" means *cats* and "doggies" means *dogs*

It follows that the correct SVO derivation will be assigned a higher prior probability by the model than the spurious OVS derivation. Thus, the child is in a position to assign high probability to the correct lexical entry for "flee".

Gleitman 1990 and Gleitman, Cassidy, Nappa, Papafragou, and Trueswell (2005) have described such learning as "syntactic bootstrapping," implying that it is the child's knowledge of the language-specific SVO syntax, as opposed to semantics, that guides lexical acquisition. However, in present terms such a grammatical influence on learning is simply emergent from the statistical model used in semantic bootstrapping in the later stages of learning. We will return to this point.

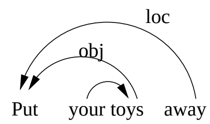
Zettlemoyer and Collins 2005 did not include a parser model as distinct from the lexical model. They get away with this because their grammar is small and English specific, defined by the artificial GeoQuery dataset (Thompson and Mooney, 1998). For small unambiguous grammars, CCG categories alone are usually enough to eliminate search. However, we know from experience with parsers of the size of those needed for the Wall Street Journal corpus that such a model will be necessary once the child's grammar begins to approach adult—or even CHILDES—size.

It is not clear whether the child uses a head-word-dependency model of the kind used by Collins (1999) and discussed in appendix C as a parsing oracle, or uses semantics and world knowledge directly, as proposed by Crain and Steedman 1985; Altmann and Steedman 1988 and others, or some interpolation of the two, such as the semantics sketched in the previous appendix.

B.3.1 A Computational Model of Child Language Acquisition

Learners of the above form has been implemented as a computer program by Kwiatkowski et al. (2012) and Abend et al. (2017), using the Variational Bayes form of the EM algorithm (Sato, 2001; Beal, 2003; Hoffman, Blei, and Bach, 2010) to learn a mean field approximation to the complete joint distribution of derivations and model parameters, given a sentence-meaning pair. The programs have been trained and tested on the "Eve" section of the CHILDES corpus (MacWhinney, 2000).

B.3.1.1 The data Sagae, MacWhinney, and Lavie (2004) have tagged the Eve section of the CHILDES corpus with dependency graphs. For example:

(6) 

The dependency graphs are lacking in certain information the learner needs, so they are mapped onto something more like a predicate-argument structure or first-order formula, such as the following:[6]

(7) $put\text{-}away\,(of\,eve\,toys)\,eve$

Crucially, we ignore any alignment of the terminals in the depedency graph and the corresponding formula with the words in the sentence, treating the sentence as a foreign language with no known relation to the elements of the logical form.

Nevertheless, it is important to note that the English-centered form of the CHILDES data will cause the learner to learn constructions that are *more analytic* in other languages than English as multi-word items. For example, consider French *Milou traverse la rue à la course!* ("Milou runs across the road!")

(8)
$$\begin{array}{ccc}
\text{Milou} & \text{traverselarueàla}\textsc{course} & ! \\
\overline{S/(S\backslash NP) : \lambda p.p\ milou} & \overline{S\backslash NP : \lambda y.run\,(across\,road)y} & \\
\hline
\multicolumn{2}{c}{S : run\,(across\,road)\ milou} & >
\end{array}$$

Thus, we are using the dependency notation as a proxy for the language-independent conceptual "Language of Mind" that the real child has access to.

Nevertheless, we can learn any construction in any language that is *less* analytic than English, such as French *Range tes jouets!* ("Put away your toys!"):

(9)
$$\begin{array}{ccc}
\textsc{Range} & \textsc{tes jouets} & ! \\
\overline{S/NP : \lambda x.put\,away\,x\ you} & \overline{NP : toys} & \\
\hline
\multicolumn{2}{c}{S : put\,away\,toys\,you} & >
\end{array}$$

However, we would clearly prefer a better, less language-dependent semantic representation for content-words.

Given the mapping from chapter 2 defining the set of syntactic categories licensed by the Universal Grammar, the acquisition model outlined above can be run on the corpus to provide an indication of our model's performance on real child directed speech for comparison with the real child's performance.

The Eve section of the corpus (Brown, 1973) consists of around twelve thou-

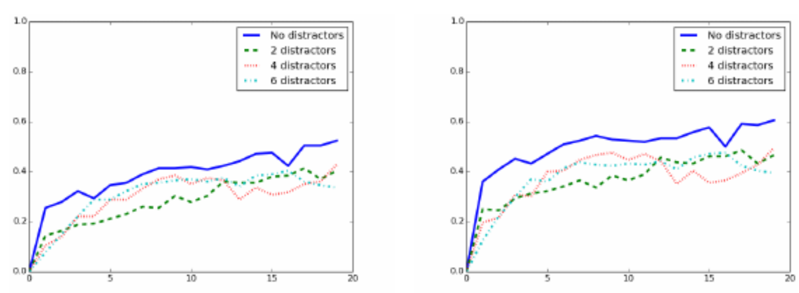
6. Such logical forms are of course even more ridiculously simplified than those we have seen in the earlier chapters.

sand sentences. Of these, we exclude around 60% which consist of semantically unhelpful utterances like "MMMMMM!" and "DOGGIE DOGGIE DOGGIE" together with some very long sentences. Abend et al. hold out the chronologically last section 20 of the corpus as unseen testmaterials. They then train the learner incrementally on sections 1-19, testing on section 20 after each, and thereby obtaining learning curves over the whole dataset.

It should be pointed out that this is a very tough testing regime. Our learner is only given what we estimate to be around 2% of the data that Eve herself had been exposed to by the time of session 20. It will become apparent that this was only just enough data to obtain reasonable stable learning.

B.4 The experiment (Abend et al., 2017)

The figures in this section show learning curves for various constructions and their head words. The curves are of two kinds. Some are for learning general types, such as that the English transitive verb bears the SVO type $(S\backslash NP)/NP$ rather than any of the alternatives corresponding to SOV, VSO, OSV, VOS, OVS. Other curves compare frequent versus infrequent members of such general classes, showing that late learning is qualitatively different, showing the step-like learning characteristic of structural bootstrapping, in contrast to the more continuous early learning, from which syntactic bootstrapping is emergent. Still other curves compare rates of learning across categories such as nouns and verbs. In all cases, each curves is shown with increasing numbers of irrelevant "distractor" meanings involved. In all cases learning is slower with distractors present, but still converges.



(a) (b)

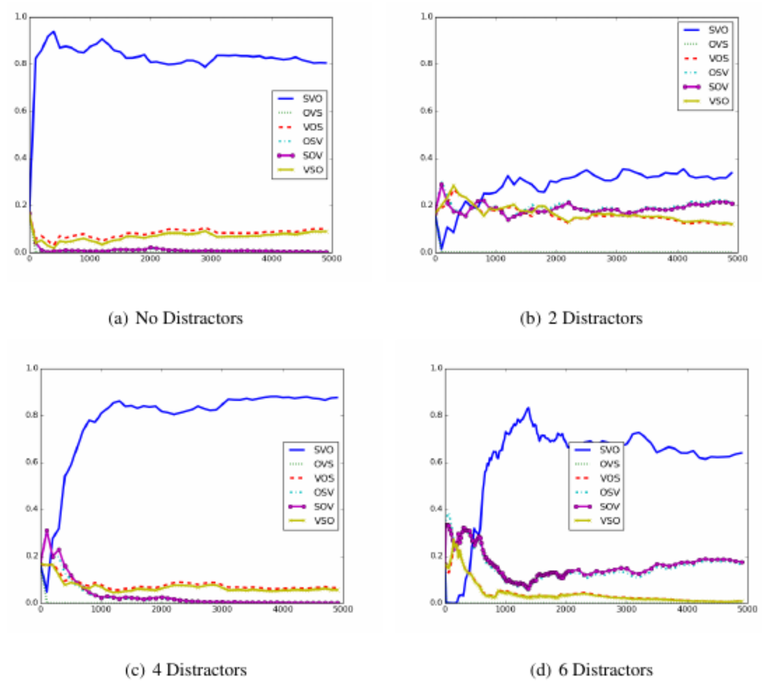
Figure B.1: Overall Learning Curves (from Abend et al. (2017))

Figure B.1 shows overall learning of correct categories for all lexical item, with separate curves for increasing numbers of distractor meaning training on sentence-meaning pairs from the first *i* sessions, testing on accuracy in assigning meanings to unseen session 20 sentences, (a) without, and (b) with, "guessing" syntax and semantics of unseen words in the test-set on the basis of the seen words.

The improvement of the latter is a measure of the extent and earliness of emergent syntactic bootstrapping.

Overall learning level is low, because learning is from only 2% of the real child "Eve"'s actual exposure to these data.

This is barely enough data for robust learning: the model seems to get unlucky with the 2-distractor condition.



(a) No Distractors (b) 2 Distractors

(c) 4 Distractors (d) 6 Distractors

Figure B.2: Semantically Bootstrapping the SVO Category-type for transitives (from Abend et al. (2017))

Figure B.2 shows that the learner rapidly learns which words are the verbs, and that they bear the SVO category, even with six distractor meanings. Again, something about the two distractor condition makes learning slower, but SVO is still ahead.

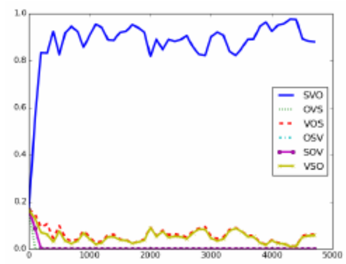


Figure B.3: Learning SVO with 2 Distractors: Slower Learning Rate (from Abend et al. (2017))

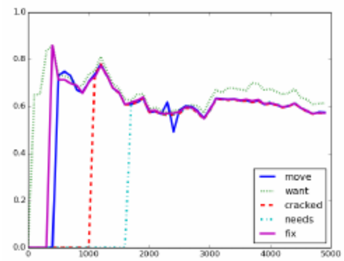
Figure B.3 shows that changing to a slower learning rate, allows successful learning in the two-distractor condition. However, it is likely that a similar instability with then show up in some other condition. Learning rate is not a critical parameter of the model: the problem is the shortage of training data: we are operating with barely enough to ensure stable learning.

Figure B.4 shows individual learning curves for decreasingly frequent verbs "want" > "move" > "fix" > "crack" > "needs". The measure here is Correct Production Probability (CPP). (The unstable 2-distractor condition is omitted as distracting.)

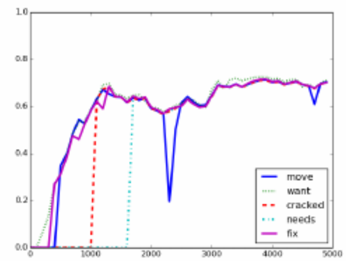
The rarer verbs are learned later, and so the chance is increased that so much grammar has been learned by the time that they are encountered that everything *else* about the sentences has been learned with sufficient probability mass that the novel lexical item reaches the average CPP of the grammar so far *in one step*, giving the appearance of syntactic bootstrapping. However, syntactic bootstrapping of this simple kind is simply an emergent property of semantic bootstrapping.

Figure B.5 demonstrates the quantitative effect of this kind of syntactic bootstrapping as the area between the curves of average CPP for 99 frequent and 23 infrequent verbs (frequency threshold of 10 for the latter.)

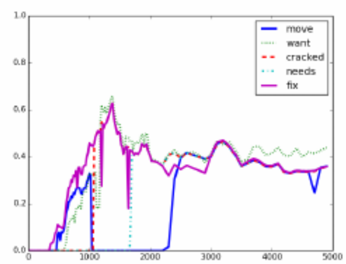
Kwiatkowski et al. (2012) and Abend et al. (2017) showshow similar semantic and syntactic bootstrapping effects, including one-trial late learning, for other grammatical categories including nouns, determiners and prepositions, and a number of similarities and differences in the rate and onset of learning of nouns and verbs



(a) No Distractors



(b) 4 Distractors

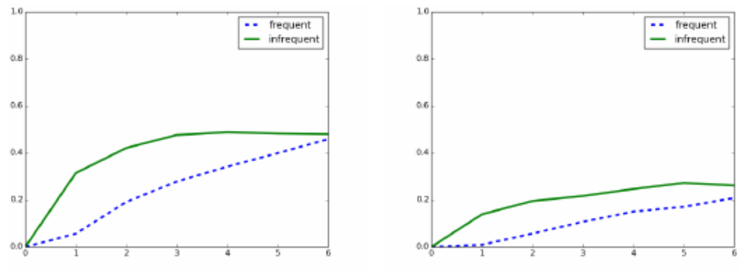


(c) 6 Distractors

Figure B.4: Syntactic Bootstrapping of Less Frequent Verbs (from Abend et al. (2017))

B.5 Against Parameters

Like the related proposals of Siskind, Villavicencio, Zettlemoyer and Collins, and the somewhat different probabilistic approach of Yang 2002, this proposal considerably simplifies the logical problem of language acquisition. In particular, it allows us to eliminate the Subset Principle of Berwick (1985), and attendant requirements for ordered presentation of unambiguous parametric triggers, both of which appear to present serious problems for the language learner (Angluin 1980; Becker 2005; Fodor and Sakas 2005). Nor does this move contradict widely-held assumptions concerning the "poverty of the stimulus", and in particular the assumption that no negative evidence is available to the child. The child's progression from the universal superset grammar to the language-specific target grammar is entirely determined by positive evidence, which raises the probability of repeatedly supported hypotheses at the expense



(a) 4 Distractors

(b) 6 Distractors

Figure B.5: Quantitative Evaluation of Syntactic Bootstrapping Effect (from Abend et al. (2017))

of unsupported ones. The incorrect hypotheses that are eliminated in this way include any that are introduced by error and noise. The only evidence that the child needs in order to learn their language is a reasonable proportion of utterances relavant to the context, involving sentences which are sufficiently short for them to deal with.

The theory presented here resembles the proposal of Fodor, Bever, and Garrett 1974 and Fodor 1998 as developed in Sakas and Fodor (2001) and Niyogi (2006) in that it treats the acquisition of grammar as arising from parsing with a universal "supergrammar". As in that proposal, both parameters and triggers are simply properties of the language-specific grammar itself—in their case, rules over independently learned parts of speech, in present terms, a parsing model of the kind discussed in appendix C for English, but applied over the entire space of possibilities allowed by universal grammar which are consistent with the data.

It differs in assuming that the unordered logical form for the utterance is mostly available, with tolerable degrees of error and ambiguity. This means that the problem of syntactically ambiguous sentences to which Structural Triggers Learning (STL, Fodor et al., 2007) is heir does not arise.

It also differs in the algorithm by which it converges on the target grammar. Rather than learning rules in an all or none fashion on the basis of unambiguous sentences that admit of only one analysis, it adjusts probabilities in a model of all elements of the grammar for which there is positive evidence for *all* processable utterances. In this respect, it more closely resembles the proposal of Yang (2002). However, it differs from the latter in the rate of learning (cf. 2002:fg.2.1), and differs from both in eschewing the view that grammar learning is parameter setting.

If the parameters are implicit in the rules or categories themselves, and you can learn the rules or categories directly, why should the child or the theory bother with parameters at all? For the child, all-or-none parameter-setting is counterproductive, as it will make it hard to learn the many languages which have inconsistent settings of parameters across lexical types and exceptional lexical items, as in German and Dutch verb-finality.

The fact that languages show violable tendencies to consistency for values of parameters like headedness across categories for related semantic types such as verbs and prepositions probably stems from considerations of overall encoding efficiency for the grammar as a whole, of the kind captured in notions like Minimal Description Length (MDL) or more elegantly in the Tolerance Principle of Yang (2016). Such considerations may be relevant to comparing entire grammars for the purpose of explaining language change, as in the work of Briscoe (2000). Their presence will under the present theory make the task of learning easier, by raising prior probabilities in the model for rules and categories that actually do recur. But it is less clear that representing them explicitly, rather than leaving them implicit in th model, will help the individual child learning a specific grammar, word-by-word.

B.6 A More Realistic Lexicon

If children's exposure to language were merely confined to recitations of propositions they already had in mind, it would be a dull affair. It is not even clear why they would bother to learn language at all, as Clark (2004) points out in defence of a PAC learning model.[7]

However, the worked example above is deliberately simplified in respect of the child's syntax and semantics. We know from Fernald, Taeschner, Dunn, Papousek, Boysson-Bardies, and Fukui (1989) and Fernald (1993) that infants are sensitive to interpersonal meanings of intonation from a very early age. In English, intonation contour is used to convey a complex system of information-structural elements, including topic/comment markers and given/newness markers (Bolinger 1965; Halliday 1967a; Ladd 1996), and is exuberantly used in speech by and to infants. It is this part of the meaning that constitutes the whole point of the exercise for the child, providing the motiva-

7. It is equally unclear why they would bother to learn language in the *absence* of any interpretation, as assumed in PAC learning, a point which Clark doesn't address.

tion that Clark questions.

For example, it is likely that the child's representation of the utterance "MORE DOGGIES! is more like (10), which uses the notation of chapter 6, in which [S] represents speaker modality (contributed by the LL% boundary tone), ρ represents rheme, and $+$ represnts polarity (both contributed by the H* accents).

(10)
$$\begin{array}{ccc}
\textrm{MORE} & \textrm{DOGGIES} & !\\
\textrm{H}* & \textrm{H}* & \textrm{LL\%}\\
\hline
NP^{\uparrow}_{\rho}/N_{\rho} & N_{\rho} & X_{\phi}\backslash_{\star}X_{\eta}\\
:\lambda p\lambda x.p(*more\,x) & :*dogs & :\lambda g.\eta\, g\,\textrm{S}\\
\hline
\multicolumn{2}{c}{NP^{\uparrow}_{\rho}} & >\\
\multicolumn{2}{c}{:\lambda p.p(*more*dogs)} & \\
\hline
\multicolumn{3}{c}{NP^{\uparrow}_{\phi}:\rho\,\lambda p.p(*more*dogs)\,\textrm{S}} <
\end{array}$$

—which the child interprets as meaning something like "Mum makes the property afforded by more dogs common ground."

The semantics of speech acts goes beyond the immediate concerns of this paper, and is discussed by Steedman and Petrick (2007), who note that the inference system that the semantics supports is closely related to that involved in planning with sensing actions.

The set of type-raised NP categories licenced by UG that is schematized in (10) as $NP^{\uparrow}$ denotes the set of all order-preserving functions over functions-over-NP onto the results of applying those functions to the original NP. It includes categories of the following two forms, where T is a variable over all category types::

(11) $\mathrm{T}/(\mathrm{T}\backslash NP):\lambda p\lambda x.px$
$\mathrm{T}\backslash(\mathrm{T}/NP):\lambda p\lambda x.px$

We also need the following related non-order-preserving "extracting" categories, in which S_x indicates a distinct type of clause:

(12) $S_x\backslash(S\backslash NP):\lambda p\lambda x.px$
$S_x/(S/NP):\lambda p\lambda x.px$

While, up until now, we have only seen one syntactic type per semantic type in the child's lexicon for English, in general a single semantic type may be realized by many syntactic types in a single language, and this is the case for English NPs. Such ambiguity is perfectly compatible with the learning procedure defined earlier: it just means that there will be several categories with substantial conditional probability mass $P(\sigma|\phi)$

It may seem surprising that a language should allow so much ambiguity in such a basic linguistic category type as NP. However, this is simply the same proliferation of syntactic types that would be disambiguated in a language with overt morphological *case*. English just ahppens to be a language which has so-called structural case, implicit in linear order. We shall see that the child will not find this a problem. But first we need to consider the role of intonation in the child's grammar.

While intonation has been shown to be even more markedly discrepant from traditional syntactic structure in child-directed and child-originated speech (Fisher and Tokura 1996; Gerken, Jusczyk, and Mandel 1994; Gerken 1996; Zubizarreta and Nava 2011) than in adult dialog, in CCG intonation structure is, as we saw in chapter 6, united with a freer notion of derivational structure. Consider the child in a similar situation faced with the following utterance, from Fisher and Tokura (1996), repeated from chapter 6:

(13)

$$
\begin{array}{cccc}
\text{You} & \text{LIKE} & & \text{the doggies!} \\
 & \text{H}* & \text{L} & \text{LL\%} \\
\hline
S/(S\backslash NP) & (S\backslash NP)/NP & X_\phi\backslash_\star X_\eta & S_\phi\backslash(S_\phi/NP_\phi) \\
:\lambda p.p\;you & *like & :\lambda g.\eta\,(g\,\mathrm{S}) & :\theta(\lambda q.q\,dogs)\,\mathrm{S} \\
\multicolumn{2}{c}{\overline{S/NP:\lambda x.*like\,x\,you}\;{>}\mathbf{B}} & & \\
\multicolumn{3}{c}{\overline{S_\phi/NP_\phi:\rho(\lambda x.*like\,x\,you)\,\mathrm{S}}\;{<}} & \\
\multicolumn{4}{c}{\overline{S_\phi:\theta(\lambda p.p\;dogs)\mathrm{S}(\rho(\lambda x.*like\,x\,you)\mathrm{S})}\;{<}} \\
\multicolumn{4}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\
\multicolumn{4}{c}{S: like\,dogs\,you}
\end{array}
$$

"Mum supposes the properties the dogs afford to be common ground, Mum makes it common ground it's what I like."

Fisher points out that the L intermediate phrase boundary that she observed after the verb makes the intonation structure inconsistent with standard assumptions about surface constituency. However, this intonation structure is isomorphic to the CCG derivation above, which delivers the corresponding theme/rheme information partition directly.

Thus, here too, the availability of the full semantic interpretation, including information-structural information, directly reveals the target grammar. In this case, since the derivation requires the use of the forward composition rule, indexed $>\mathbf{B}$, the child gets information not only about the probability of the verb, the nominative, and the accusative categories of English, but also about the probability of applying the composition rule to the first two categories, the probability that the subject of "like" will be headed by "you", and its object be headed by “doggies”. Thus, the child can build the parsing model in parallel with learning the grammar.

The involvement of the rules of composition requires the inclusion of slash-types in the learning (the application rules apply to all types.) We assume that categories are learned conservatively, with types only as general as the derivations they are seen with require. That is, if X/Y is only seen in purely applicative derivations, it will be learned as $X/_{\star}X$. However, derivation (13) is impossible without the forward harmonic composition rule. We therefore generalize the type for the nominative learned so far to be $S/_{\diamond\star}(S/NP)$, compatible with that rule. Similarly, the involvement of crossing composition in a derivation such as Heavy NP Shift will induce generalisation to $\times$ or $\cdot$ slash-type on the relevant argument.

B.7 Smoothing and Generalization

A standard assumption in wide-coverage parsing using treebank grammars is that the grammar must be generalized and the statistical model must be smoothed with respect to unseen words and word-category pairs. Since all language-specific information in CCG resides in the lexicon, this amounts to predicting unseen word-category pairs and head-word-dependencies.

Generalizing grammars is a tricky business: Fodor and Sakas offer as an example the observation that the child should assume on the basis of seen topicalizations in English that all NPs can undergo topicalization. However, they should not assume on the basis of observations of negative placement with repect to auxiliaries that the same process can apply to all verbs.

This problem looks rather different from the present perspective. Since we are learning a probabilistic instance of universal grammar, the grammar is already generalized, and predicts all possible word-category pairs. Since topicalization is a lexically-specified contruction in CCG, when the child hears the following as its first example of the construction, it still has available all possible categories for “doggies”, including the preposing topicalized one that supports this derivation:

(14) $$\begin{array}{ll}
\text{D{\scriptsize OGGIES}} \quad , & \text{you} \ \text{{\scriptsize LIKE}} \quad ! \\
\text{L+H}* \quad \text{LH\%} & \text{H}* \quad \text{LL\%} \\
\overline{S_\phi/(S_\phi/NP_\phi)}^{<} & \overline{S_\phi/NP_\phi}^{<} \\
:\theta\,(\lambda p.p*dogs)\,\text{H} & :\rho\,(\lambda x.*like\,x\,you)\,\text{S} \\
\multicolumn{2}{l}{\overline{S_\phi : \theta\,(\lambda p.p*dogs)\,\text{H}\,(\rho\,(\lambda x.*like\,x\,you)\,\text{S})}^{>}} \\
\multicolumn{2}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\
\multicolumn{2}{c}{S : like\,dogs\,you}
\end{array}$$

"I suppose what property dogs (as opposed to something else) afford to be common ground, Mum makes it common ground it's me liking them (as opposed to anything else)."

So the conditional probability of this category given this type $P(S_\phi/(S_\phi/NP_\phi) \mid ((e,t),t))$ will grow and become available to other words of that type, supporting generalization.

We must correspondingly assume that the non-generalization of the negative category is based on a semantically distinct type of verb, the raising modal auxiliaries.

B.8 Limiting Overgeneralization

Children do of course overgeneralize in their productions (Braine, 1971; Bowerman, 1988). Ambridge, Pine, Rowland, Chang, and Bidgood (2013) identify three types of overgeneralization found in child utterance:

1. Sense overextension: "doggy" denotes all animals, "daddy" all men.
2. Morphological overgeneralization and overanalysis: "spyer" for "spy"; "hoove" as a verb from "hoover"; "runned" as the past tense of "run".
3. Overgeneralization of lexical rules: "(Don't) giggle me" for "make me giggle"; "(Shall I) whisper you something" for "whisper something to you".

To understand how children escape from such overgeneralizations, we need only assume that they are paying atttention to the utterances of adult speakers and building a probabilistic model in the manner just described. Adult utterance will provide very little evidence that "doggies" extends to other animals, so to reverse the first kind of overgeneralization, all the child has to do is to learn the features that distinguish dogs on further occasions of child-directed use, and to include those in the lexical semantic sense of the word.

Similarly, even if the child includes the words "spyer", "hoove", and transitive "giggle" in their lexicon, they will not attract any probability mass for those meanings on the evidence of adult speech, in contrast to “spy”, “hoover”, and periphrastic “make giggle”.[8]

By contrast, we have children acquiring the English relative clause should never overgeneralize relativization and other *wh*-constructions to allow embedded subject extraction. While the existence of object extraction will force them to allow $\diamond$ type for the slash on the S/S' complement of verbs like “think”, they will never see any evidence for any further generalization to $\times$ or $\cdot$ modality, which would be necessary to allow **that-trace* violations. Thus they should never show such violations, not indeed any true embedded subject extractions until they learn the subject extracting category (39) for the bare complement verbs from chapter 9. (They may on the other hand repair such extraction with resumptive elements of various kinds, as adult speakers do in nonstandard relatives like “#a book which I don’t know where *it* is”. The elicitation studies of Thornton (1990); Thornton and Crain (1994) and Crain and Thornton (1998) showed no **that-trace* violations in children’s production of embedded subject *wh*-questions, although they some children did produce adult embedded subject extractions with bare-complement verbs, while others produced non-adult relative forms with intermediate wh-elements, such as the following:

(15) a. Who do you think who is in the box?
b. Who do you think who Cookie Monster likes?

(16) a. What do you think who is ate this?
b. What do you think where this froggy lives

As Thornton and Crain point out, these forms are reminiscent of “*wh*-copying” relatives in German, and French embedded subject extractions. They interpret these non-standard relativisers as residues of successive cyclic *wh*-movement. However, as we saw in section 9.6, these constructions and some related phenomena in Irish can be captured without movement, within the same degrees of freedom as standard embedded subject-extraction.[9]

8. There is some anecdotal evidence that children distinguish their own nonce productions from the adult language, and that they reject similar productions if adults tease them by uttering them deliberately (M.Bowerman, p.c.).
9. Stromswold, 1995:41 reports a possible **that-trace* violation in a spontaneous child utterance *What do they say that is in Holland*. However, as she points out, other explanations of this datum are possible, including Thornton and Crain’s and the present one.

B.9 Grounding Semantics in Use and Interaction

One might ask at this point how the child or machine comes to have access to the logical form *more dogs* (or whatever), and why she does not entertain other candidates, such as *more tails*. As Quine (1960) pointed out, this is a different kind of question, whose answer lies in the nature of the child's sensory-motor interactions with the world, and depends on hundreds of millions of years of vertebrate evolution evolution, rather than on learning in the individual child.

Nevertheless, this observation carries a warning that the semantics that emerges from that interaction and those evolutionary processes may be very unlike the semantics that naive logicist assumptions suggest, and that is used in the corpus for Abend et al.'s experiment. For example, the logical form that the child brings to (13) is likely be something more like *give pleasure you dogs*, so that the lexical entry for "like" of type $(e,(e,t))$ is something like (17), exhibiting the same "quirky" relation between (structural) nominative case and an underlying dative role that Icelandic exhibits morphologically for the corresponding verb:

(17) like := $(S\backslash NP)/NP : \lambda x \lambda y.give\,pleasure\,y\,x$

Similarly, it is quite possible that the childs initial representation of the meaning of "more" is as a predicate $S/NP : more$, and that it is the resulting prior on the conditional probability $P(S/NP|e \to t)$ that is generalized to "allgone", leading to transient non-standard orders like "Allgone milk". Or "all gone" may be misanalysed as a proto-determiner like "no more." These questions are much harder to investigate. While one can annotate corpora such as CHILDES with logical forms, as Villavicencio and Abend et al. did, one has very little idea of what relation such logical forms bear to a psychologically real adult semantics, let alone a child's. This fact makes quantitative testing of the present theory difficult.

One we way around this is to do as the linguists do, and meditate on the huge collection of phenomena to do with binding, case, classification, tense, evidence, and aspect, and so on, that seem to dimly reveal an underlying system of meanings, in the hope of discerning the real semantics. This is a very hard problem, and progress seems slow.

Another alternative is to investigate the question qualitatively, using simulated language learners. Since large corpora of artificial logical forms such as database queries annotated with sentences are unlikely to become available, and everyone believes that the semantics is determined by the child's sensory-motor experience of acting in the physical world, this makes the use of physically grounded robots particularly interesting. Projects of this kind are under investigation by a number of groups, including those led by Luc Steels, Deb Roy, and Geert-Jan Kruijff. These groups are looking at emergence of agreed vocabulary among prelinguistic agents (Steels and Baillie 2003; Steels 2004), plans and plan-recognition as a basis for situated language understanding (Roy 2005; Gorniak and Roy 2007), and context-dependent spatial models for natural language semantics (Kelleher, Kruijff, and Costello 2006). However, these projects so far rely on forms of semantics that are designed top-down, using the robot tasks as a forcing function, rather than on a semantics developed bottom-up from action representations themselves. Delivering semantic representations that are grounded in the same sense that mechanisms developed over hundreds of millions of years of evolution is much harder. Appendix A argues that the combinators **B** and **T** that do most of the projective syntactic work in CCG are directly related to operations of seriation and affordance in the planner. This suggests that mechanisms for state-based deliberative planning of the kind investigated by Petrick and Bacchus (2002, 2004) may offer a way towards a more distinctively action-based semantics for natural language (cf. Steedman 2009, Geib and Steedman 2007).

It is also likely that the child's acquisition of lexical semantics is piecemeal, and based on multiple examples of use, as proposed in the previous appendix. That is, it seems quite likely that the child first learns that *conquer* means some kind of *attack*, and only later acquires the specific components that distinguish the former from the latter. The denotation-based method for acquiring complex semantics discussed there would enable the child to learn that *liking things* entails *things pleasing* and *being pleased by things*.

B.10 Conclusion

This appendix has argued that syntax is learned on the basis of universally available preexisting prelinguistic conceptual structures, afforded by the situation of adult utterance, using a generative statistical model over a universal set of grammatical possibilities. The existence of the model itself helps the child to rapidly acquire a correct grammar even in the face of competing ambiguous candidate interpretations and consequent error.

In equating language-specific grammar with a statistical model for parsing with universal grammar, the proposal bears an intriguing relation to the Maximum Spanning Tree (MST) parser (McDonald, Pereira, Ribarov, and Hajic

2005; McDonald and Pereira 2006b,a). This parser searches for the maximum-valued spanning tree-forming subgraph of a totally connected graph over the words of the string, using a perceptron-like maximum-margin discriminative model trained using pairs of strings and dependency trees. It has been applied to parsing "non-projective" or long-range dependencies, including crossing dependencies. It works best when the features over which the model is trained are grammar-like features such as position with respect to the verb, or morphological features. In particular, Çakıcı (2009) has shown that using CCG categories as features in a dependency-model of Turkish improves performance over the baseline in McDonald and Pereira (2006b). MST could therefore be seen as offering an alternative, discriminative, version of the present approach, according to which it could be used to learn weights for a language-specific set of features or categories drawn from a larger universal set.[10]

The fact that the onset of syntactically productive language at the end of the Piagetian sensory-motor develomental phase is accompanied by an explosion of advances in qualitatively different "operational" cognitive abilities suggests that the availability of language has a feedback effect, facilitating access to concepts that the child would not otherwise have access. Early work by Oléron (1953) and Furth (1961) on specific cognitive deficits concerning non-perceptually evident concepts arising in deaf children who had been linguistically deprived by being denied access to sign supports this view.

This means that Gleitman's (1990) influential suggestion that it is the availability of syntax that enables the child to "syntactically bootstrap" lexical entries for verbs (such as "think") that are not situationally evident is essentially correct. However, it is the availability to the child of *a model of the relation between language-specific syntax and universal semantics* that makes this possible. It follows that the effects observed by Gleitman, must have the character of *directing the child's attention* to alternatives and distinctions that are in principle available to them, but which they might otherwise overlook, by sheer force of Bayesian priors on the conditional probability of a syntactic category given a semantic type, as seems to be implicit in Gleitman et al. (2005). In that sense, we should probably refer to this effect as "grammatical" bootstrapping, since it is an effect that is simultaneously both syntactic and semantic.[11]

10. Ambati, Deoskar, and Steedman (2013); Ambati, Deoskar, Johnson, and Steedman (2015); Ambati, Deoskar, and Steedman (2016b) and Ambati (2016) have shown a similar beneficial effect of categories as features for the MALT parser (Nivre, 2006).

11. This chapter is dedicated to the memory of Cathy Urwin (1949-2012), who made the unpublished observations of the child Emma described in section B.2.1

Appendix C
Natural Language Processing

One might seek to develop a more elaborate relation between statistical and syntactic structure than the simple order of approximation method we have rejected.
—*Syntactic Structures* Noam Chomsky, 1957:17,n.4

Parsers of all kinds can be thought of as consisting of three elements: first, a grammar for the language, defining an unbounded set of sentences and their meanings; second, an algorithm that systematically applies the rules of the grammar to the sentences of the language; third, an oracle to tell the algorithm which rules of the grammar are most likely to yield correct analyses (since in many cases it will be impracticable to explore all possibilities licensed by the grammar, and which of the completed analyses is most probably intended.

C.1 The Grammar

It can be difficult to identify the components of this tripartite structure in actual computational processors. In particular, in some neural-network-based language processors the grammar may be "hidden" in the connectivity of the network and hard to separate from the "language model" or oracle implicit in the weights on connections, both of which have usually been learned automatically. The implicit grammar may also be a "covering grammar" that is not the same as the linguists' grammar—that is the descriptively adequate grammar that supports the semantics. (In fact, it is conceivable that the meaning representation itself may at least in part be represented distributionally, as has been suggested by Baroni and Lenci, 2010.)

An interesting processor of this kind is reported by Vinyals, Kaiser, Koo, Petrov, Sutskever, and Hinton (2015), Building on work treating machine translation as a string transduction problem (Brown, Cocke, Della-Pietra, Della-Pietra, Jelinek, Lafferty, Mercer, and Roossin, 1990), this work treats parsing as the problem of transducing strings of English into labeled bracketed strings representing syntactic structure in supervised training data, using

LSTM (Hochreiter and Schmidhuber, 1997) with an attention mechanism, including training on a large corpus of automatically treebank-parsed data, and achieved results comparable to grammar-based treebank PCFG parsers. Although this is a surprising result, it is as much a reflection on the weakness of treebank parsers and the leniency of the EvalB measure of parser performance as on the quality of the parse trees that resulted. (A 100% EvalB score doesn't mean the parse is 100% linguistically correct.)

If one asks what the grammar is in such a parser, it isn't really the one represented by the bracketed strings. It seems rather to be a hidden high-order Markov process that is doing a fairly good job of covering the context-free grammar that is explicit in the treebank parser output that it was trained on. This implicit grammar is quite unlike the grammar it covers, and in fact there is no distinction between the algorithm and the grammar, while the oracle simply consists in the learned weights or probabilities of the various transitions.

To the extent that the output of the covering parser is correct according to the linguistic grammar it covers, so that its labeled bracketed output can be interpreted as a logical form, such a parser can support a traditional semantics. However, to actually *interpret* that logical form—say, in order to answer the corresponding question—requires a quite different grammar and automaton. This raises the question of why we didn't parse the strings with that grammar and automaton in the first place. A reasonable answer to that question might be that the high degree of ambiguity in the grammar made the LSTM parser the most efficient way to obtain the logical form. Another possible answer might that the LSTM can be used to directly buld a quite different kind of non-symbolic meaning representation that can do thes same job—for example, a vector-based representation derived by linear-algebraic composition from word-meanings derived from collocations (Osgood, Suci, and Tannenbaum, 1957; Landauer and Dumais, 1997; Mikolov, Yih, and Zweig, 2013c, *passim)*. However, the case would need to be made.

In the present appendix, we will confine our attention to parsers with an explicit grammar aligned with a compositional semantics, of the kind of which CCG is one representative. For the algorithm, we will follow earlier work in CCG and LFG in assuming that it is subject to the following condition:

(1) The parsing algorithm can only be defined over the types and rules that are licensed by the grammar.

This condition says that if the grammar licenses a derivation for a substring (i.e projects it onto a type via the combinatory rules)—say "Frankie gave" or

"Albert a book"—then the algorithm can build the corresponding representational structure (i.e logical form) and the oracle can pronounce on its plausibility relative to other analyses of that substring. If the grammar does not license a derivation/type for a string—say "give the" or "Albert a"—then the algorithm cannot build any structure, and the oracle has nothing to say.

This condition is referred to in *SP* as the "Strict Competence Hypothesis". It amounts to an assumption of transparency between the processor and the grammar that seems to be crucial to the (related) requirements for child language acquisition and the evolution of language considered in later appendices.

The strict competence assumption also has direct consequences for properties of the parser that are of interest to psycholinguists and cognitive modelers, as well as for engineers interested in bringing syntax and semantics into language models for automatic speech recognition (ASR) and statistical machine translation (SMT).

For example, if we want our parser to be incremental, as it must be if it is to contribute to such language-modeling, then the strict competence assumption says that if we use a standard context-free right-branching grammar for English, according to which none of the above strings are constituents, then our parser will be necessarily be almost entirely non-incremental. If we use a CCG grammar, according to which both "Frankie gave" and "Albert a book" are typable constituents, then more incrementality will be available for English, and even for entirely verb-final languages. However, the nonconstituent non-typability of "file without" or "Albert a" will limit incrementality to be less than word-by-word. In the parser of Vinyal et al., on the other hand, *all* of the above substrings, and in particular all prefix strings are typable, so an LSTM parser algorithm could in principle be completely incremental word-by-word.

The question then is whether CCG grammars are incremental *enough* to act as a model of human sentence processing, and to contribute maximally to language models for practical ASR and SMT. This is simply an empirical question: it is not obvious that an oracle based on either head-word-dependency statistics or semantic interpretation will have any more useful information after processing the prefix "Frankie gave Albert a" than it had when it had only gotten as far as "Frankie gave Albert". Studies showing incrementality in human sentence processing (Marslen-Wilson, 1973; Chelba and Jelinek, 2000; Demberg, Keller, and Koller, 2013) do not seem to have examined the question at this fine a grain.

C.2 The Algorithm

Parsers may approach the string and the grammar in a great many different ways, applying the rules either top-down from the start symbol(s) or bottom-up from the terminals or words, and from the start of the sentence to its end, that is left-to-right in English orthography, or from right to left, or any mixture of the above.

The central problem for such parsers, given the huge degree of ambiguity in natural grammars, is how to avoid duplicating analyses. For example, the following sentence has multiple analyses:

(2) I saw a man with a pocket telescope.

However, the subsequences “a man” and “a pocket telescope” have the same analysis as an NP in several of those analyses. It would therefore be efficient to only analyse them once and re-use the results in different analyses. Similarly, the perfectly legal NP “a pocket” may not be used in any analysis: it would be nice to either not build that NP at all, or at least to only waste time building it once.

There are several techniques for handling this problem. The most general techniques lie in the use of divide-and-conquer algorithms using charts or tables to store the results of previous analyses of particular substrings, of which the simplest is the bottom-up CKY algorithm. A second technique, borrowed from programming language compilation is to try to preempt search by using a lookahead of some fixed number of symbols, on the assumption that the language will be deterministically parsable with that information. (LR(k) is an example of such an algorithm. Programming languages are usually deterministic given a lookahead of 1.)

In the early days of computational linguistics, considerable attention was paid to psycholinguistic evidence that promised to tell us which of these alternatives was pursued by the human parser. For example, the Marcus (1980) parser can be seen as as investigating the possibility that natural language is LR(3)-parsable.[1]

However, once the machines got big enough to actually try the exercise on realistically-sized standard datasets, such as the Penn Wall Street Journal Tree-

1. Marcus’ conclusion was that human language was not LR(k) for any k, which is perhaps not suprising, since human processors do not seem to need any lookahead (Marslen-Wilson, 1973). However, a lookahead of three words in natural languages is remarkably informative, and many computational language models currently in use in ASP and SMT, and in greedy transition-based parsers use exactly this lookahead—see below.

bank (Marcus, Santorini, and Marcinkiewicz, 1993), it became clear that most of the grammars that had been built up to that time were far to small to cover realistic data, and also that it was impractical to parse grammars of the necessary size by exhaustive search. Whatever grammar and algorithm were used, the assistance of an effective oracle is essential to eliminate unpromising avenues and/or to rank the large number of analyses that natural grammars usually allow. Accordingly, the more recent literature has tended to use the simplest algorithms, usually either CKY, or Shift-Reduce aka Transition-Based parsing, both discussed in application to CCG in *SP*:ch.9. CKY is a table- or chart-driven algorithm, and the CCG parsers of Hockenmaier and Steedman (2002b) and Clark and Curran (2007b) are based on it.

Transition-based Shift-Reduce parsers resemble LR parsers in being driven by a table of transitions conditioned on the state of the stack and the next k items in the stack. They also have a chart of partially completed analyses. Both require the use of an oracle to limit search. A CCG parser of theis kind is described in *SP*:ch.9, and the CCG parsers of Ambati et al. (2015, 2016b) are of this kind. Ambati (2016) discusses a version of the latter parser which is truly incremental in the sense of not using the standard but psychologically implausible three-word lookahead. This version of the parser, which also uses a narrow beam search of width 16, and whose performance is comparable to the version with lookahead is under investigation as a psychological model of human sentence processing.

C.3 The Oracle

There are two types of mechanism that can act as an oracle for the natural language parsing algorithm, telling it which alternatives to favor and which to disfavor or discard entirely.

C.3.1 The semantic oracle

One psychologically plausible but computationally problematic mechanism depends on access to semantic interpretations. If the previous example (2) had been the following, the access to an interpretation may tell us that the attachment of "with a pocket telescope" as a VP adjunct is more consistent with our knowledge of the world than the interpretation as an NP adjunct, as well either of the analyses where "saw" is interpreted as a cutting action.

(3) I saw a bird with a pocket telescope.

Early research by Marslen-Wilson (1973) showed that human subjects show effects of similar semantic anomalies with very short latency (less than a word) during the actual course of sentence processing, using speech shadowing as a measure. Altmann and Steedman (1988) showed similar effects of referential context on modifier attachment, using a reading time measure.

However, no one knows how to build a semantics of this kind that can be applied at the required scale, or how to to represent world knowledge in a way that will make it practically usable for such a task in realistic time. In the next section we will look at some proposals for such a semantics, but in the meantime there is an alternative more practical oracular mechanism that is widely used.

C.3.2 Statistical parsing models

The alternative is a statistical parsing model representing the likelihood of such aspects of the derivation as a "with" adjunct headed by "telescope" modifying a noun group headed by "bird" in comparison with it modifying a verb group headed by "saw". together with the likelhoods of seeing birds vs. sawing birds, and all other dependencies in the various analyses.

Such parsing models are called "head-word-dependency" models, and they are obtained from corpora by counting frequencies of all the events of all the kinds that are allowed by the grammar itself, where an "event" is the application of any rule of the grammar. (Or course we have to "smooth" the model to allow for the fact that in any finite corpus some of the events allowed by the grammar will not actually occur.)

Usually, such corpora are hand-annotated with correct grammatical analyses ("supervised" learning). (It is in theory possible to induce grammars and models (or some approximation to them), given sufficient amounts of unlabeled data, ("unsupervised" learning), but in practice this has not worked well.)

The first such corpus was the one million word Penn Wall Street Journal corpus of news text(Marcus et al., 1993). A CCG version of the corpus was automatically obtained from the original by Hockenmaier (2003a); Hockenmaier and Steedman (2007), and was improved in respect of NP structure by Vadas and Curran (2007, 2008) and Honnibal, Curran, and Bos (2010a); Honnibal, Kummerfeld, and Curran (2010b).

Such "generative" models are very powerful, because they approximate a mixture of semantics and world-knowledge, reducing inference to purely numerical computation. They are very attractive in psychological terms because they are entirely determined by the grammar and the data. For the same reason, they can be used to guide generation as well as parsing. Hockenmaier's parser uses a generative model, and so does the model of language acquisition in the next appendix.

There is a second "discriminative" kind of statistical model, based on the Perceptron (Rosenblatt, 1962) that is not limited in this way, and for that reason can be more powerful than the generative model. This class of models can be ignored for present purposes, but the Clark and Curran (2007b) parser uses such a model, as do most CCG parsers since.

Probabilistic models can be used to assign (conditional) probabilities to analyses using Bayes Theorem in two ways. First, "inside" estimates of probability can be assigned to elements in the chart, and to transitions in the shift-reduce transition table. Second, very low probabilities can be used as a criterion for dropping analyses entirely from further analyses, in a procedure called "beam search". In the case of a chart-based algorithm such as CKY, this amounts to pruning the chart. In the case of the shift-reduce parser, which lends itself to a probabilistic model for the transition table itself and to "greedy" incremental parsing (Nivre, 2003, 2004, 2006), this may amount to just keeping the the most likely analyses of some left prefix of the string. (Some beam is needed just to deal with lexical ambiguity of the most recent words.)

The danger of beam search is that the beam is quite likely to end up only containing very similar analyses that all stem from an early, possibly wrong, decision about the most promising analyses. However, some pruning is usually a necessary tactic.

C.3.3 Supertagging

Lexicalized grammars such as LTAG and CCG, in which there is a large number of category types carrying a large amount of language-specific syntactic information lend themselves to a technique known as "supertagging". Supertagging is so-called, not because of any specific relation to TAG grammars (although they were first proposed by Bangalore and Joshi (1999) for TAG parsing), but by analogy to part-of-speech (PoS)-tagging. Supertagging also chooses the most contextually likely categories for lexically ambiguous words, differing only in assigning between hundreds and thousands of category types, rather than the fifty or so Brown/WSJ PoS tags. Like a PoS-tagger, a supertagger is an essentially Markovian sequence model, independent of the grammar, that is used as a front end to parsing to generate *n*-best lists of categories for input to the parser.

Clark and Curran (2007b) discovered a particularly effective mode for supertagging, called *adaptive* supertagging. Rather than taking the *n* best categories for an ambiguous word, the supertagger proposed only categories within a narrow highest-probability beam. Only if it proved impossible to find a parse with those highest ranking categories was the beam widened to allow lower-ranked alternatives to be considered. This has the advantage of increasing speed very greatly, without significant impact on accuracy.[2]

Lewis and Steedman (2014c) showed that supertagging could be made more effective by using collocation-based vector word embeddings (Turian, Ratinov, and Bengio, 2010; Collobert, Weston, Bottou, Karlen, Kavukcuoglu, and Kuksa, 2011; Mikolov, Chen, Corrado, and Dean, 2013a) as features in a CRF model, which could be tuned over large amounts of unlabeled text. Part of the improvement in supertagging accuracy came from the fact that embedding supertagger could be trained without using PoS tags as features. (Automatically-obtained PoS tags are notoriously inaccurate because of poor quality training data, and are a major source of error in parsing models that use them.) Embeddings-based supertagging is also more robust to out-of-domain parsing, because of the diversity of the unlabeled text used to tune them. Lewis and Steedman (2014a) showed that embeddings-enhanced adaptive supertagging was accurate enough for the remainder of the parsing task to be done by exhaustive A^* search. Xu, Auli, and Clark (2015) show further improvements from the use of similar features in a Recurrent Neural Network (RNN) model. Lewis, Lee, and Zettlemoyer (2016) show that an LSTM based CCG supertagger lends itself to a threaded implementation using a GPU, with an attendant increase in speed of over an order of magnitude.

Socher, Bauer, Manning, and Ng (2013) make a similar use of embeddings as features analogous to head-words or Brown clusters in PCFG parsing models, learning embedding-specific rules for composing embeddings to yield embedding features for parents. While Socher et al. treat the vectors themselves as meaning representations, leaving the usual questions about how logical operators can be represented in distributional terms, Krishnamurthy and Mitchell (2014) use the embedding strictly as features of the model, building standard logical form meanings using λ-terms in the way used elsewhere in this book. LSTM neural network models have also been used as models for transition-based Shift-Reduce parsing (Chen and Manning, 2014; Dyer, Ballesteros, Ling, Matthews, and Smith, 2015; Dyer, Kuncoro, Ballesteros, and Smith, 2016), following earlier transition-based shift-reduce parsers using

2. This effect seems to arise because a successful parse with high probability lexical categories is going to be preferred in any case over most parses with lower-probability lexical categories—so the latter are a waste of effort unless they are the *only* parse.

neural network-models (Henderson, 2003; Henderson, Merlo, Musillo, and Titov, 2008). Kuncoro, Ballesteros, Kong, Dyer, Neubig, and Smith (2017) presents a fully incremental version without lookahead.

It is an empirical question whether the latter kinds of model or the supertagger front-end will turn out to be the most effective way of combining statistical and grammar-based parsing, but it seems likely that the way forward lies in the hybrid methods described in this section.

C.4 Discussion

All of the parsing techniques described above are quite generally applicable to a wide variety of grammar formalisms, such as context-free and depencdency parsers, as well as CCG. The advantage of the latter is mainly to do with its treatment of long-range syntactic dependency and the close relation it embodies between syntax and semantics.

Because minmalist grammars involve potentially non-monotonic structure-changing operations, a little more needs to be said about them. There have been many parsers based on some version of minimalist programmatic grammar, some computationally oriented (Fong, 1991; Lin, 1993, 1994), and some more psychologically motivated (Weinberg, 1998). Among the most interesting for the present purpose are those of Harkema (2001); Stabler (2011), and Torr and Stabler (2016), which are based on the "Bare Phrase Structure" of Chomsky (1995a) which can be seen as reducing the theory to pure categorial grammar, with the addition of movement (defined monotonically in terms of structure-preserving copying) in place of CCG combinatory rules (cf. Berwick and Epstein, 1995a,b). The parsers of Harkema and Stabler, and in particular Torr, make this explicit, with a lexicalized notation analysed in the Discussion section of chapter 2 that is a notational variant of the categorial slash. While the rest of the grammar makes use of movement, and is of greater expressive power (at at least the level of fulll LCFRS, the computational interpretation of the program in MG is illuminating, and a treebank parser and parsing model is proposed by Torr (2019).

CCG itself has been widely applied in computational NLP applications, particularly those requiring syntax to support semantics: Almaghout, Jiang, and Way 2010, 2011; Almaghout 2012; Artzi, Das, and Petrov 2014; Artzi, Lee, and Zettlemoyer 2015; Auli 2011; Auli and Lopez 2011a,b; Birch, Osborne, and Koehn 2007; Clark and Curran 2007b,a; Dun, Sun, and Wan 2015; Espinosa, White, and Mehay 2008; Espinosa, Rajkumar, White, and Berleant

2010; Garrette, Erk, and Mooney 2014; Garrette, Dyer, Baldridge, and Smith 2015; Giuliani and Knoll 2007, 2008; Glass and Yates 2014; Hassan, Sima'an, and Way 2008; Hermann and Blunsom 2013; Honnibal et al. 2010a,b; Kato and Matsubara 2015; Krishnamurthy and Mitchell 2013, 2014; Lee, Lewis, and Zettlemoyer 2016; Lewis et al. 2016; Maillard, Clark, and Grefenstette 2014; Matuszek, Herbst, Zettlemoyer, and Fox 2012b; Matuszek, Fitzgerald, Zettlemoyer, Bo, and Fox 2012a; Mehay and Brew 2012; Misra and Artzhi 2016; Nădejde, Reddy, Sennrich, Dwojak, Junczys-Dowmunt, Koehn, and Birch 2017; Ng and Curran 2012; Rajeh, Li, and Ayedh 2016; Uematsu, Matsuzaki, Hanaoka, Miyao, and Mima 2013; Wang, Kwiatkowski, and Zettlemoyer 2014; White and Baldridge 2003; White 2004; Wu, Zhang, and Zong 2016; Xu, Clark, and Zhang 2014; Xu et al. 2015; Xu 2016; Yoshikawa, Noji, and Matsumoto 2017; Zettlemoyer and Collins 2005.

Many of these computational applications exploit the "surface compositional" semantics of CCG. In particular, Hockenmaier (2003b); Clark and Curran (2004); Lewis and Steedman (2014c,a); Lewis et al. (2016) provide publicly available efficient parsers trained on WSJ; Birch et al. (2007); Hassan, Sima'an, and Way (2009); Mehay and Brew (2012); Nădejde et al. (2017) use CCG in statistical machine translation; Prevost (1995); White (2006) apply it to sentence realization Briscoe (2000); Kwiatkowski, Zettlemoyer, Goldwater, and Steedman (2010); Kwiatkowski et al. (2012); Krishnamurthy and Mitchell (2012); Abend et al. (2017) apply it to semantic parsing and modeling child language acquisition; Bos and Markert (2005); Lewis and Steedman (2013a) apply it to open-domain question answering and entailment.

Appendix D

Towards a Form-Independent Semantics

Man kann für eine *große* Klasse von Fällen der Benützung des Wortes "Bedeutung"—wenn auch nicht für *alle* Fälle seiner Benützung—dieses Wort so erklären: Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache.[1]
—(*Philosophische Untersuchungen*, Ludwig Wittgenstein, 1953:¶43)

The previous appendix has described some ways in which CCG and other near-context-free grammars can be efficiently parsed with somewhat usable accuracy, including the buliding of logical form, even for unbounded dependencies if we care about them, at speeds of hundreds or even thousands of sentences a second. Google is parsing everything we type at it. So, why don't we have real QA, where we ask a question such as "Is the president in Washington DC today?", have it mapped to an equivalent query and get an precise answer based on a semantic net of knowledge built offline by semantic parsers, rather than a bunch of snippets from pages whose words and linkages offer some chance of answering our question when we ourselves do the reading?

The central problem in using parsers to answer questions from unrestricted text like this is that the answer to our question is very likely to be there somewhere, but that it is almost certainly there in a form which is not the same as that suggested by the form of our question. For example, the question "Is the President in Washington DC?" is in fact answered by the statement in today's paper that "Obama has arrived at the White House". However, understanding this requires inferences that "having arrived" at a place at a time entails "being at" that place at that time, that being at the White House entails being in Washington, and so on. We draw all of these inferences effortlessly ourselves when we read the latter sentence. However, the standard semantics for that sentence, of the kind that we have used up to this point in this book says that the semantics of our question is something like $present\,(in\,washington\,sk_{president})$, while that of the text is $present\,(perfect\,(arrived\,whitehouse\,obama))$

1. "For a *large* class of occasions of use of the word 'meaning'—though not for *every* occasion of its use—this word can be defined thus: the meaning of a word is its usage in the language."

D.1 Decompositional Lexical Semantics

Many linguists, starting with the generative semanticists of the late '60s have tried to build a form-independent semantics. The following are various attempts to specify the meaning of the sentence "Bugs kill plants" in something like the language of mind:

(1) Montague, 1973: $\forall x[bug\, x \Rightarrow \exists y[plants\,(y) \wedge kill\, y\, x]]$
McCawley, 1968: $[_S CAUSE\ BUGS [_S BECOME [_S NOT [_S ALIVE\ PLANTS]]]]$
Dowty, 1979b: $[CAUSE[DO\ BUGS\ \varnothing][BECOME \neg [ALIVE\ PLANTS]]]$
Talmy, 2000: $Bugs\ ARE\text{-}the\text{-}AUTHOR\text{-}OF[plants\ RESULT\text{-}TO\text{-}die]$
Van Valin, 2005: $[do\,(bugs\,\varnothing)]CAUSE[BECOME[dead\,(plants)]]$
Goddard, 2010: *BUGS* do something to *PLANTS*; because of this, something happens to *PLANTS* at the same time; because of this, something happens to *PLANTS*'s body; because of this, after this *PLANTS* are not living anymore.

Other related representations are graphical, such as that of Schank, 1972, in which the left-right arrow $\longleftrightarrow$ represents the subject dependency, while the double up-arrow $\uparrow\uparrow$ represents the causal dependency of the plants' death upon the ACT of the bugs:

(2)
$$\begin{array}{ccc} \text{bugs} & \longleftrightarrow & \text{do} \\ & \uparrow\uparrow & \\ \text{plants} & \longleftrightarrow & \text{die} \end{array}$$

Such a semantics offers the attraction of being language-independent, and has the considerable advantage of being immediately compatible with inference using first-order logical opeators such as negation. Thus, one could in principle deduce an answer to the question "Are the plants alive?" from the text "The bugs killed the plants", or the equivalent in another language, or be used to support machine translation. However, such semantics was confined to small fragments, and remained somewhat language-specific (Dorr, Passonneau, Farwell, Green, Habash, Helmreich, Hovy, Levin, Miller, Mitamura, Rambow, and Siddharthan, 2010).

Related attempts at a decompositional semantics have been more recently realized semi automatically as computational lexical resources, including WordNet (Fellbaum, 1998), FrameNet (Baker, Fillmore, and Lowe, 1998), VerbNet/PropBank (Hwang, Bhatia, Bonial, Mansouri, Vaidya, Xue, and Palmer, 2010), BabelNet (Navigli and Ponzetto, 2012), AMR (Banarescu, Bonial, Cai, Georgescu, Griffitt, Hermjakob, Knight, Koehn, Palmer, and Schneider, 2012), and the named entities of the Google Knowledge Graph (Singhal, 2012).

Many of these resources may help to identify the fact that “the President” and “Mr. Obama” refer to the same entities, and that the White house is a location in Washington.

However, such hand-built semantic resources are invariably incomplete, in the sense that they leave out many relations, usually because such resources are built consciously or unconsciously for human users, and omit entailments that humans find too obvious ever need to state.

For example, at the time of writing, the FrameNet entry for the verb “arrive” tells us a great deal about the verb “arrive”, but omits the information that the *result* of the *theme* arriving at the *goal* is that the former is situated *at* the latter, which is what the relation *perfect* in the logical form for the text in our running example needs to access in order to know that the text does actually answer the question.[2]

Of course, this lacuna is easy enough to fix, but there are many more (such as that *not being at the goal already* is a precondition of *arriving*). It is hard to believe that such resources will ever be complete enough to support our hypothetical question answerer.

D.2 Semantic Primitives as Hidden Variables

This realization prompts the following thought: why not let parsing and machine learning do the work of completing the semantics instead, using the “machine reading” approach of Etzioni, Banko, and Cafarella (2007) and Mitchell, Cohen, Hruschka, Talukdar, Betteridge, Carlson, Dalvi Mishra, Gardner, Kisiel, and Krishnamurthy (2015) to mine “hidden” entailment relations such as that between *arriving* and *being at* a place?

There are two active approaches to this problem. The first, originating woith the “Semantic Differential” of Osgood et al. (1957), treats the meaning of content words as a location in a high-dimensional vector space, in wh8ich the dimensions can be thought of as all the other content words of the language, although this is a space of such high dimensionality and such sparse occupancy that it must be reduced by some technique (the original semantic differential used Principal Components Analysis). This reduction must be such as to preserve the proximity of words that occur in similar contexts. Closeness in the space then represents relateness in meaning (although relatedness tends to in-

2. `https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Arriving`

clude antonymy as well as synonymy).

The attraction of such representations is that one can think of their composition into phrasal and sentence levels as accomplished by linear-algebraic operations like vector addition and multiplication (Church and Hanks, 1989; Smolensky, 1990; Landauer and Dumais, 1997; Lin, 1998; Baroni and Zamparelli, 2010; Grefenstette and Sadrzadeh, 2011; Padó and Lapata, 2007; Mikolov, Sutskever, Chen, Corrado, and Dean, 2013b; Bordes, Usunier, Garcia-Duran, Weston, and Yakhnenko, 2013; Mitchell and Steedman, 2015; Guu, Miller, and Liang, 2015; Neelakantan, Roth, and McCallum, 2015; Weir, Weeds, Reffin, and Kober, 2016, *passim*).

Vector "embeddings" representing all the contexts a word has been encountered in can be trained by unsupervised methods over vast ampounts of text, and can be very effective in disambiguating unseen words. We noted that when used as features to tune a supervised parsing model, they can be very effective in deciding which seen events in the supervised model most resemble unseen events in unseen text (Henderson, 2003; Henderson et al., 2008; Chen and Manning, 2014; Lewis and Steedman, 2014c,a; Dyer et al., 2015, 2016)

However it is doubtful that we should think of vectore representations as word *meanings*. In particular, it remains completely unclear how to make such representations compatible with the logical operators such as negation that are crucial to tasks such as question answering.

Moro and Navigli (2012), Navigli and Ponzetto (2012), Nakashole, Weikum, and Suchanek (2012), and Grycner and Weikum (2014); Grycner, Weikum, Pujara, Foulds, and Getoor (2015) show that relational ontologies, including multilingual ones, can be built by mining text concerning recognizable named entities.

Lewis and Steedman (2013a,b, 2014b) and Lewis (2015) propose to mine text concerning typed named entites such as the person named by *Mr. Obama* and the office named by *President* for consistent directional entailments, such as that if we read about a person *being elected* to an office, we often also read about them running for that office (but not *vice versa*). Such candidate entailments will be probabilistic and noisy, and are inherently distributional (since *the president* is sometimes a person and sometimes an office). But Lewis and Steedman (2014b) follow Berant, Alon, Dagan, and Goldberger (2015) in exploiting the transitivity of entailment to to make cleaner entailment graph out of the candidate entailments using integer linear programming.

For example, the typed named-entity technique is applied to (errorfully) estimate local probabilities of entailments using an asymmetric similarity measure such as Weeds precision (Weeds and Weir, 2003), giving data that might look like the following simplified example for pairs of countries *xy*:

(3) a. $p(\textit{conquer}\,x\,y \Rightarrow \textit{invade}\,x\,y) = \textit{0.9}$
b. $p(\textit{invade}\,x\,y \Rightarrow \textit{attack}\,x\,y) = \textit{0.8}$
c. $p(\textit{invasion}\,(\textit{of}\,x)\,(\textit{by}\,y) \Rightarrow \textit{attack}\,x\,y) = \textit{0.8}$
d. $p(\textit{invade}\,x\,y \Rightarrow \textit{invasion}\,(\textit{of}\,x)\,(\textit{by}\,y)) = \textit{0.7}$
e. $p(\textit{invasion}\,(\textit{of}\,x)\,(\textit{by}\,y) \Rightarrow \textit{invade}\,x\,y) = \textit{0.7}$
f. $p(\textit{conquer}\,x\,y \Rightarrow \textit{attack}\,x\,y) = \mathbf{0.4}$
g. $p(\textit{conquer}\,x\,y \Rightarrow \textit{conqueror}\,(\textit{of}\,x)\,y) = \textit{0.7}$
h. $p(\textit{conqueror}\,(\textit{of}\,x)\,y \Rightarrow \textit{conquer}\,x\,y) = \textit{0.7}$
i. $p(\textit{bomb}\,x\,y \Rightarrow \textit{attack}\,x\,y) = \textit{0.7}$
(etc.)

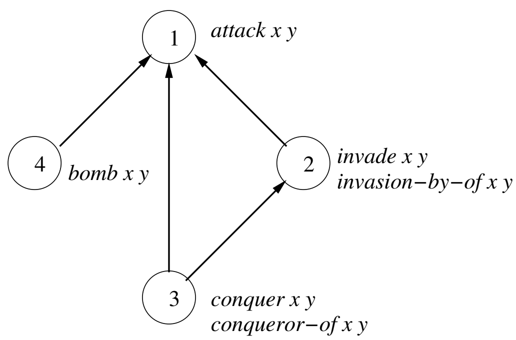


Figure D.1: A simple entailment graph for relations between countries.

These local entailment probabilities are used to construct an entailment graph shown in figure D.2 using integer linear programming with a prior $p = 0.25$ with the global constraint that *entailment graphs must be closed under transitivity*.

Thus, (3f) will be correctly included despite low observed frequency because it is supported by the transitivity of entailment, while other low frequency spurious local entailments will be dropped.

"Cliques" within the entailment graphs—that is, groups of relations that all mutually entail each other, and are therefore paraphases—can be collapsed to a single cluster relation identifier, such as $\textit{rel}_{399}$.

On the basis of this graph of entailments, we can take the categorial lexicon used throughout this book, and transform it into something better adapted for question answering, by replacing the monolithic Montague-style predicates by conjunctions of paraphrase cluster identifiers for their meaning and all their entailments. For example, the verbs related to the *attack* entailment graph will now look something like the following:

$$
\begin{array}{rl}
\text{(4) attack} & := (S\backslash NP)/NP : \lambda x \lambda y.rel_1\, x\, y \\
\text{bomb} & := (S\backslash NP)/NP : \lambda x \lambda y.rel_1\, x\, y \wedge rel_4\, x\, y \\
\text{invade} & := (S\backslash NP)/NP : \lambda x \lambda y.rel_1\, x\, y \wedge rel_2\, x\, y \\
\text{conquer} & := (S\backslash NP)/NP : \lambda x \lambda y.rel_1\, x\, y \wedge rel_2\, x\, y \wedge rel_3\, x\, y \\
\text{conqueror} & := VP_{pred}/PP_{of} : \lambda x \lambda p \lambda y.p\, y \wedge rel_1\, x\, y \wedge rel_2\, x\, y \wedge rel_3\, x\, y
\end{array}
$$

Such logical forms immediately support correct inference under negation, such as that *conquered* entails *attacked* and *didn't invade* entails *didn't conquer*.

To answer a question "Did Y invade X?" we therefore look for sentences whose logical forms subsume the conjunctive logical form $rel_2\, x\, y \wedge rel_1\, x\, y$, or satisfy its negation $\neg rel_2\, x\, y \vee \neg rel_1\, x\, y$. Similarly, the fact that that *conqueror-of* is a paraphrase of *conquer* $= rel_2$, means we can immediately infer that *conqueror-of* entails *attack* $= rel_1$.

An example of open-domain questions succesfully answered from unseen text is shown below. (Following Poon and Domingos (2008), the questions were artificially generated by replacing arguments in parsed web text to generate wh-question which were then answered on the basis of unseen text of the same genre. See Lewis and Steedman (2013a); Lewis (2015) for further details and experiments.)

(5)

Question	**Answer**	**From Unseen Sentence:**
What did Delta merge with?	Northwest	The 747 freighters came with Delta's acquisition of Northwest
What spoke with Hu Jintao?	Obama	Obama conveyed his respect for the Dalai Lama to China's president Hu Jintao during their first meeting
What arrived in Colorado?	Zazi	Zazi flew back to Colorado...
What ran for Congress?	Young	... Young was elected to Congress in 1972

It should be apparent at this point that we can collect local entailments between expressions in languages other than English, provided that we can parse and type the named entities in the language concerned. Lewis and Steedman (2013b) report an extension of the paraphrase/entailment semantics to French, and apply it to the task of reordering Moses phrase-based statistical machine translations from French sentences to English. The bilingual semantics is evaluated by parsing the top 50 English translations and reordering them according to how well they preserve the multilingual entailment-based semantics obtained by parsing the original French. Where this process prefers a translation that is different from Moses' own top-ranked translation, bilingual judges are asked which they prefer. In 39% of cases where there is a difference, the judges prefer the reranked alternative, compared to the 5% of cases in which they prefer the Moses 1-best. (Many of the remaining 56% of cases in which there was no preference are ones in which the difference between the candidates was a matter of a syntactic attachment which was not available to the judges from presentations of the mere strings.)

An example of a successful reordering of Moses SMT translations is the following:

(6)

Source:	Le Princess Elizabeth arrive à Dunkerque le 3 août 1999
SMT 1-best:	The Princess Elizabeth is to manage to Dunkirk on 3 August 1999.
Reranked 1-best:	The Princess Elizabeth arrives at Dunkirk on 3 August 1999.

See Lewis and Steedman (2013b) for detailed results and further experiments.

D.3 Temporal Semantics

A great variety of semantic information can potentially be mined in this way. If the text that we are mining is datelined, as news material usually is, then we should be able to work out that the entailments associated with people visting places in the graph in figure D.3 are temporally (or rather, causally) ordered, and that *being there* is the result of *arriving*, and therefore an entailment of *having arrived*, as in the example with which this section began. Certain finer distinction, such as that between the present and the futurate can be drawn on the basis of temporal modifiers, as in *visits/is visiting Hawai'i next week*, whose automatic extraction has been investigated by Chambers, Cassidy, McDowell, and Bethard (2014), using supervised learning over labeled resourced such as UzZaman, Llorens, Derczynski, Allen, Verhagen, and Pustejovsky (2013).

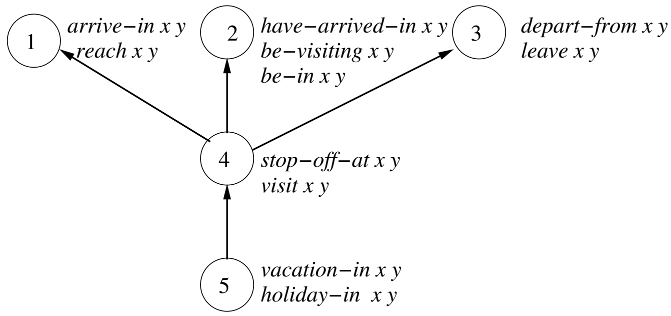


Figure D.2: A temporal entailment graph for people visiting places.

We may also expect to find aspectual "coercions" of the kind discussed by Moens and Steedman (1987, 1988) and Pustejovsky (1998), such as futurate *is visiting*, and nominals *starting the visit*, *finishing his vacation*, and the like.

These finer distinctions in varieties of entailment between relations can be discovered automatically from data like the following concerning particular pairs of named entities, from the University of Washington NewsSpike corpus (Zhang and Weld, 2013): VERBATIM COMMENTED OUT TEMPORARILY

In such data, we find that statements that so-and-so *is visiting*, *is in* and the perfect *has arrived in* such and such a place, occur in stories with the same datestamp, whereas *is arriving*, *is on her way to*, occur in preceding stories, while *has left, is on her way back from, returned*, etc. occur in later ones.

This information provides a basis for inference that *visiting* entails *being in*, that the latter is the consequence of *arriving*, and that *arrival* and *departure* coincide with the beginning and end of the progressive state of *visiting*.

D.3.1 Neo-Reichenbachian Semantics of Temporality

In order to capture the semantics behind these intuitions, we follow Moens and Steedman (1988); Horn (1990); Steedman (1997), and Portner (2003) in assuming a neo-Reichenbachean semantics of tense, aspect, and modality.

Reichenbach (1947) identified three temporal entities underlying the semantics of the tensed verb group. These were: S—the *speech time*, or the time of the speech-act itself; R—the *reference time*, or the time referred to; and E—the *event time*, the time of the eventuality identified by the main verb.

These entities may all be temporally disjoint, or may coincide or overlap. For example, in the case of the pluperfect in (7a), S is the time of utterance, R is at the time before S that we are talking about, and E is at a time before R:

(7) a. My luggage had arrived.
  b. We leave at dawn.
  c. He is driving to London.

In (7b), on the other hand, R the reference time is in the future, after S, and R and E coincide at dawn. In (7c), S and R coincide .

Tense—past in (7a), the futurate present in (7b), and the simple present in (7c)—defines the relation between R and S as precedence in (a), succession in b and as identity in (c).[3]

Reichenbach seems to have concieved of S, R, and E as undifferentiated monolithic intervals and their relations as purely temporal. However, the relation between R and E, which is defined by progressive and perfect Aspect, respectively marked in English by the auxiliary verbs *be* and *have*, is not a purely to temporal relation between intervals.

The effect of the progressive auxilliary *be* is to turn the eventuality into a progressive state. The identity of that state depends on the type of the eventuality. Events are anatomized by Moens and Steedman (1988), following Vendler (1957), as falling into four types or *aktionsarten*, as follows:

(8)

Name	Type	Example	Grammatical test
Accomplishments	+telic, +durative	drive to London	#for an hour/in an hour/#at dawn
Achievements	+telic, −durative	arrive in London	#for an hour/in an hour/at dawn
Activities	−telic, +durative	drive	for an hour/#in an hour/#at dawn
Point	−telic, −durative	start/stop driving	#for an hour/#in an hour/at dawn

The symbol "#" on a test such as combination of "drive to London" with "at dawn" means that the combination is impossible *without a change in the type of the event*, in this case to something like "start to drive to London". (It is important to remember this point, because almost any of these combinations is possible with such "coercions" to different event types.)

The effect of simple past tense on these event types is simply to identify the entire extent of the eventuality time E with the anterior reference time. For the telic accompishments and achievements, this entails that the goal of the event—in this case being in London—was achieved.

3. More accurately, past tense defines the reference time as *distinct from* the situation of utterance, since past tense is also a marker of counterfactuctual modality, as in its use in counterfactual conditions (Isard, 1974).

(9) a. He drove to London.
b. He arrived in London.
c. He drove.

The effect of the progressive auxiliary is to turn the core eventuality into a *progressive state*. The type of this state is determined by the above eventuality types, as in the following examples:

(10) a. He was driving
b. He was driving to London
c. He was arriving in London

The progressive of an activity (10a) says that the R is anterior to S and that the eventuality E is a progressive state of *him driving* holding at R, where the start and stop points of E are undefined.

The progressive of an accomplishment (10b) is almost identical to (a). The progressive state is *his driving with the goal of being in London*, and it holds at the anterior reference time. However, it is not entailed that the goal was achieved: it is perfectly consistent to continue "but the car broke down and he never got to London."

The progressive of an achievement (10c) says that the progressive state holding at R was not the *arriving* but an inferrable activity that would normally result in arrival, such as *his driving the last part of the route to London*.

Thus the three examples int (10) have rather similar truth conditions. One might think of this as the progressive auxilliary turning everything into the nearest related activity. Often this is what Moens and Steedman (1988) called the *preparatory activity*, but it may also be iteration of the core event, as in "I am dating Jackie"

Since most states do not have associated preparatory activities, and nor can they iterate, they can only combine with the progressive by rather extreme coercions to events. Thus (a) below seemws to refer to *repeatedly showing that you now the answer when asked*, while the slogan from McDonalds (b) seems to refer to repeated feelings of pleasure at some iterative activity:

(11) a. #I am knowing the answer (these days).
b. #I'm lovin' it.

The perfect auxiliary has a similar effect of mapping events onto states. The states in question are what Moens and Steedman (1988) called the *consequent state* of the core event.

(12) a. I have driven to London
b. I have arrived in London
c. #I have driven

Thus, the perfect of an accomplishment (a) and an achievement (b) are true just in case the consequent state (in this case, *my being in London*) hold at the (present) reference time. The perfect of an activity is only acceptable to the extent that the activity has an accessible consequent state, such as *my probably still remembering how to drive*.[4]

Like the first-order logical operators such as quantifiers and negation, the tense operators relating S and R can be hand-coded into the lexicon, either via the morphology or semi-automatically for unanalysed verbs.

However, this is unlikely to be possible for the aspectual operators. There are no equivalents of gazeteers or WordNet where we can look up the consequent state of preparatory activity of *arrival*. However it does seem possible in principle to learn that a typed multi-word relation *PERSON have arrived in PLACE* entails *PERSON be in PLACE* using the method outlined above.

However, there are limits to what we can expect text mining of this kind to discover. It is also important in answering questions about prople's whereabouts to understand that *nothing can be in more than one place at a time*. This knowledge is probably too banal to be mentioned in text, ever.

Moreover, this is not the kind of knowledge we want to put into the lexical entailments of *being there*. It is the kind of non-linguistic knowledge that we share with other animals. (My cat of blessed memory was quite clear on the point that a mouse could only be in one place at a time.) It is the kind of knowledge that we hard-wire into the knowledge representation for our robots planning systems, as in the STRIPS solution to the Frame Problem in planning (Fikes and Nilsson, 1971), which among other things builds in such knowledge as that, if you move something, then you don't change the containment relations with things that it may contain. (This has the pleasing consequence that if my bicycle is on a train in London, and the train goes to Edinburgh, I know that my bicycle is in Edinburgh and not in London without ever having to invoke an axiom that things can only be in one place at a time.) We should probably build that in to the natural logical "language of mind" that underpins language acquisition and language evolution in the same way, rather than making it explicit in the lexicon. (We will return to the topic of the language of mind in appendices B and A.)

A semantics of this kind, acquired by machine-learning of "hidden" seman-

4. Of course, in context, other coercions that those suggested here may be possible.

tic primitives from text, builds common-sense entailment into logical form itself, so that simple entailments can be derived directly from sentential meanings, raher than by theorem proving search. It can be seen as a lexicalized implementation of Carnap's 1952 "meaning postulates", which he proposed to make the repository of the knowledge that *PERSON being a batchelor* entails *PERSON being male* and *unmarried*. It can also be seen as a practical implementation of Wittgenstein's 1953:¶43 identification of the meaning of (content) words with "use", cited in the epigraph to this appendix.

Such a semantics, if refined considerably further than it has been so far, especially by the use of multilingual data (Lewis and Steedman, 2013b; Christodoulopoulos and Steedman, 2015), might ultimately approach the hidden conceptual language to which the child must have access in order to hang language-specific grammar onto it during first language acquisition, as considered in the next appendix.

Appendix E

The Devil's Dictionary

DICTIONARY, n. A malevolent literary device for cramping the growth of a language and making it hard and inelastic. *This* dictionary, however, is a most useful work.
—*The Devil's Dictionary* Ambrose Bierce, 1911

Linguists work with a lot of abstract concepts. Sometimes its hard to remember what the distinctions are. Sometimes they use the same term in different ways. Sometimes they use familar terms with unfamiliar and even counterintuitive meanings. This glossary is intended to help the perplexed, by giving the definition used in this book, and indicating where the reader may encounter other conflicting usages.

Absolute Construction
Accent metrical, pitch
Accusative: see Case
Adequacy: (of a theory)
Adjective
Adjunct
Adverb adverbial modifier
Agreement
Aktionsart
Algorithm
Anaphor:
Affix: hopping; prefix suffix
Argument Structure
Aspect
Base Generation
Case
Clitic: pro-; en-
Combinator
Comment/Rheme
Competence
Complexity

Complement
Compositional Semantics
Computation
Concord
Conservative (of quantifier)
Construal
Construction
Context
Context-free
Context-sensitive
Control: adjunct; arbitrary
Do-support
Deep Structure
Dependency
Determiner
Embedding
Endotypic
Equivalence, of Theories
Equivalence, Weak/Strong
Exceptional Case Marking
Exotypic
Given/background (Information)
Graph
Govern: government; governor
Head see govern
Head-word dependency model
Information Structure
Intonation
Island Effect
Labeled/Unlabelled data
Lambda Binding
Lambda Calculus
Lambda Term
Lambek Calculus
Lexical Redundancy Rule
Lexical Rule
Linear Indexed Grammar
Logical Form
Logophor

Mechanism
Meter metrical
Model logical psychological computational
Mood (modality)
Movement
Multiple Context-free
New/Contrastive (information)
Notational Variant: (of Theories)
Node:terminal, nonterminal
Noun
Oracle
Parse parser parsing
Parasitic Gap/Extraction
Perfect/Imperfect
Perfective/Imperfective
Performance
Phrase Stucture
Polymorphic
Preposition
Predicate
Predicate-Argument Structure
Progressive
Proposition
Prosody
Raising
Reciprocal (pronoun)
Reference
Referent
Reflexive (pronoun)
Sense
Slash
Spell-out
Stress (metrical)
Subjacency
Subject
Supervised/Unsupervised learning
Surface Structure
Surface-Compositional Semantics
Surface-Compositional Semantics, Direct

Syntax!Narrow
Telic/Atelic
Topic/Theme
Tree
Verb
Voice passive, optative, reflexive, reciprocal

References

Abbott, Barbara, 1993. A pragmatic account of the definiteness effect in existencial sentences. *Journal of Pragmatics* 19:39–55.

Abeille, Anne and Yves Schabes, 1989. Parsing idioms with a lexicalized tree-adjoining grammar. In *Proceedings of the 4th International Meeting of the European Association for Computational Linguistics*, 1–9. ACL.

Abels, Klaus, 2008. Towards a restrictive theory of (remnant) movement. *Linguistic Variation Yearbook* 7:53–120.

Abels, Klaus and Ad Neeleman, 2009. Universal 20 without the LCA. In Josep Brucart, Anna Gavarró, and Jaume Solà, eds., *Merging Features: Computation, Interpretation and Acquisition*, 60–80. Oxford: Oxford University Press.

Abend, Omri, Tom Kwiatkowski, Nathaniel Smith, Sharon Goldwater, and Mark Steedman, 2017. Bootstrapping language acquisition. *Cognition* 164:116–143.

Ades, Anthony and Mark Steedman, 1982. On the order of words. *Linguistics and Philosophy* 4:517–558.

Adger, David, 2003. *Core Syntax: A Minimalist Approach*. Oxford: Oxford University Press.

Adger, David, 2007. Stress and phasal syntax. *Linguistic Analysis* 33:238–266.

Adger, David, 2013. *A Syntax of Substance*. MIT Press.

Aho, Alfred, 1968. Indexed grammars—an extension of context-free grammars. *Communications of the Association for Computing Machinery* 15:647–671.

Aho, Alfred, 1969. Nested stack automata. *Communications of the Association for Computing Machinery* 16:383–406.

Ajdukiewicz, Kazimierz, 1935. Die syntaktische Konnexität. In Storrs McCall, ed., *Polish Logic 1920–1939*, 207–231. Oxford: Oxford University Press. Trans. from *Studia Philosophica*, 1, 1-27.

Almaghout, Hala, 2012. *CCG-Augmented Hierarchical Phrase-Based Statistical Machine Translation*. Ph.D. thesis, Dublin City University.

Almaghout, Hala, Jie Jiang, and Andy Way, 2010. CCG augmented hierarchical phrase based machine-translation. In *Proceedings of the 7th International Workshop on Spoken Language Translation*, 211–218. Paris.

Almaghout, Hala, Jie Jiang, and Andy Way, 2011. CCG contextual labels in hierarchical phrase-based SMT. In *15th International Conference of the European Association for Machine Translation*, 281–288. EAMT.

Altmann, Gerry and Mark Steedman, 1988. Interaction with context during human sentence processing. *Cognition* 30:191–238.

Ambati, Bharat Ram, 2016. *Transition-based Combinatory Categorial Grammar Parsing for English and Hindi*. Ph.D. thesis, University of Edinburgh.

Ambati, Bharat Ram, Tejaswini Deoskar, Mark Johnson, and Mark Steedman, 2015. An incremental algorithm for transition-based CCG parsing. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 53–63. Denver.

Ambati, Bharat Ram, Tejaswini Deoskar, and Mark Steedman, 2013. Using CCG categories to improve Hindi dependency parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 604–609. Sofia.

Ambati, Bharat Ram, Tejaswini Deoskar, and Mark Steedman, 2016a. Hindi CCGbank: CCG treebank from the Hindi dependency treebank. *Language Resources and Evaluation* 51:1–34.

Ambati, Bharat Ram, Tejaswini Deoskar, and Mark Steedman, 2016b. Shift-reduce ccg parsing using neural network models. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 447–453. San Diego, CA.

Ambridge, Ben and Adele Goldberg, 2008. The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics* 19:357–389.

Ambridge, Ben, Julian Pine, Caroline Rowland, Franklin Chang, and Amy Bidgood, 2013. The retreat from overgeneralization in child language acquisition: Word learning, morphology, and verbc argument structure. *Wiley Interdisciplinary Reviews: Cognitive Science* 4:47–62.

Angluin, Dana, 1980. Inductive inference of formal languages from positive data. *Information and Control* 45:117–135.

Aoun, Joseph and Yen-hui Audrey Li, 1989. Scope and constituency. *Linguistic Inquiry* 20:141–172.

Artzi, Yoav, Dipanjan Das, and Slav Petrov, 2014. Learning compact lexicons for CCG semantic parsing. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1273–1283. Doha: ACL.

Artzi, Yoav, Kenton Lee, and Luke Zettlemoyer, 2015. Broad-coverage CCG semantic parsing with AMR. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1643–1648. Lisbon, Portugal: ACL.

Asher, Nicholas, 1993. *Referring to Abstract Objects in Discourse*, *Studies in Linguistics and Philosophy*, vol. 50. Dordrecht: Kluwer Academic Publishers.

Asudeh, Ash and Richard Crouch, 2002. Coordination and parallelism in glue semantics: Integrating discourse cohesion and the element constraint. In Miriam Butt and Tracy Holloway King, eds., *Proceedings of the LFG02 Conference*, 19–39. Stanford, CA: CSLI Publications.

Auli, Michael, 2011. *Combined Models for CCG Parsing*. Ph.D. thesis, University of Edinburgh.

Auli, Michael and Adam Lopez, 2011a. A comparison of loopy belief propagation and dual decomposition for integrated CCG supertagging and parsing. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 470–480. Portland, OR: ACL.

Auli, Michael and Adam Lopez, 2011b. Efficient CCG parsing: A* versus adaptive supertagging. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 1577–1585. Portland, OR: ACL.

Bach, Emmon, 1976. An extension of classical Transformational Grammar. In *Problems in Linguistic Metatheory: Proceedings of the 1976 Conference at Michigan State University*, 183–224. Lansing, MI: Michigan State University.

Bach, Emmon, 1979. Control in Montague Grammar. *Linguistic Inquiry* 10:513–531.

Bach, Emmon, 1980. In defense of passive. *Linguistics and Philosophy* 3:297–341.

Bach, Emmon and Barbara Partee, 1980. Anaphora and semantic structure. In *Papers from the 10th Regional Meeting of the Chicago Linguistic Society, Parasession on Pronouns and Anaphora*. Chicago: University of Chicago, Chicago Linguistic Society.

Bachrach, Asaf and Roni Katzir, 2009. Right-node raising and delayed spell-out. In Kleanthes Grohmann, ed., *InterPhases: Phase-Theoretic Investigations of Linguistic Interfaces*, 283–316. Oxford University Press.

Baker, C. Lee, 1979. Syntactic theory and the projection problem. *Linguistic Inquiry* 10:533–581.

Baker, Collin, Charles Fillmore, and John Lowe, 1998. The berkeley framenet project. In *Proceedings of the 17th international Conference on Computational Linguistics*, 86–90. Montreal, Quebec, Canada: ACL.

Baker, Mark, 2015. *Case: Its Principles and its Parameters*. Cambridge University Press.

Baldridge, Jason, 1998. *Local Scrambling and Syntactic Asymmetries in Tagalog*. Master's thesis, University of Pennsylvania.

Baldridge, Jason, 2002. *Lexically Specified Derivational Control in Combinatory Categorial Grammar*. Ph.D. thesis, University of Edinburgh.

Banarescu, Laura, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider, 2012. Abstract meaning representation (AMR) 1.0 specification.

Bangalore, Srinivas and Aravind Joshi, 1999. Supertagging: An approach to almost parsing. *Computational Linguistics* 25:237–265.

Bannard, Colin, Elena Lieven, and Michael Tomasello, 2009. Modeling children's early grammatical knowledge. *PNAS* 106:17284–17289.

Bar-Hillel, Yehoshua, 1953. A quasi-arithmetical notation for syntactic description. *Language* 29:47–58.

Barker, Chris, 2012. Quantificational binding does not require c-command. *Linguistic Inquiry* 43:614–633.

Barker, Chris and Chung-Chieh Shan, 2014. *Continuations and Natural Language*. Oxford University Press.

Barlow, Horace, 1961. Possible principles underlying the transformation of sensory messages. In Walter Rosenblith, ed., *Sensory Communication*, 217–234. New York: Wiley.

Baroni, Marco and Alessandro Lenci, 2010. Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics* 36:673–721.

Baroni, Marco and Roberto Zamparelli, 2010. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 1183–1193. Cambridge, MA: ACL.

Beal, Matthew, 2003. *Variational Algorithms for Approximate Bayesian Inference*. Ph.D. thesis, University of London.

Beavers, John, 2004. Type-inheritance Combinatory Categorial Grammar. In *Proceedings of the International Conference on Computational Linguistics*, 57–63. Geneva: ACL.

Beavers, John and Ivan Sag, 2004. Coordinate ellipsis and apparent non-constituent coordination. In *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar*, 48–69. Stanford: CSLI.

Bech, G, 1955. *Studien über das deutsche Verbum infinitivum, Band I.* Kopenhagen: Munksgaard.

Becker, Misha, 2005. Raising, control, and the subset principle. In *Proceedings of the 24th West Coast Conference on Formal Linguistics*, 52–60. Somerville, MA: Cascadilla Proceedings Project.

Beghelli, Filippo and Tim Stowell, 1997. Distributivity and negation: The syntax of *each* and *every*. In Anna Szabolcsi, ed., *Ways of Scope-Taking*, 71–107. Dordrecht: Kluwer.

Bekki, Daisuke, 2010. *Nihongo Bunpoo no Keesiki Riron: Katsuyoo taikee, Toogo koozoo, Imi goosee* [Formal theory of Japanese grammar: The system of conjugation, syntactic structure, and semantic composition]. Nihongo Kenkyuu Soosyo [Japanese Frontier Series] 24. Tokyo: Kurosio Publishers.

Benmamoun, Elabbas, 2000. Agreement asymmetries and the PF interface. In Jacqueline Lecarme, Jean Lowenstamm, and Ur Shlonsky, eds., *Research in Afroasiatic Grammar*, 23–40. Amsterdam: John Benjamins.

Berant, Jonathan, Noga Alon, Ido Dagan, and Jacob Goldberger, 2015. Efficient global learning of entailment graphs. *Computational Linguistics* 42:221–263.

Berman, Arlene, 1974. *Adjectives and Adjective Complement Constructions in English*. Ph.D. thesis, Department of Linguistics, Harvard University.

Bernardi, Raffaella, 2002. *Reasoning with Polarity in Categorial Type-Logic*. Ph.D. thesis, Universiteit Utrecht.

Berwick, Robert, 1985. *The Acquisition of Syntactic Knowledge*. Cambridge, MA: MIT Press.

Berwick, Robert and Noam Chomsky, 2016. *Why Only Us: Language and Evolution*. Cambridge, MA: MIT Press.

Berwick, Robert and Samuel Epstein, 1995a. Computational minimalism: The convergence of 'Minimalist' syntax and Categorial Grammar. In A. Nijholt, G. Scollo, and R. Steetkamp, eds., *Algebraic Methods in Language Processing 1995: Proceedings of the Twente Workshop on Language Technology 10, jointly held with the First Algebraic Methodology and Software Technology (AMAST) Workshop on Language Processing*. Enschede, The Netherlands: Faculty of Computer Science, Universiteit Twente.

Berwick, Robert and Samuel Epstein, 1995b. Merge: The Categorial Imperative. In A. Nijholt, G. Scollo, and R. Steetkamp, eds., *Proceedings of the First Algebraic Methodology and Software Technology (AMAST) Workshop on Language Processing*. Enschede, The Netherlands: Faculty of Computer Science, Universiteit Twente.

Birch, Alexandra, Miles Osborne, and Philipp Koehn, 2007. CCG supertags in factored translation models. In *Proceedings of the 2nd Workshop on Statistical Machine Translation*, 9–16. held in conjunction with ACL, Prague: ACL.

Boas, Hans and Ivan Sag, 2012. *Sign-based Construction Grammar*. Stanford, CA: CSLI Publications.

Bobaljik, Jonathan and Susi Wurmbrand, 2009. Case in GB/Minimalism. In Andreì Mal'chukov and Andrew Spencer, eds., *The Oxford Handbook of Case*, 44–58. Oxford Univerisity Press.

Boeckx, Cedric, 2008. *Bare Syntax*. Oxford: Oxford University Press.

Boeckx, Cedric, 2012. *Syntactic Islands*. Cambridge University Press.

Boeckx, Cedric, Norbert Hornstein, and Jairo Nunes, 2010. *Control as Movement*. Cambridge University Press.

Bok-Bennema, Reineke, 1991. *Case and Agreement in Inuit*. Berlin: Walter de Gruyter.

Bolhuis, Johan, Ian Tattersall, Noam Chomsky, and Robert Berwick, 2014. How could language have evolved? *PLoS Biol* 12:e1001934.

Bolinger, Dwight, 1958. A theory of pitch accent in English. *Word* 14:109–149. Reprinted as Bolinger 1965, 17–56.

Bolinger, Dwight, 1961. Contrastive accent and contrastive stress. *Language* 37:83–96. Reprinted as Bolinger 1965, 101-117.

Bolinger, Dwight, 1965. *Forms of English*. Cambridge, MA: Harvard University Press.

Bolinger, Dwight, 1972a. Accent is predictable (if you're a mind-reader). *Language* 48:633–644.

Bolinger, Dwight, ed., 1972b. *Intonation: Selected Readings*. Harmondsworth: Penguin Books.

Bolinger, Dwight, 1977. *Meaning and Form*. Longman.

Bordes, Antoine, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko, 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems*, 2787–2795.

Boroditsky, Lera, 2001. Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive Psychology* 43:1–22.

Borsley, Robert, Maggie Tallerman, and David Willis, 2007. *The Syntax of Welsh*. Cambridge: Cambridge University Press.

Bos, Johan and Katja Markert, 2005. Recognising textual entailment with logical inference. In *Proceedings of the 2005 Conference on Empirical Methods in Natural Language Processing (EMNLP 2005)*, 628–635.

Bošković, Željko, 2002. On multiple wh-fronting. *Linguistic inquiry* 33:351–383.

Bošković, Željko and Daiko Takahashi, 1998. Scrambling and Last Resort. *Linguistic inquiry* 29:347–366.

Bowerman, Melissa, 1973a. *Early Syntactic Development*. Cambridge University Press.

Bowerman, Melissa, 1973b. Structural relationships in children's utterances: Syntactic or semantic? In Timothy Moore, ed., *Cognitive Development and the Acquisition of Language*, 197–213. Academic Press.

Bowerman, Melissa, 1988. The 'no negative evidence' problem: How do children avoid constructing an overly general grammar? In John Hawkins, ed., *Explaining Language Universals*, 73–101. Basil Blackwell.

Bowerman, Melissa and Soonja Choi, 2001. Shaping meanings for language: Universal and language-specific in the acquisition of spatial semantic categories. In Melissa Bowerman and Stephen Levinson, eds., *Language Acquisition and Conceptual Development*, 475–511. Cambridge University Press.

Boxwell, Stephen and Chris Brew, 2010. A pilot arabic ccgbank. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, 1881–. Valletta: ELRA.

Bozşahin, Cem, 1998. Deriving predicate-argument structure for a free word order language. In *Proceedings of International Conference on Computational Linguistics*, 167–173. Montreal: ACL.

Bozşahin, Cem, 2002. The combinatory morphemic lexicon. *Computational Linguistics* 28:145–186.

Bozşahin, Cem, 2012. *Combinatory Linguistics*. Berlin: de Gruyter.

Braine, Martin, 1971. On two types of models of the internalization of grammars. In Dan Slobin, ed., *The Ontogenesis of Grammar*, 153–186. New York: Academic Press.

Braun, Bettina, 2006. Phonetics and phonology of thematic contrast in German. *Language and Speech* 49:451–493.

Bresnan, Joan, 1971. Sentence stress and syntactic transformations. *Language* 47:257–281.

Bresnan, Joan, 1977. Variables in the theory of transformations part i:bounded versus unbounded transformations. In Peter Culicover, Thomas Wasow, and Adrian Akmajian, eds., *Formal Syntax*, 157–196. New York: Academic Press.

Bresnan, Joan, 1978. A realistic Transformational Grammar. In Morris Halle, Joan Bresnan, and George Miller, eds., *Linguistic Structure and Psychological Reality*, 1–59. Cambridge, MA: MIT Press.

Bresnan, Joan, ed., 1982. *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press.

Bresnan, Joan, Ronald Kaplan, Stanley Peters, and Annie Zaenen, 1982. Cross-serial dependencies in Dutch. *Linguistic Inquiry* 13:613–636.

Briscoe, Ted, 2000. Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. *Language* 76:245–296.

Brody, Michael, 1995. *Lexico-Logical Form: A Radically Minimalist Theory*. Cambridge, MA: MIT Press.

Brown, Gillian, 1983. Prosodic structure and the given/new distinction. In Anne Cutler, D. Robert Ladd, and Gillian Brown, eds., *Prosody: Models and Measurements*, 67–77. Berlin: Springer-Verlag.

Brown, Peter, John Cocke, Stephen Della-Pietra, Vincent Della-Pietra, Fredrick Jelinek, John Lafferty, Robert Mercer, and Paul Roossin, 1990. A statistical approach to machine translation. *Computational Linguistics* 16:79–85.

Brown, Roger, 1973. *A First Language: the Early Stages*. Cambridge, MA: Harvard University Press.

Browning, Marguerite, 1987. *Null operator constructions*. Ph.D. thesis, Massachusetts Inst. of Technology Cambridge.

Browning, Marguerite, 1996. CP recursion and *that-t* effects. *Linguistic Inquiry* 237–255.

Browning, Robert, 1855. Andrea del sarto. In *Men and Women*. Chapman and Hall.

Bruening, Benjamin, 2014. Precede-and-command revisited. *Language* 90:342–388.

Büring, Daniel, 1997a. The great scope inversion conspiracy. *Linguistics and Philosophy* 20:175–194.

Büring, Daniel, 1997b. *The Meaning of Topic and Focus: The 59th Street Bridge Accent*. London: Routledge.

Büring, Daniel, 2005. *Binding Theory*. Cambridge: Cambridge University Press.

Büring, Daniel, 2016a. *Intonation and Meaning*. Oxford University Press.

Büring, Daniel, 2016b. Unalternative semantics. In *Proceedings of the 25th Conference on Semantics and Linguistic Theory (SALT)*, 550–575. LSA.

Burkart, Judith, Sarah Blaffer Hrdy, and Carel Van Schaik, 2009. Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology: Issues, News, and Reviews* 18:175–186.

Bush, Robert and Frederick Mosteller, 1955. *Stochastic Models for Learning*. New York: Wiley.

Butt, Miriam, 2006. *Theories of Case*. Cambridge University Press.

Butt, Miriam, 2009. Case in lexical-functional grammar. In Andrei Malchukov and Andrew Spencer, eds., *The Oxford Handbook of Case*, 59–71. Oxford University Press.

Butt, Miriam and Tracy Holloway King, 1991. Semantic case in Urdu. In *Papers from the General Session at the 27th Regional Meeting of the Chicago Linguistics Society*, 31–45.

Butt, Miriam and Tracy Holloway King, 2004. The status of case. In Veneeta Dayal and Anoop Mahajan, eds., *Clause Structure in South Asian languages*, 153–198. Springer.

Bybee, Joan, 2013. Usage-based theory and exemplar representations of constructions. In Thomas Hoffman and Graeme Trousdale, eds., *The Oxford Handbook of Construction Grammar*, 49–69. Oxford University Press.

Calhoun, Sasha, 2010. The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language* 86:1–42.

Calhoun, Sasha, 2012. The theme/rheme distinction: Accent type or relative prominence? *Journal of Phonetics* 40:329–349.

Calhoun, Sasha, 2015. The interaction of prosody and syntax in Samoan focus marking. *Lingua* 165:205–229.

Cann, Ronnie, Ruth Kempson, and Lutz Marten, 2005. *The Dynamics of Language*. Syntax and Semantics 35. San Diego: Elsevier Academic Press.

Carlson, Greg, 1977a. *Reference to Kinds in English*. Ph.D. thesis, University of Massachusetts, Amherst.

Carlson, Greg, 1977b. A unified analysis of the English bare plural. *Linguistics and Philosophy* 1:413–458.

Carlson, Greg, 1987. Same and different: Some consequences for syntax and semantics. *Linguistics and Philosophy* 10:531–565.

Carlson, Greg and Jeff Pelletier, eds., 1995. *The Generic Book*. Chicago: University of Chicago Press.

Carnap, Rudolf, 1952. Meaning postulates. *Philosophical Studies* 3:65–73. Reprinted as Carnap, 1956:222-229.

Carnap, Rudolf, ed., 1956. *Meaning and Necessity*. Chicago: University of Chicago Press, 2nd edn.

Carpenter, Bob, 1992. Categorial grammars, lexical rules, and the english predicative. In Robert Levine, ed., *Formal Grammar: Theory and Implementation*, 168–242. Oxford University Press.

Carpenter, Bob, 1997. *Type-Logical Semantics*. Cambridge, MA: MIT Press.

Casadio, Claudia, 2004. Quantifiers and scope in pregroup grammar. Ms., University G.D'Annunzio, Chieti.

Cattell, Ray, 1976. Constraints on movement rules. *Language* 18–50.

Çakıcı, Ruket, 2005. Automatic induction of a CCG grammar for Turkish. In *Proceedings of the Student Workshop, 43rd Annual Meeting of the ACL, Ann Arbor MI*, 73–78. ACL.

Çakıcı, Ruket, 2009. *Parser Models for a Highly Inflected Language*. Ph.D. thesis, University of Edinburgh.

Cha, Jeongwon and Geunbae Lee, 2000. Structural disambiguation of morpho-syntactic categorial parsing for Korean. In *Proceedings of the 18th International Conference on Computational Linguistics*, 1002–1006. Saarbrücken: ACL.

Chambers, Nathanael, Taylor Cassidy, Bill McDowell, and Steven Bethard, 2014. Dense event ordering with a multi-pass architecture. *Transactions of the Association of Computational Linguistics* 2:273–284.

Charniak, Eugene, 1997. Statistical parsing with a context-free grammar and word statistics. In *Proceedings of the 14th National Conference of the American Association for Artificial Intelligence, Providence, RI., July*, 598–603.

Charniak, Eugene, 2000. A maximum-entropy-inspired parser. In *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics*, 132–139. Seattle, WA.

Chaves, Rui, 2012. On the grammar of extraction and coordination. *Natural Language & Linguistic Theory* 30:465–512.

Chelba, Ciprian and Frederick Jelinek, 2000. Structured language modeling. *Computer Speech & Language* 14:283–332.

Chen, Danqi and Christopher Manning, 2014. A fast and accurate dependency parser using neural networks. In *Proceedings of the International Conference on Empirical Methods in Natural Language Processing*, 740–750.

Chen-Main, Joan, 2006. *On the Generation and Linearization of Multi-Dominance Structures*. Ph.D. thesis, John Hopkins University, Baltimore, MD.

Chierchia, Gennaro, 1984. *Topics in the Syntax and Semantics of Infinitives and Gerunds*. Ph.D. thesis, U. Mass. Amherst.

Chierchia, Gennaro, 1989. Anaphora and attitudes *de se*. In Renata Bartsch, Johan van Benthem, and Peter van Emde Boas, eds., *Semantics and Contextual Expression*, 1–32. De Gruyter Mouton.

Choe, Jae, 1987. LF movement and Pied Piping. *Linguistic Inquiry* 18:348–353.

Chomsky, Noam, 1955/1975. *The Logical Structure of Linguistic Theory*. Chicago: University of Chicago Press.

Chomsky, Noam, 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, Noam, 1964. *Current Issues in Linguistic Theory*. The Hague: Mouton.

Chomsky, Noam, 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam, 1973. Conditions on transformations. In Stephen Anderson and Paul Kiparsky, eds., *A Festschrift for Morris Halle*, 232–286. New York: Holt, Rinehart & Winston. Reprinted in Chomsky (1977a).

Chomsky, Noam, 1975. *Reflections on Language*. New York: Pantheon.

Chomsky, Noam, 1977a. *Essays on Form and Interpretation*. Amsterdam: North Holland.

Chomsky, Noam, 1977b. On *wh*-movement. In Peter Culicover, Thomas Wasow, and Adrian Akmajian, eds., *Formal Syntax*, 71–132. New York: Academic Press.

Chomsky, Noam, 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, Noam, 1982. *Concepts and Consequences of the Theory of Government and Binding*. Cambridge, MA: MIT Press.

Chomsky, Noam, 1986a. *Barriers*. Cambridge, MA: MIT Press.

Chomsky, Noam, 1986b. *Knowledge of Language*. New York: Praeger.

Chomsky, Noam, 1993. A minimalist program for linguistic theory. In Kenneth Hale and Samuel Jay Keyser, eds., *The View from Building 20*, 1–52. Cambridge, MA: MIT Press. Reprinted in Chomsky (1995b).

Chomsky, Noam, 1995a. Bare phrase structure. In Gert Webelhuth, ed., *Government and Binding Theory and the Minimalist Program*, 383–439. Oxford: Blackwell.

Chomsky, Noam, 1995b. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam, 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels, and Juan Uriagereka, eds., *Step by Step: Essays in Minimalist Syntax in Honor of Howard Lasnik*, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam, 2001. Derivation by phase. In Michael Kenstowicz, ed., *Ken Hale: a Life in Language*, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam, 2001/2004. Beyond explanatory adequacy. In Adriana Belletti, ed., *Structures and Beyond: The Cartography of Syntactic Structures*, 104–131. Oxford University Press.

Chomsky, Noam, 2005. Three factors in language design. *Linguistic Inquiry* 36:1–22.

Chomsky, Noam, 2007. Approaching UG from below. In Uli Sauerland and Hans-Martin Gärtner, eds., *Interfaces + recursion = language? Chomsky's Minimalism and the View from Syntax-Semantics*, 1–29. Berlin: de Gruyter.

Chomsky, Noam, 2008. On phases. In Robert Freidin, Carlos Otero, and Maria Luisa Zubizarreta, eds., *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, 133–166. Cambridge, MA: MIT Press.

Chomsky, Noam, 2013. Problems of projection. *Lingua* 130:33–49.

Chomsky, Noam and Morris Halle, 1968. *The Sound Pattern of English*. New York: Harper and Row.

Chomsky, Noam, Morris Halle, and Fred Lukoff, 1956. On accent and juncture in English. *For Roman Jakobson* 65–80.

Chomsky, Noam and Marcel-Paul Schützenberger, 1963. The algebraic theory of context-free languages. In Paul Braffort and David Hirschberg, eds., *Computer Programming and Formal Systems*, 118–161. Amsterdam: North-Holland.

Christodoulopoulos, Christos and Mark Steedman, 2015. A massively parallel corpus: The Bible in 100 languages. *Language Resources and Evaluation* 49:to appear.

Chung, Sandra, 2013. Syntactic identity in sluicing: How much and why. *Linguistic Inquiry* 44:1–44.

Chung, Sandra, William Ladusaw, and James McCloskey, 1995. Sluicing and logical form. *Natural Language Semantics* 3:239–282.

Chung, Sandra and James McCloskey, 1983. On the interpretation of certain island facts in GPSG. *Linguistic Inquiry* 14(4):704–713.

Church, Kenneth and Patrick Hanks, 1989. Word association norms, mutual information, and lexicography. In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, 76–83. ACL.

Cinque, Guglielmo, 1990. *Types of $\bar{A}$-dependencies*. MIT Press, Cambridge, MA.

Cinque, Guglielmo, 1993. A null theory of phrase and compound stress. *Linguistic Inquiry* 24:239–297.

Cinque, Guglielmo, 2005. Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry* 36:315–332.

Cinque, Guglielmo and Luigi Rizzi, 2008. The cartography of syntactic structures. *Studies in Linguistics:CISCL Working Papers on Language and Cognition* 2:43–59.

Citko, Barbara, 2011. Multidominance. In Cedric Boeckx, ed., *The Oxford Handbook of Linguistic Minimalism*, 119–142. OUP Oxford.

Citko, Barbara, 2017. Right node raising. In *The Wiley Blackwell Companion to Syntax*, 1–33. Wiley, 2 edn.

Citko, Barbara and Martina Gračanin-Yuksek, 2020. *Merge: Binarity in (Multidominant) Syntax*. Cambridge, MA: MIT Press.

Clark, Alexander, 2004. Grammatical inference and first language acquisition. In *CoLing Workshop on Psycho-computational Models of Human Language Acquisition*.

Clark, Eve, 1973. What's in a word? on the child's acquisition of semantics in his first language. In Timothy Moore, ed., *Cognitive Development and the Acquisition of Language*, 65–110. Academic Press.

Clark, Herbert, 1996. *Using Language*. Cambridge: Cambridge University Press.

Clark, Herbert and Catherine Marshall, 1981. Definite reference and mutual knowledge. In Aravind Joshi, Bonnie Webber, and Ivan Sag, eds., *Elements of Discourse Understanding*, 10–63. Cambridge: Cambridge University Press.

Clark, Robin and Ian Roberts, 1993. A computational model of language learnability and language change. *Linguistic Inquiry* 24:299–345.

Clark, Stephen and James R. Curran, 2004. Parsing the WSJ using CCG and log-linear models. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, 104–111. Barcelona: ACL.

Clark, Stephen and James R. Curran, 2007a. Formalism-independent parser evaluation with CCG and DepBank. In *Proceedings of the Annual Meeting of the ACL*, 248–255. Prague: ACL.

Clark, Stephen and James R. Curran, 2007b. Wide-coverage efficient statistical parsing with CCG and log-linear models. *Computational Linguistics* 33:493–552.

Cocke, John and Jacob T. Schwartz, 1970. *Programming Languages and Their Compilers*. New York: New York University, Courant Institute.

Cohen, Philip, 1978. *On Knowing What to Say: Planning Speech Acts*. Ph.D. thesis, University of Toronto.

Collins, Michael, 1997. Three generative lexicalized models for statistical parsing. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, 16–23. Madrid: ACL.

Collins, Michael, 1999. *Head-Driven Statistical Models for Natural Language Parsing*. Ph.D. thesis, University of Pennsylvania.

Collins, Peter, 2005. Exclamative clauses in English. *Word* 56:1–17.

Collobert, Ronan, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa, 2011. Natural language processing (almost) from scratch. *The Journal of Machine Learning Research* 12:2493–2537.

Comrie, Bernard, 1989. *Language Universals and Linguistic Typology*. Oxford: Blackwell, 2nd edn.

Cooper, Robin, 1983. *Quantification and Syntactic Theory*. Dordrecht: Reidel.

Cormack, Annabel and Neil Smith, 2004. Backward control in Korean and Japanese. *Working Papers in Linguistics* 16:57–83.

Cowart, Wayne, 1997. *Experimental Syntax*. Sage.

Crain, Stephen and Mark Steedman, 1985. On not being led up the garden path: The use of context by the psychological parser. In David Dowty, Lauri Kartunnen, and Arnold Zwicky, eds., *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*, 320–358. Cambridge: Cambridge University Press.

Crain, Stephen and Rosalind Thornton, 1998. *Investigations in Universal Grammar*. Cambridge, MA: MIT Press.

Culicover, Peter and Ray Jackendoff, 2005. *Simpler Syntax*. Oxford: Oxford University Press.

Culicover, Peter and Michael Rochemont, 1990. Extraposition and the complement principle. *Linguistic Inquiry* 21:23–47.

Culicover, Peter W, 1993. Evidence against ECP accounts of the *that-t* effect. *Linguistic Inquiry* 557–561.

Curry, Haskell and Robert Feys, 1958. *Combinatory Logic: Vol. I*. Amsterdam: North Holland.

Dalrymple, Mary, 1993. *The Syntax of Anaphoric Binding*. CSLI Monograph 36. Stanford, CA: CSLI Publications.

Dang, Hoa Trang, Joseph Rosenzweig, and Martha Palmer, 1997. Associating semantic components with levin classes. In *Proceedings of Interlinga Workshop, MTSUMMIT97, San Diego,CA, October*, 11–18.

Darwin, Charles, 1871. *The Descent of Man*. John Murray.

Davidson, Donald and Gilbert Harman, eds., 1972. *Semantics of Natural Language*. Dordrecht: Reidel.

Davis, Anthony, 1992. Missing heads and empty subjects: Underspecification in Hausa VPs. In *Papers from the 28th Regional Meeting of the Chicago Linguistic Society*, 94–108. CLS.

De Kuthy, Kordula and Detmar Meurers, 2001. On partial constituent fronting in German. *The Journal of Comparative Germanic Linguistics* 3:143–205.

De Vries, Mark, 2009. On multidominance and linearization. *Biolinguistics* 3:344–403.

Deacon, Terence, 1997. *The Symbolic Species*. New York: Norton.

Deal, Amy, 2009. The origin and content of expletives: Evidence from “selection”. *Syntax* 12:285–323.

Demberg, Vera, Frank Keller, and Alexander Koller, 2013. Incremental, predictive parsing with psycholinguistically motivated tree-adjoining grammar. *Computational Linguistics* 39:1025–1066.

den Dikken, Marcel, 1994. Minimalist verb (projection) raising. *Groninger Arbeiten zur germanistischen Linguistik* 37:71–88.

Den Dikken, Marcel, 1995. *Particles: On the Syntax of Verb-particle, Triadic, and Causative Constructions*. Oxford University Press.

Dench, Alan, 1987/1995. *Martuthunira: A Language of the Pilbara Region of Western Australia*. Ph.D. thesis, Australian National University, Canberra.

Di Sciullo, Anna Maria and Cedric Boeckx, 2011. *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*. Oxford University Press.

Dik, Simon, 1968. *Coordination: Its Implications for the Theory of General Linguistics*. Amsterdam: North Holland.

Diogo, A de A and Masaya Yoshida, 2007. A problem for the preposition stranding generalization. *Linguistic Inquiry* 38:349–362.

Dixon, R.M.W., 1972. *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.

Doron, Edit, 2000. VSO and left-conjunct agreement: Biblical Hebrew vs. Modern Hebrew. In Andrew Carnie and Eithne Guilfoyle, eds., *The Syntax of Verb Initial Languages*, 79–95. Oxford: Oxford University Press.

Dorr, Bonnie, Rebecca Passonneau, David Farwell, Rebecca Green, Nizar Habash, Stephen Helmreich, Edward Hovy, Lori Levin, Keith Miller, Teruko Mitamura, Owen Rambow, and Advaith Siddharthan, 2010. Interlingual annotation of parallel text corpora: A new framework for annotation and evaluation. *Natural Language Engineering* 16:197–243.

Dougherty, Ray, 1970. A grammar of coordinate conjoined structures, i. *Language* 46:850–898.

Dowty, David, 1979a. Dative movement and Thomason's extensions of Montague Grammar. In Steven Davis and Marianne Mithun, eds., *Linguistics, Philosophy, and Montague Grammar*, 153–222. Austin: University of Texas Press.

Dowty, David, 1979b. *Word Meaning in Montague Grammar*. Dordrecht: Reidel, 1st edn.

Dowty, David, 1981. Quantification and the lexicon: A reply to Fodor and Fodor. In Michael Moortgat, Harry van de Hulst, and Teun Hoekstra, eds., *The Scope of Lexical Rules*, 79–106. Foris, Dordrecht.

Dowty, David, 1985/1988. Type-raising, functional composition, and nonconstituent coordination. In Richard Oehrle, Emmon Bach, and Deirdre Wheeler, eds., *Categorial Grammars and Natural Language Structures*, 153–198. Dordrecht: Reidel.

Dowty, David, 1991a. Thematic proto-roles and argument selection. *Language* 547–619.

Dowty, David, 1991b. *Word Meaning in Montague Grammar*. Dordrecht: Kluwer, 2nd edn. (1st edition, Reidel 1979).

Dowty, David, 2003. The dual analysis of adjuncts/complements in categorial grammar. In Ewald Lang, Claudia Maienborn, and Cathry Fabricius-Hansen, eds., *Modifying Adjuncts*, 33–66. Berlin: Mouton de Gruyter.

Dryer, Matthew, 2009. On the order of demonstrative, numeral, adjective, and noun: An alternative to Cinque. In *Conference on Theoretical Approaches to Disharmonic Word Orders*. Newcastle. Handout.

Dun, Yantao, Weiwei Sun, and Xiaojun Wan, 2015. A data-driven, factorization parser for ccg dependency structures. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 1545–1555. Beijing: ACL.

Dunbar, Robin, 1992. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution* 22:469–493.

Dyer, Chris, Miguel Ballesteros, Wang Ling, Austin Matthews, and Noah Smith, 2015. Transition-based dependency parsing with stack long short-term memory. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, 334–343.

Dyer, Chris, Adhiguna Kuncoro, Miguel Ballesteros, and Noah Smith, 2016. Recurrent neural network grammars. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 199–209. ACL.

Emonds, Joseph, 1976. *A Transformational Approach to English Syntax*. New York: Academic Press.

Emory, Nathan and Nicola Clayton, 2004. The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science* 306:1903–1907.

Engdahl, Elisabet, 1983. Parasitic gaps. *Linguistics and Philosophy* 6:5–34.

Epstein, Samuel, 1999. Un-principled syntax: The derivation of syntactic relations. In Samuel Epstein and Norbert Horstein, eds., *Working Minimalism*, 317–345. Cambridge, MA: MIT Press.

Epstein, Samuel, Erich Groat, Ruriko Kawashima, and Hisatsugu Kitahara, 1998. *A Derivational Approach to Syntactic Relations*. Oxford: Oxford University Press.

Epstein, Samuel and Daniel Seely, 2006. *Derivations In Minimalism*. Cambridge: Cambridge University Press.

Erteschik-Shir, Nomi, 2007. *Information Structure: The Syntax-Discourse Interface*. Oxford: Oxford University Press.

Espinosa, Dominic, Rajakrishnan Rajkumar, Michael White, and Shoshana Berleant, 2010. Further meta-evaluation of broad-coverage surface realization. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 564–574. ACL.

Espinosa, Dominic, Michael White, and Dennis Mehay, 2008. Hypertagging: Supertagging for surface realization with CCG. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, 183–191. Columbus, OH: ACL.

Etzioni, Oren, Michele Banko, and Michael Cafarella, 2007. Machine reading. In *Proceedings of AAAI Spring Symposium on Machine Reading*. Menlo Park, CA: AAAI Press.

Evers, Arnold, 1975. *The Transformational Cycle in Dutch and German*. Ph.D. thesis, University of Utrecht.

Faltz, Leonard, 2000. A semantic basis for Navajo syntactic typology. In Theodore Fernald and Paul Platero, eds., *The Athabaskan Languages: Perspectives on a Native American Language Family*, 28–50. New York: Oxford University Press.

Farkas, Donka, 1997. Towards a semantic typology of noun phrases. In *Empirical Issues in Formal Syntax and Semantics: Selected Papers from the Duexième Colloque de Syntaxe et Sémantique de Paris, October 1997*. The Hague: Thesus.

Fellbaum, Christiane, ed., 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Felser, Claudia, 2004. *Wh*-copying, phases, and successive cyclicity. *Lingua* 114:543–574.

Fernald, Anne, 1993. Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development* 64:657–667.

Fernald, Anne, T. Taeschner, J. Dunn, M. Papousek, B. Boysson-Bardies, and I. Fukui, 1989. A cross-language study of prosodic modifications in mothers' and fathers' speech to infants. *Journal of Child Language* 16:477–501.

Fikes, Richard and Nils Nilsson, 1971. STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence* 2:189–208.

Fillmore, Charles, 1968. The case for case. In Emmon Bach and Robert Harms, eds., *Universals in Linguistic Theory*, 1–88. New York: Henry Holt.

Fisher, Cynthia, Geoffrey Hall, Susan Rakowitz, and Lila Gleitman, 1994. When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua* 92:333–375.

Fisher, Cynthia and Hisayo Tokura, 1996. Prosody in speech to infants: Direct and indirect acoustic cues to syntactic structure. In James Morgan and Katherine Demuth, eds., *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, 343–363. Erlbaum.

Fitch, Tecumseh, 2014. Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews* 11:329–364.

Fitch, Tecumseh and David Reby, 2001. The descended larynx is not uniquely human. *Proceedings of the Royal Society of London B: Biological Sciences* 268:1669–1675.

Fodor, Janet Dean, 1998. Unambiguous triggers. *Linguistic Inquiry* 29:1–36.

Fodor, Janet Dean, 2009. Syntax acquisition: An evaluation measure after all? In Massimo Piatelli-Palmarini, Juan Uriagereka, and P. Salaburu, eds., *Of Minds and Language: A Dialogue with Noam Chomsky in the Basque Country*, 256–277. Oxford: Oxford University Press.

Fodor, Janet Dean and William Sakas, 2005. The subset principle in syntax: Costs of compliance. *Journal of Linguistics* 41:513–569.

Fodor, Janet Dean, William Sakas, and Arthur Hoskey, 2007. Implementing the subset principle in syntax acquisition: Lattice-based models. In *Proceedings of the European Cognitive Science Conference*. Delphi.

Fodor, Jerry, 1966. How to learn to talk: Some simple ways. In Frank Smith and George Miller, eds., *The Genesis of Language*, 105–122. Cambridge: MIT Press.

Fodor, Jerry, 1975. *The Language of Thought*. Cambridge, MA: Harvard.

Fodor, Jerry, Thomas Bever, and Merrill Garrett, 1974. *The Psychology of Language*. New York: McGraw-Hill.

Fodor, Jerry and Janet Dean Fodor, 1980. Functional structure, quantifiers, and meaning postulates. *Linguistic Inquiry* 11:759–770.

Fong, Sandiway, 1991. *Computational Properties of Principle-Based Grammatical Theories*. Ph.D. thesis, Massachusetts Institute of Technology.

Fox, Danny, 1999. Reconstruction, binding theory, and the interpretation of chains. *Linguistic Inquiry* 30:157–196.

Fox, Danny, 2002. Antecedent-contained deletion and the copy theory of movement. *Linguistic Inquiry* 33:63–96.

Frampton, John and Sam Gutmann, 2002. Crash-proof syntax. In Samuel Epstein and Daniel Seely, eds., *Derivation and Explanation in the Minimalist Program*, 90–105. Blackwell Publishing.

Frank, Robert, 2002. *Phrase Structure Composition and Syntactic Dependencies*. Cambridge, MA: MIT Press.

Frank, Robert, 2006. Phase theory and tree adjoining grammar. *Lingua* 116:145–202.

Frishberg, Dave, 1962. Peel me a grape. Swiftwater Music.

Frost, Robert, 1914. Mending wall. In *North of Boston*. London: David Nutt.

Fukui, Naoki, 1986. *A Theory of Category Projection and its Applications*. Ph.D. thesis, Massachusetts Institute of Technology.

Fukui, Naoki, 1995. *Theory of Projection in Syntax*. CSLI Publications.

Furth, Hans, 1961. The influence of language on the development of concept formation in deaf children. *Journal of Abnormal and Social Psychology* 63:386–389.

Fyleman, Rose, 1917. There are fairies at the bottom of our garden. *Punch* .

Garrette, Dan, Chris Dyer, Jason Baldridge, and Noah Smith, 2015. Weakly-supervised grammar-informed bayesian ccg parser learning. In *Proceedings of the 29th National Conference of the Association for the Advancement of Artificial Intelligence*, XXX.

Garrette, Dan, Katrin Erk, and Raymond Mooney, 2014. A formal approach to linking logical form and vector-space lexical semantics. In Harry Bunt, Johan Bos, and Stephen Pulman, eds., *Computing Meaning*, 27–48. Springer.

Gazdar, Gerald, 1980. A cross-categorial semantics for coordination. *Linguistics and Philosophy* 3:407–409.

Gazdar, Gerald, 1981. Unbounded dependencies and coordinate structure. *Linguistic Inquiry* 12:155–184.

Gazdar, Gerald, 1988. Applicability of indexed grammars to natural languages. In Uwe Reyle and Christian Rohrer, eds., *Natural Language Parsing and Linguistic Theories*, 69–94. Dordrecht: Reidel.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum, and Ivan Sag, 1984. Foot features and parasitic gaps. In *Sentential Complementation*, 83–94. Dordrecht: Foris.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum, and Ivan Sag, 1985. *Generalized Phrase Structure Grammar*. Oxford: Blackwell.

Geach, Peter, 1970. A program for syntax. *Synthèse* 22:3–17. Reprinted as Davidson and Harman 1972:483–497.

Geib, Christopher and Mark Steedman, 2007. On natural language processing and plan recognition. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, 1612–1617.

Gerken, LouAnn, 1996. Prosodic structure in young children's language production. *Langauge* 72:683–712.

Gerken, LouAnn, Peter Jusczyk, and Denise Mandel, 1994. When prosody fails to cue syntactic structure. *Cognition* 51:237–265.

Gibson, Edward and Kenneth Wexler, 1994. Triggers. *Linguistic Inquiry* 25:355–407.

Gibson, James, 1966. *The Senses Considered as Perceptual Systems*. Boston, MA: Houghton-Mifflin Co.

Gibson, James, 1977. The theory of affordances. In Robert Shaw and John Bransford, eds., *Perceiving, Acting, and Knowing*, 67–82. Lawrence Erlbaum.

Gillies, William, 1993. Scottish gaelic. In Martin Ball, ed., *The Celtic Languages*, 145–227. Routledge.

Giuliani, Manuel and Alois Knoll, 2007. Integrating multimodal cues using grammar based models. In Constantine Stephanidis, ed., *Proceedings of the 10th International Conference on Multimodal Interfaces (ICMI)*, *Lecture Notes in Computer Science*, vol. 4555, 165–172. Berlin: Springer.

Giuliani, Manuel and Alois Knoll, 2008. Multiml: A general purpose representation language for multimodal human utterances. In *Proceedings of Human Computer Interaction International, vol. II*, 165–172. Beijing: ACM.

Glass, Jesse and Alexander Yates, 2014. Empirically-motivated generalizations of CCG semantic parsing learning algorithms. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, 348–357. Gothenburg.

Gleitman, Lila, 1990. The structural sources of verb meanings. *Language Acquisition* 1:1–55.

Gleitman, Lila, Kimberly Cassidy, Rebecca Nappa, Anna Papafragou, and John C. Trueswell, 2005. Hard words. *Language Learning and Development* 1:23–64.

Goddard, Cliff, 2010. The natural semantic metalanguage approach. In Bernd Heine and Heiko Narro, eds., *The Oxford Handbook of Linguistic Analysis*, 459–484. Oxford: Oxford University Press.

Goldberg, Adele, 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: Chicago University Press.

Goldberg, Adele, 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Goldberg, Adele, 2013a. Argument structure constructions versus lexical rules or derivational verb templates. *Mind & Language* 28:435–465.

Goldberg, Adele, 2013b. Constructionist approaches. In Thomas Hoffmann and Graeme Trousdale, eds., *The Oxford Handbook of Construction Grammar*, 15–31. New York: Oxford University Press.

Goldberg, Adele, 2019. Explain me this. In *Explain Me This*. Princeton University Press.

Goldsmith, John, 1985. A principled exception to the Coordinate Structure Constraint. In *Proceedings of the 21st Regional Meeting of the Chicago Linguistic Society*. Chicago: CLS.

Goldwater, Sharon, Thomas Griffiths, and Mark Johnson, 2009. A bayesian framework for word segmentation: Exploring the effects of context. *Cognition* 112:21–54.

Goodall, Grant, 1983. A three-dimensional analysis of coordination. In *Proceedings of the 19th meeting of the Chicago Linguistic Society*. Chicago: CLS.

Goodall, Grant, 1987. *Parallel Structures in Syntax*. Cambridge: Cambridge University Press.

Gopnik, Alison, 2001. Theories, language, and culture: Whorf without wincing. In *Language Acquisition and Conceptual Development*, 45–69. Cambridge University Press.

Gorniak, Peter and Deb Roy, 2007. Situated language understanding as filtering perceived affordances. *Cognitive Science* 31. 197-231.

Greenberg, Joseph, 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph Greenberg, ed., *Universals of Language*, 73–113. Cambridge, MA: MIT Press.

Grefenstette, Edward and Mehrnoosh Sadrzadeh, 2011. Experimental support for a categorical compositional distributional model of meaning. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 1394–1404. Edinburgh: ACL.

Grice, Paul, 1967/1989. *Studies in the Way of Words*. Harvard University Press. The William James Lectures, 1967.

Grimshaw, Jane, 1990. *Argument Structure*. Cambridge, MA: MIT Press.

Grohmann, Kleanthes, John Drury, and Juan Carlos Castillo, 2000. No more epp. In *Proceedings of 19th WCCFL*, 139–152.

Grover, Claire, 1995. *Rethinking some Empty Categories: Missing Objects and Parasitic Gaps in HPSG*. Ph.D. thesis, University of Essex.

Grycner, Adam and Gerhard Weikum, 2014. Harpy: Hypernyms and alignment of relational paraphrases. In *Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers*, 2195–2204. ACL.

Grycner, Adam, Gerhard Weikum, Jay Pujara, James Foulds, and Lise Getoor, 2015. Relly: Inferring hypernym relationships between relational phrases. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 971–981. ACL.

Gundel, Jeanette, 1974. *The Role of Topic and Comment in Linguistic Theory*. Ph.D. thesis, University of Texas, Austin.

Gundel, Jeanette and Torsten Fretheim, 2001. Topic and focus. In Laurence Horn and Gregory Ward, eds., *Handbook of Pragmatic Theory*. Oxford: Blackwell.

Gunlogson, Christine, 2001. *True to Form: Rising and Falling Declaratives in English*. Ph.D. thesis, University of California at Santa Cruz. In preparation.

Gunlogson, Christine, 2002. Declarative questions. In Brendan Jackson, ed., *Proceedings of* Semantics and Linguistics Theory XII, 144–163. Ithaca, NY: CLC Publications.

Gussenhoven, Carlos, 1983. *On the Grammar and Semantics of Sentence Accent*. Dordrecht: Foris.

Guu, Kelvin, John Miller, and Percy Liang, 2015. Traversing knowledge graphs in vector space. In *Proceedings of the Conference on Empirical Methods in Natural Language Processsing*, 318–327.

Haegeman, Liliane, 1992. *Theory and Description in Generative Syntax: A Case Study in West Flemish*. Cambridge: Cambridge University Press.

Haegeman, Liliane, 1995. *The Syntax of Negation*. Cambridge University Press.

Haegeman, Liliane and Henk van Riemsdijk, 1986. Verb Projection Raising, scope, and the typology of rules affecting verbs. *Linguistic Inquiry* 17:417–466.

Haider, Hubert, 2003. V-clustering and clause union: Causes and effects. In Pieter Seuren and Gerhard Kempen, eds., *Verb Constructions in German and Dutch*, 91–126. John Benjamins.

Hale, Kenneth, 1983. Warlpiri and the grammar of non-configurational languages. *Natural Language & Linguistic Theory* 1:5–47.

Hale, Kenneth and Samuel Jay Keyser, 2002. *Prolegomenon to a Theory of Argument Structure*. Linguistic Inquiry Monograph 39. Cambridge, MA: MIT Press.

Halliday, Michael, 1963. The tones of English. *Archivum Linguisticum* 15:1.

Halliday, Michael, 1967a. *Intonation and Grammar in British English*. The Hague: Mouton.

Halliday, Michael, 1967b. Notes on transitivity and theme in English, part ii. *Journal of Linguistics* 3:199–244.

Hankamer, Jorge and Ivan Sag, 1976. Deep and surface anaphora. *Linguistic Inquiry* 7:391–428.

Harkema, Hendrik, 2001. *Parsing Minimalist Languages*. Ph.D. thesis, University of California Los Angeles.

Hart, Johan't, René Collier, and Antonie Cohen, 1990. *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody*. Cambridge University Press.

Hasegawa, Nobuko, 1985. On the so-called "zero pronoun" in Japanese. *The Linguistic Review* 4:289–342.

Haspelmath, Martin, 2004. Coordinating constructions: An overview. In Martin Haspelmath, ed., *Coordinating Constructions*, 3–39. John Benjamins.

Haspelmath, Martin, 2008. Parametric versus functional explanations of syntactic universals. In Theresa Biberauer, ed., *The Limits of Syntactic Variation*, 75–107. Amsterdam: John Benjamins.

Hassan, Hany, Khalil Sima'an, and Andy Way, 2008. A syntactic language model based on incremental CCG parsing. In *Spoken Language Technology Workshop (SLT)*, 205–208. IEEE.

Hassan, Hany, Khalil Sima'an, and Andy Way, 2009. A syntactified direct translation model with linear-time decoding. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 1182–1191. Singapore: ACL.

Haug, Dag, 2017. Backward control in Ancient Greek and Latin participial adjuncts. *Natural Language & Linguistic Theory* 35:99–159.

Hauser, Marc, Noam Chomsky, and Tecumseh Fitch, 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298:1569–1579.

Hawkins, John, 1983. *Word Order Universals*. New York: Academic.

Hayes, J. Milton, 1911. The green eye of the yellow god. Reynolds &Co.

Hays, David, 1964. Dependency theory: a formalism and some observations. *Language* 40:511–525.

Hays, David, 1967. *Introduction to Computational Linguistics*. New York: American Elsevier.

Heim, Irene, 1990. E-type pronouns and donkey anaphora. *Linguistics and Philosophy* 13:137–177.

Heim, Irene and Angelika Kratzer, 1998. *Semantics in Generative Grammar*. Oxford: Blackwell.

Henderson, James, 2003. Neural network probability estimation for broad coverage parsing. In *Proceedings of the 10th conference of the European Chapter of the Association for Computational Linguistics-Volume 1*, 131–138. ACL.

Henderson, James, Paola Merlo, Gabriele Musillo, and Ivan Titov, 2008. A latent variable model of synchronous parsing for syntactic and semantic dependencies. In *Proceedings of the 12th Conference on Computational Natural Language Learning*, 178–182. ACL.

Hendriks, Petra, 2004. *Either, both* and *neither* in coordinate structures. In Alice ter Meulen and Werner Abraham, eds., *The Composition of Meaning: From Lexeme to Discourse*, 115–138. Amsterdam: John Benjamins.

Hepple, Mark, 1990. *The Grammar and Processing of Order and Dependency: A Categorial Approach*. Ph.D. thesis, University of Edinburgh.

Hermann, Karl and Phil Blunsom, 2013. The role of syntax in vector space models of compositional semantics. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 894–904. Sofia, Bulgaria: ACL.

Heycock, Caroline and Anthony Kroch, 1994. Verb movement and coordination in a dynamic theory of licensing. *The Linguistic Review* 11:257–283.

Hicks, Glyn, 2009. Tough-constructions and their derivation. *Linguistic Inquiry* 40:535–566.

Higashikawa, Masahiko and Fred Minifie, 1999. Acoustical-perceptual correlates of "whisper pitch" in synthetically generated vowels. *Journal of Speech, Language, and Hearing Research* 42:583–591.

Higginbotham, James, 1983. Logical form, binding, and nominals. *Linguistic inquiry* 14:395–420.

Higgins, F. Roger, 1973. *The Pseudocleft Construction in English*. Ph.D. thesis, MIT, Cambridge, MA.

Hilliard, Austin and Stephanie White, 2009. Possible precursors of syntactic components in other species. In Derek Bickerton and Eörs Szathmáry, eds., *Biological Foundations and Origin of Syntax*, 161–184. Cambridge, MA: MIT Press.

Hinton, Geoffrey, ed., 1990. *Connectionist Symbol Processing*. Cambridge, MA: MIT Press/Elsevier. Reprint of *Artificial Intelligence, 46,1-2*.

Hirsch, Aron, 2017. A case for conjunction reduction. *submitted* .

Hirsch, Aron and Michael Wagner, 2015. Right node raising, scope, and plurality. In *Proceedings of the 20th Amsterdam Colloquium*, 187–196.

Hirschberg, Julia and Janet Pierrehumbert, 1986. Intonational structuring of discourse. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, 136–144. New York: ACL.

Hobbs, Jerry, 1990. The pierrehumbert-hirschberg theory of intonational meaning made simple: Comments on pierrehumbert and hirschberg. In Philip Cohen, Jerry Morgan, and Martha Pollack, eds., *Intentions in Communication*, 313–323. Cambridge, MA: MIT Press.

Hochreiter, Sepp and Jürgen Schmidhuber, 1997. Long short-term memory. *Neural Computation* 9:1735–1780.

Hockenmaier, Julia, 2003a. *Data and Models for Statistical Parsing with CCG*. Ph.D. thesis, School of Informatics, University of Edinburgh.

Hockenmaier, Julia, 2003b. Parsing with generative models of predicate-argument structure. In *Proceedings of the 41st Meeting of the Association for Computational Linguistics, Sapporo*, 359–366. San Francisco: Morgan-Kaufmann.

Hockenmaier, Julia, 2006. Creating a CCGbank and a wide-coverage CCG lexicon for German. In *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics*, 505–512. Sydney: ACL.

Hockenmaier, Julia and Mark Steedman, 2002a. Acquiring compact lexicalized grammars from a cleaner treebank. In *Proceedings of the Third International Conference on Language Resources and Evaluation*, 1974–1981. Las Palmas, Spain.

Hockenmaier, Julia and Mark Steedman, 2002b. Generative models for statistical parsing with Combinatory Categorial Grammar. In *Proceedings of the 40th Meeting of the Association for Computational Linguistics*, 335–342. Philadelphia.

Hockenmaier, Julia and Mark Steedman, 2007. CCGbank: a corpus of CCG derivations and dependency structures extracted from the Penn Treebank. *Computational Linguistics* 33:355–396.

Hoffman, Beryl, 1995. *Computational Analysis of the Syntax and Interpretation of "Free" Word-Order in Turkish*. Ph.D. thesis, University of Pennsylvania. Publ. as IRCS Report 95–17. Philadelphia: University of Pennsylvania.

Hoffman, Matthew, David Blei, and Francis Bach, 2010. Online learning for latent dirichlet allocation. *Advances in Neural Information Processing Systems* 23:856–864.

Honnibal, Matthew, James Curran, and Johan Bos, 2010a. Rebanking CCGbank for improved NP interpretation. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, 207–215. Uppsala, Sweden: ACL.

Honnibal, Matthew, Jonathan Kummerfeld, and James Curran, 2010b. Morphological analysis can improve a CCG parser for English. In *Proceedings of the International Conference on Computational Linguistics*, 445–453. Beijing: ACL.

Horn, Laurence, 1990. Showdown at truth-value gap: Burton-Roberts on presupposition. *Journal of Linguistics* 26(2):483–503.

Horning, James, 1969. *A Study of Grammatical Inference*. Ph.D. thesis, Stanford.

Hornstein, Norbert, 1995. *Logical Form: From GB to Minimalism*. Oxford: Blackwell.

Hornstein, Norbert, 1999a. Minimalism and quantifier raising. In Samuel David Epstein and Norbert Hornstein, eds., *Working Minimalism*, 46–75. MIT Press.

Hornstein, Norbert, 1999b. Movement and control. *Linguistic Inquiry* 30:69–96.

Hornstein, Norbert, 2001. *Move! A Minimalist Theory of Construal*. Blackwell Publishing.

Hornstein, Norbert, 2003. On control. In Randall Hendrick, ed., *Minimalist Syntax*, 6–81. Wiley.

Hornstein, Norbert, 2009. *A Theory of Syntax: Minimal Operations and Universal Grammar*. Cambridge: Cambridge University Press.

Hornstein, Norbert and Jairo Nunes, 2002. On asymmetries between parasitic gap and across-the-board constructions. *Syntax* 5:26–54.

Hornstein, Norbert and Amy Weinberg, 1981. Case theory and preposition stranding. *Linguistic Inquiry* 12:55–91.

Hrdy, Sarah Blaffer, 2007. Evolutionary context of human development: The cooperative breeding model. In Catherine Salmon and Todd Schackelford, eds., *Family Relationships: An Evolutionary Perspective*, 39–68. New York: Oxford University Press.

Hrdy, Sarah Blaffer, 2009. *Mothers and Others*. Cambridge, MA: Belnap/Harvard University Press.

Huang, C.-T. James, 1982a. *Logical Relations in Chinese and the Theory of Grammar*. Ph.D. thesis, MIT.

Huang, C.-T. James, 1982b. Move wh in a language without wh movement. *Linguistic Review* 1:369–416.

Huber, Ludwig and Gayla Gajdon, 2006. Technical intelligence in animals: The kea model. *Animal Cognition* 9:295–305.

Huddleston, Rodney and Geoffrey Pullum, 2002. *Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Hudson, Richard, 1976. Conjunction reduction, gapping, and right-node raising. *Language* 535–562.

Hull, Clark, 1943. *Principles of Behavior*. Appleton-Century-Crofts.

Hume, David, 1738. *Treatise of Human Nature*. Oxford: Clarendon Press. Ed. Selby-Bigge, 1902.

Huybregts, Riny, 1984. The weak inadequacy of context-free phrase-structure grammars. In Ger de Haan, Mieke Trommelen, and Wim Zonneveld, eds., *Van Periferie naar Kern*, 81–99. Dordrecht: Foris.

Hwang, Jena, Archna Bhatia, Clare Bonial, Aous Mansouri, Ashwini Vaidya, Nianwen Xue, and Martha Palmer, 2010. Propbank annotation of multilingual light verb constructions. In *Proceedings of the Fourth Linguistic Annotation Workshop*, 82–90. Association for Computational Linguistics.

Hyams, Nina, 1986. *Language Acquisition and the Theory of Parameters*. Dordrecht: Reidel.

Isard, Stephen, 1974. What would you have done if .... *Theoretical Linguistics* 1:233–255.

Ishii, Toru, 2004. The phase impenetrability condition, the vacuous movement hypothesis, and *that-t* effects. *Lingua* 114:183–215.

Jackendoff, Ray, 1972. *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.

Jackendoff, Ray, 1990. *Semantic Structures*. Cambridge, MA: MIT Press.

Jackendoff, Ray, 2002. *Foundations of Language*. Oxford: Oxford University Press.

Jacobson, Pauline, 1992a. Flexible categorial grammars: Questions and prospects. In Robert Levine, ed., *Formal Grammar*, 129–167. Oxford: Oxford University Press.

Jacobson, Pauline, 1992b. The lexical entailment theory of control and the *tough* construction. In Ivan Sag and Anna Szabolcsi, eds., *Lexical Matters*, 269–300. Stanford, CA: CSLI Publications.

Jacobson, Pauline, 1994. Binding connectivity in copular sentences. In *Semantics and Linguistic Theory*, vol. 4, 161–178. Cornell University.

Jacobson, Pauline, 1999. Towards a variable-free semantics. *Linguistics and Philosophy* 22:117–184.

Jacobson, Pauline, 2004. Direct compositionality: Is there any reason why not? In *Paper to Workshop in Linguistics and Philosophy*. University of Michigan.

Jacobson, Pauline, 2006. Constituent structure. In Keith Brown, ed., *The Encyclopedia of Language and Linguistics*, vol. 3, 58–71. Amsterdam: Elsevier, 2 edn.

Jacobson, Pauline, 2007. Direct compositionality and variable-free semantics: The case of "Principle B" effects. In Chris Barker and Pauline Jacobson, eds., *Direct Compositionality*, 191–236. Oxford: Oxford University Press.

Jarvis, Erich, 2004. Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences* 1016:749–777.

Jelinek, Eloise, 1984. Empty categories, case, and configurationality. *Natural Language & Linguistic Theory* 2:39–76.

Jespersen, Otto, 1922. *Language: Its Nature, Development, and Origin*. Allen and Unwin.

Jespersen, Otto, 1924. *The Philosophy of Grammar*. Allen and Unwin.

Johannessen, Janne Bondi, 2005. The syntax of correlative adverbs. *Lingua* 115:419–443.

Johnson, Kyle, 2000. How far will quantifiers go? In Roger Martin, David Michaels, and Juan Uriagereka, eds., *Step by Step*, 187–210. Cambridge, MA: MIT Press.

Johnson, Kyle, 2020. Principle C. *submitted* .

Johnson, Mark, 1986. A GPSG account of VP structure in German. *Linguistics* 24:871–882.

Joshi, Aravind, 1985. Tree-Adjoining Grammars: How much Context-Sensitivity id Required to Provide Sensible Structural Descriptions? In David Dowty, Lauri Karttunen, and Arnold Zwicky, eds., *Natural Language Parsing*, 206–250. Cambridge: Cambridge University Press.

Joshi, Aravind and Yves Schabes, 1997. Tree-Adjoining Grammars. In Grzegorz Rozenberg and Arto Salomaa, eds., *Handbook of Formal Languages*, vol. 3, 69–124. Berlin: Springer.

Joshi, Aravind, K. Vijay-Shanker, and David Weir, 1991. The convergence of mildly context-sensitive formalisms. In Peter Sells, Stuart Shieber, and Tom Wasow, eds., *Processing of Linguistic Structure*, 31–81. Cambridge, MA: MIT Press.

Jurka, Johannes, 2013. Subject islands in German revisited. In Jon Sprouse and Norbert Hornstein, eds., *Experimental Syntax and Island Effects*, 265–285. Cambridge University Press.

Kaisse, Ellen, 1985. *Connected Speech: The Interaction of Syntax and Phonology*. Orlando, FL: Academic Press.

Kamp, Hans and Uwe Reyle, 1993. *From Discourse to Logic*. Dordrecht: Kluwer.

Karttunen, Lauri, 1976. Discourse referents. In James McCawley, ed., *Syntax and Semantics*, vol. 7, 363–385. Academic Press.

Karttunen, Lauri and Stanley Peters, 1979. Conventional implicature. In Choon-Kyu Oh and David Dinneen, eds., *Syntax and Semantics 11: Presupposition*, 1–56. New York: Academic Press.

Kato, Yoshihide and Shigeki Matsubara, 2015. Incremental semantic construction using normal form CCG derivation. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, 269–278. Denver, CO: ACL.

Kay, Paul, 2002. An informal sketch of a formal architecture for construction grammar. *Grammars* 5:1–19.

Kay, Paul and Charles Fillmore, 1999. Grammatical constructions and linguistic generalizations: the *what's x doing y?* construction. *Language* 1–33.

Kay, Paul and Ivan Sag, 2012. Cleaning up the big mess: Discontinuous dependencies and complex determiners. In Hans Boas and Ivan Sag, eds., *Sign-Based Construction Grammar*, 229–256. Stanford, CA: CSLI Publications.

Kayne, Richard, 1975. *French Syntax*. Cambridge, MA: MIT Press.

Kayne, Richard, 1981. ECP extensions. *Linguistic Inquiry* 93–133. Reprinted as Kayne 1983:ch.3.

Kayne, Richard, 1983. *Connectedness and Binary Branching*. Dordrecht: Foris.

Kayne, Richard, 1994. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.

Keenan, Edward and Bernard Comrie, 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8:63–99.

Kehler, Andrew, 2002. *Coherence, Reference, and the Theory of Grammar*. Stanford, CA: CSLI Publications.

Keine, Stefan and Ethan Poole, 2018. Not all reconstruction effects are syntactic. *Ms., University of Southern California and University of California, Los Angeles* .

Kelleher, John D., Geert-Jan M. Kruijff, and Fintan J. Costello, 2006. Proximity in context: An empirically grounded computational model of proximity for processing topological spatial expressions. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, 745–752. Sydney, Australia: ACL.

Keller, William, 1988. Nested Cooper storage. In Uwe Reyle and Christian Rohrer, eds., *Natural Language Parsing and Linguistic Theory*, 432–447. Dordrecht: Reidel.

Kempson, Ruth and Annabel Cormack, 1981. Ambiguity and quantification. *Linguistics and Philosophy* 4:259–309.

Kennedy, 1882. *The Public School Latin Primer*. Longmans, Green and Co. Revised ed. 1930.

Kennedy, Christopher and Jason Merchant, 2000. Attributive comparative deletion. *Natural Language & Linguistic Theory* 18:89–146.

Kim, Jong-Bok and Peter Sells, 2011. The big mess construction: interactions between the lexicon and constructions. *English Language and Linguistics* 15:335–362.

É. Kiss, Katalin, 1991. Logical structure in syntactic structure: The case of Hungarian. In James Huang and Robert May, eds., *Logical Structure and Linguistic Structure: Cross-Linguistic Perspectives*, 111–147. Dordrecht: Kluwer.

É. Kiss, Katalin, 1994. Sentence structure and word-order. In Ferenc Kiefer and Katalin É. Kiss, eds., *Syntax and Semantics 27: The Syntactic Structure of Hungarian*, 1–90. San Diego, CA: Academic Press.

Klein, Ewan and Ivan Sag, 1985. Type-driven translation. *Linguistics and Philosophy* 8:163–201.

Kluender, Robert, 1998. On the distinction between strong and weak islands: A processing perspective. In Peter Culicover and Louise McNally, eds., *Syntax and Semantics 29: The Limits of Syntax*, 241–280.

Köhler, Wolfgang, 1925. *The Mentality of Apes*. New York: Harcourt Brace and World.

Koller, Alexander and Marco Kuhlmann, 2009. Dependency trees and the strong generative capacity of CCG. In *Proceedings of the 12th Conference of the European Chapter of the ACL*, 460–468. Athens: ACL.

Komagata, Nobo, 1999. *Information Structure in Texts: A Computational Analysis of Contextual Appropriateness in English and Japanese*. Ph.D. thesis, University of Pennsylvania.

Koopman, Hilda, 1983. ECP effects in main clauses. *Linguistic Inquiry* 346–350.

Koopman, Hilda, 1996. The spec head configuration. *Syntax at sunset: UCLA working papers in syntax and semantics* 1(37-65).

Koopman, Hilda and Anna Szabolcsi, 2000. *Verbal Complexes*. Cambridge, MA: MIT Press.

Koster, Jan, 1986. *Domains and Dynasties*. Dordrecht: Foris.

Kratzer, Angelika, 1988/1995. Stage-level and individual-level predicates. In Gregory Carlson and Geoffrey Pelletier, eds., *The Generic Book*, 125–175. Chicago University Press.

Kratzer, Angelika, 1998. Scope or pseudo-scope: Are there wide-scope indefinites? In Susan Rothstein, ed., *Events in Grammar*, 163–196. Dordrecht: Kluwer.

Kratzer, Angelika and Elisabeth Selkirk, 2007. Phase theory and prosodic spellout: The case of verbs. *The Linguistic Review* 24:93–135.

Krishnamurthy, Jayant and Tom Mitchell, 2012. Weakly supervised training of semantic parsers. In *Proceedings of Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 754–765. Jeju Island, Korea: ACL.

Krishnamurthy, Jayant and Tom Mitchell, 2013. Vector space semantic parsing: A framework for compositional vector space models. In *Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality*, 1–10. Sofia: ACL.

Krishnamurthy, Jayant and Tom Mitchell, 2014. Joint syntactic and semantic parsing with combinatory categorial grammar. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1188–1198. Baltimore, MD.

Kruijff-Korbayová, Ivana and Mark Steedman, 2003. Discourse and information structure. *Journal of Logic, Language, and Information* 12:249–259.

Kubota, Yusuke, 2010. *(In)flexibility of Constituency in Japanese in Multi-Modal Categorial Grammar with Structured Phonology*. Ph.D. thesis, Ohio State University.

Kubota, Yusuke and Robert Levine, 2020. *Type-Logical Syntax*. MIT Press.

Kuhlmann, Marco, Alexander Koller, and Giorgio Satta, 2010. The importance of rule restrictions in CCG. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 534–543. Uppsala: ACL.

Kuhlmann, Marco, Alexander Koller, and Giorgio Satta, 2015. Lexicalization and generative power in CCG. *Computational Linguistics* 41:187–219.

Kuncoro, Adhiguna, Miguel Ballesteros, Lingpeng Kong, Chris Dyer, Graham Neubig, and Noah Smith, 2017. What do recurrent neural network grammars learn about syntax? In *Proceedings of the 15th Meeting of the European Chapter of the ACL*, 124–1258. ACL.

Kuno, Susumu, 1973a. Constraints on internal clauses and sentential subjects. *Linguistic Inquiry* 4:363–386.

Kuno, Susumu, 1973b. *The Structure of the Japanese Language*. Cambridge, MA: MIT Press.

Kuno, Susumu, 1974. The position of relative clauses and conjunctions. *Linguistic Inquiry* 5:117–136.

Kuno, Susumu, 1987. *Functional Syntax: Anaphora, Discourse, and Empathy*. Chicago: University of Chicago Press.

Kuroda, Sige-Yuki, 1965. *Generative Grammatical Studies in the Japanese Language*. Ph.D. thesis, MIT, Cambridge, MA. Pub. 1979 by Garland, New York.

Kwiatkowski, Tom, 2012. *Probabilistic Grammar Induction from Sentences and Structured Meanings*. Ph.D. thesis, University of Edinburgh.

Kwiatkowski, Tom, Sharon Goldwater, Luke Zettlemoyer, and Mark Steedman, 2012. A probabilistic model of syntactic and semantic acquisition from child-directed utterances and their meanings. In *Proceedings of the 13th Conference of the European Chapter of the ACL (EACL 2012)*, 234–244. Avignon: ACL.

Kwiatkowski, Tom, Luke Zettlemoyer, Sharon Goldwater, and Mark Steedman, 2010. Inducing probabilistic CCG grammars from logical form with higher-order unification. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 1223–1233. Cambridge, MA: ACL.

Ladd, D. Robert, 1996. *Intonational Phonology*. Cambridge: Cambridge University Press. 2nd edition revised 2008.

Ladd, D. Robert, 2008. *Intonational Phonology*. Cambridge: Cambridge University Press, 2nd edn.

Lakoff, George, 1970a. *Irregularity in Syntax*. New York: Holt, Rinehart & Winston.

Lakoff, George, 1970b. Repartee, or a reply to "negation, conjunction and quantifiers". *Foundations of Language* 6:389–422.

Lakoff, George, 1986. Frame semantic control of the coordinate structure constraint. In Anne Farley, Peter Farley, and Karl-Eric McCullough, eds., *Papers from the Parasession on Pragmatics and Grammatical Theory at the 22nd Regional Meeting of the Chicago Linguistics Society*, 152–167. Chicago: CLS.

Lambek, Joachim, 1958. The mathematics of sentence structure. *American Mathematical Monthly* 65:154–170.

Lambek, Joachim, 2001. Type grammars as pregroups. *Grammars* 4:21–39.

Landau, Idan, 2001. *Elements of Control: Structure and Meaning in Infinitival Constructions*. Springer.

Landau, Idan, 2015. *A Two-Tiered Theory of Control*. MIT Press.

Landau, Idan, 2021. *A Selectional Theory of Adjunct Control*. MIT Press.

Landauer, Thomas and S. Dumais, 1997. A solution to plato's problem: The latent semantic analysis of the acquisition, induction and representation of knowledge. *Psychological Review* 104:211–240.

Langacker, Ronald, 1988. A usage-based model. In Brygida Rudzka-Ostyn, ed., *Topics in Cognitive Linguistics*, 127–161. Amsterdam: John Benjamins.

Larson, Richard, 1985. On the syntax of disjunction scope. *Natural Language & Linguistic Theory* 3:217–264.

Larson, Richard, 1988. On the double object construction. *Linguistic Inquiry* 19:335–392.

Larson, Richard, 1990. Double objects revisited: Reply to Jackendoff. *Linguistic Inquiry* 21:589–632.

Lasnik, Howard, 2001. Subjects, objects, and the EPP. In William Davies and Stanley Dubinsky, eds., *Objects and other Subjects*, 103–121. Springer.

Lasnik, Howard, 2011. What kind of computing device is the human language faculty? In Anna Maria Di Sciullo and Cedric Boeckx, eds., *The Biolininguistic Enterprise*, 354–65. Oxford University Press.

Lasnik, Howard and Mamoru Saito, 1984. On the nature of proper government. *Linguistic Inquiry* 15:235–289.

Lasnik, Howard and Tim Stowell, 1991. Weakest crossover. *Linguistic inquiry* 687–720.

Lebeaux, David, 1991. Relative clauses, licensing, and the nature of the derivation. In *Perspectives on phrase structure: Heads and licensing*, 209–239. San Diego: Academic Press.

Lebeaux, David, 2009. *Where does Binding Theory Apply?* Mit Press.

Lechner, Winfried, 1998. Two kinds of reconstruction. *Studia Linguistica* 52:276–310.

Lechner, Winfried, 2018. A calculus for reconstruction and anti-reconstruction. In Manfred Krifka and Mathias Schenner, eds., *Reconstruction Effects in Relative Clauses*, 113–144. De Gruyter.

Lee, Jungmee and Judith Tonhauser, 2010. Temporal interpretation without tense: Korean and Japanese coordination constructions. *Journal of Semantics* 27:307–341.

Lee, Kenton, Mike Lewis, and Luke Zettlemoyer, 2016. Global neural CCG parsing with optimality guarantees. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2366–2376. ACL.

Lees, Robert, 1960. A multiply ambiguous adjectival construction in English. *Language* 207–221.

Legate, Julie, 2008. Morphological and abstract case. *Linguistic Inquiry* 39:55–101.

Legate, Julie Anne, 1999. The morphosyntax of Irish agreement. *MIT Working Papers in Linguistics* 33:219–240.

Legate, Julie Anne, 2003. Some interface properties of the phase. *Linguistic Inquiry* 34:506–515.

Lehmann, Winfred, 1973. A structural principle of language and its implications. *Language* 47–66.

Levin, Beth, 1993. *English Verb-Classes and Alternations: A Preliminary Study*. Chicago: University of Chicago Press.

Levin, Beth and Malka Rappaport-Hovav, 1995. *Unaccusativity: At the Syntax-Lexical Semantics Interface*. No. 26 in Linguistic Inquiry Monograph. Cambridge, MA: The MIT Press.

Levine, Robert, 2017. *Syntactic Analysis: An HPSG-based Approach*. Cambridge University Press.

Levine, Robert and Thomas Hukari, 2006. *The Unity of Unbounded Dependency Constructions*. Stanford, CA: CSLI Publications.

Lewis, David, 1969. *Convention: a Philosophical Study*. Cambridge, MA: Harvard University Press.

Lewis, David, 1979a. Attitudes de dicto and de se. *The philosophical review* 88(4):513–543.

Lewis, David, 1979b. Scorekeeping in a language game. *Journal of Philosophical Logic* 8:339–359.

Lewis, Mike, 2015. *Combined Distributional and Logical Semantics*. Ph.D. thesis, University of Edinburgh.

Lewis, Mike, Kenton Lee, and Luke Zettlemoyer, 2016. LSTM CCG parsing. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 221–231. San Diego CA: ACL.

Lewis, Mike and Mark Steedman, 2013a. Combined distributional and logical semantics. *Transactions of the Association for Computational Linguistics* 1:179–192.

Lewis, Mike and Mark Steedman, 2013b. Unsupervised induction of cross-lingual semantic relations. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 681–692. ACL.

Lewis, Mike and Mark Steedman, 2014a. A^* CCG parsing with a supertag-factored model. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 990–1000. Doha, Qatar: ACL.

Lewis, Mike and Mark Steedman, 2014b. Combining formal and distributional models of temporal and intensional semantics. In *Proceedings of the ACL Workshop on Semantic Parsing*, 28–32. Baltimore, MD: ACL. Google Exceptional Submission Award.

Lewis, Mike and Mark Steedman, 2014c. Improved CCG parsing with semi-supervised supertagging. *Transactions of the Association for Computational Linguistics* 2:327–338.

Lewontin, Richard, 1998. The evolution of cognition: Questions we will never answer. In *An Invitation to Cognitive Science*, vol. 4, 107–132. Cambridge, MA: MIT Press.

Liberman, Mark, 1975. *The Intonational System of English*. Ph.D. thesis, MIT. Published by Garland Press, New York, 1979.

Lieberman, Philip, Jeffrey Laitman, Joy Reidenberg, and Patrick Gannon, 1992. The anatomy, physiology, acoustics and perception of speech: Essential elements in analysis of the evolution of human speech. *Journal of Human Evolution* 23:447–467.

Lieberman, Philip and Robert McCarthy, 2007. Tracking the evolution of language and speech. *Expedition* 49:15–20.

Lin, Dekang, 1993. Principle-based parsing without overgeneration. In *Proceedings of the 31st annual meeting of the Association for Computational Linguistics*, 112–20. ACL.

Lin, Dekang, 1994. PRINCIPAR: an efficient, broad-coverage, principle-based parser. In *Proceedings of the 15th conference on Computational Linguistics-Volume 1*, 482–488. ACL.

Lin, Dekang, 1998. Automatic retrieval and clustering of similar words. In *Proceedings of the International Conference on Computational Linguistics/Annual Meeting of the Association for Computational Linguistics*, 768–774. Montreal.

Lin, Vivian I-Wen, 2002. *Coordination and Sharing at the Interfaces*. Ph.D. thesis, Massachusetts Institute of Technology.

Link, Godehard, 1983. The logical analysis of plurals and mass terms. In Rainer Bäuerle, C. Schwarze, and Arnim von Stechow, eds., *Meaning, Use, and Interpretation in Language*, 302–323. Berlin: de Gruyter. Reprinted in Link 1998.

Link, Godehard, 1998. *Algebraic Semantics in Language and Philosophy*. Stanford, CA: CSLI Publications.

MacWhinney, Brian, 2000. *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, NJ: Erlbaum.

MacWhinney, Brian, 2004. A multiple process solution to the logical problem of language acquisition. *Journal of Child Language* 31:883–914.

Maillard, Jean, Stephen Clark, and Edward Grefenstette, 2014. A type-driven tensor-based semantics for CCG. In *Proceedings of the EACL 2014 Type Theory and Natural Language Semantics Workshop (TTNLS)*, 46–54.

Maling, Joan and Soowon Kim, 1992. Case assignment in the inalienable possession construction in Korean. *Journal of East Asian Linguistics* 1:37–68.

Manaster-Ramer, Alexis, 1986. Copying in natural languages, context-freeness, and queue grammars. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, 85–89. ACL.

Manning, Christopher, 1996. *Ergativity: Argument Structure and Grammatical Relations*. Stanford, CA: CSLI Publications.

Manzini, Rita and Anna Roussou, 2000. A minimalist theory of A-movement and control. *Lingua* 110:409–447.

Marácz, László and Pieter Muysken, eds., 1989. *Configurationality: The Typology of Asymmetries*. Dordrecht: Foris.

Marcus, Mitch, 1980. *A Theory of Syntactic Recognition for Natural Language*. Cambridge, MA: MIT Press.

Marcus, Mitch, Beatrice Santorini, and M. Marcinkiewicz, 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19:313–330.

Marslen-Wilson, William, 1973. Linguistic structure and speech shadowing at very short latencies. *Nature* 244:522–523.

Mattys, Sven and Peter Jusczyk, 2001. Phonotactic cues for segmentation of fluent speech by infants. *Cognition* 78:91–121.

Mattys, Sven, Peter Jusczyk, Paul Luce, and Jim Morgan, 1999. Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology* 38:465–494.

Matuszek, Cynthia, Nicholas Fitzgerald, Luke Zettlemoyer, Liefeng Bo, and Dieter Fox, 2012a. A joint model of language and perception for grounded attribute learning. In *Proceedings of the 29th International Conference on Machine Learning (ICML)*, 1671–1678.

Matuszek, Cynthia, Evan Herbst, Luke Zettlemoyer, and Dieter Fox, 2012b. Learning to parse natural language commands to a robot control system. In *Proceedings of the 13th International Symposium on Experimental Robotics (ISER)*, 403–415.

Maxwell, John and Christopher Manning, 1996. A theory of non-constituent coordination based on finite state rules. In *Proceedings of the First LFG Conference*. Grenoble.

May, Robert, 1977. *The Grammar of Quantification*. Ph.D. thesis, MIT.

May, Robert, 1985. *Logical Form*. Cambridge, MA: MIT Press.

Maynard-Smith, John and Eörs Szathmáry, 1995. *The Major Transitions in Evolution*. Oxford: W.H.Freeman/Spektrum.

McCawley, James, 1968. Lexical insertion in a transformational grammar without deep structure. In *Papers from the 4th Regional Meeting of the Chicago Linguistics Society*, 71–80. CLS.

McCawley, James, 1982. Parentheticals and discontinuous constituent structure. *Linguistic Inquiry* 13:91–106.

McCawley, James, 1987. Some additional evidence for discontinuity. In Almerindo Ojeda and Geoffrey Huck, eds., *Discontinuous Constituency*, Syntax and Semantics 20, 185–200. New York: Academic Press.

McCloskey, James, 1979. *Transformational Syntax and Model Theoretic Semantics: A Case Study in Modern Irish*. Reidel.

McCloskey, James, 1986. Right node raising and preposition stranding. *Linguistic Inquiry* 17:183–186.

McCloskey, James, 1991. There, it, and agreement. *Linguistic Inquiry* 22:563–567.

McCloskey, James, 2002. Resumption, successive cyclicity, and the locality of operations. In Sam Epstein and Daniel Seely, eds., *Derivation and Explanation in the Minimalist Program*, 184–226. Oxford: Blackwell.

McCloskey, James, 2017. New thoughts on old questions–resumption in Irish. In James Ostrove, Ruth Kramer, and Joseph Sabbagh, eds., *Asking the Right Question: Essays in Honor of Sandra Chung*, 81–102. Santa Cruz, CA: University of California.

McDermott, Drew, Mallik Ghallab, Adèle Howe, Craig Knoblock, Ashwin Ram, Manuela Veloso, Daniel Weld, and David Wilkins, 1998. *PDDL The Planning Domain Definition Language*. DCS TR-1165, Yale Center for Computational Vision and Control, New Haven, CT.

McDonald, Ryan and Fernando Pereira, 2006a. Multilingual dependency analysis with a two-stage discriminative parser. In *Proceedings of the 10th Conference on Natural Language Learning*. New Brunswick, NJ: ACL.

McDonald, Ryan and Fernando Pereira, 2006b. Online learning of approximate dependency parsing algorithms. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics,*, 81–88. New Brunswick, NJ: ACL.

McDonald, Ryan, Fernando Pereira, Kiril Ribarov, and Jan Hajic, 2005. Non-projective dependency parsing using spanning tree algorithms. In *Proceedings of the Joint Conference on Human Laguage Technologies and Empirical Methods in Natural Language Processing HLT/EMNLP,*, 523–530. New Brunswick, NJ: ACL.

Mehay, Denis and Chris Brew, 2012. CCG syntactic reordering models for phrase-based machine translation. In *Proceedings of the Seventh Workshop on Statistical Machine Translation*, 210–221. Montréal: ACL.

Merchant, Jason, 2001. *The Syntax of Silence: Sluicing, Islands and the Theory of Ellipsis*. Oxford: Oxford University Press.

Merchant, Jason, 2006. Sluicing. In Martin Everaert and Henk van Riemsdijk, eds., *The Blackwell Companion to Syntax*, vol. IV, 271–291. Wiley.

Merchant, Jason, 2013. Voice and ellipsis. *Linguistic Inquiry* 44:77–108.

Meyer-Eppler, W, 1957. Realization of prosodic features in whispered speech. *Journal of the Acoustical Society of America* 29:104–106. Reprinted as Bolinger 1972b:385–390.

Michaelis, Jens, 2001. Transforming linear context-free rewriting systems into minimalist grammars. In Phillipe de Groote, Glyn Morrill, and Christian Retoré, eds., *Logical Aspects of Computational Linguistics (LACL'01)*, Lecture Notes in Computer Science 2099, 228–244. Berlin: Springer.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean, 2013a. Efficient estimation of word representations in vector space. In *International Conference on Learning Representations*.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeff Dean, 2013b. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, 3111–3119.

Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig, 2013c. Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 746–751. Atlanta: ACL.

Miller, George, 1967. Project grammarama. In George Miller, ed., *The Psychology of Communication*, 125–187. New York: Basic Books.

Miller, George, Eugene Galanter, and Karl Pribram, 1960. *Plans and the Structure of Behavior*. New York: Holt.

Misra, Dipendra Kumar and Yoav Artzhi, 2016. Neural shift-reduce ccg semantic parsing. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 1775–1786. Austin, TX: ACL.

Mitchell, Jeff and Mark Steedman, 2015. Orthogonality of syntax and semantics within distributional spaces. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, 1301–1310. Beijing: ACL.

Mitchell, Tom, William Cohen, Estevam Hruschka, Partha Talukdar, Justin Betteridge, Andrew Carlson, Bhavana Dalvi Mishra, Matthew Gardner, Bryan Kisiel, and Jayant Krishnamurthy, 2015. Never-ending learning. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2302–2310.

Miyagawa, Shigeru, 1997. Against optional scrambling. *Linguistic Inquiry* 28:1–25.

Miyagawa, Shigeru, 2017. Topicalization. *Gengo Kenkyu (Journal of the Linguistic Society of Japan)* 152:1–29.

Moens, Marc and Mark Steedman, 1987. Temporal ontology in natural language,. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics, Stanford, CA, July*, 1–7. Reprinted in Javier Gutieérrez-Rexach (ed.), *Semantics: Critical Concepts in Linguistics* Vol. IV: The Semantics of Predicates and Inflection, 290-305, Routledge.

Moens, Marc and Mark Steedman, 1988. Temporal ontology and temporal reference. *Computational Linguistics* 14:15–28. Reprinted in Inderjeet Mani, James Pustejovsky, and Robert Gaizauskas (eds.) *The Language of Time: A Reader*. Oxford University Press, 93-114.

Monahan, Philip, 2003. *Backward Object Control in Korean*. Ph.D. thesis, University of Florida.

Montague, Richard, 1973. The proper treatment of quantification in ordinary English. In Jaakko Hintikka, Julius Moravcsik, and Patrick Suppes, eds., *Approaches to Natural Language: Proceedings of the 1970 Stanford Workshop on Grammar and Semantics*, 221–242. Dordrecht: Reidel. Reprinted as Thomason 1974:247-279.

Moortgat, Michael, 1988. *Categorial Investigations*. Ph.D. thesis, Universiteit van Amsterdam. Published by Foris, Dordrecht, 1989.

Moro, Andrea and Roberto Navigli, 2012. Wisenet: Building a Wikipedia-based semantic network with ontologized relations. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, 1672–1676. ACM.

Morrill, Glyn, 1994. *Type-Logical Grammar*. Dordrecht: Kluwer.

Morrill, Glyn, 1995. Discontinuity in Categorial grammar. *Linguistics and Philosophy* 18:175–219.

Morrill, Glynn, 2011. *Categorial Grammar: Logical Syntax, Semantics, and Processing*. Oxford: Oxford University Press.

Mueller, Hans-Friedrich, ed., 2012. *Caesar: Selections from his* Commentarii de Bello Gallico. Bolchazy-Carducci.

Müller, Gereon, 1998. *Incomplete Category Fronting: A Derivational Approach to Remnant Movement in German*. Springer.

Müller, Gereon, 2010. On deriving CED effects from the PIC. *Linguistic Inquiry* 41:35–82.

Müller, Stefan, 1996. Yet another paper about partial verb phrase fronting in German. In *Proceedings of the 16th International Conference on Computational Linguistics-Volume 2*, 800–805. ACL.

Müller, Stefan, 2007. Qualitative korpusanalyse für die grammatiktheorie: Introspektion vs. korpus. In Werner Zifonun, Gisella und Kallmeyer, ed., *Sprachkorpora–Datenmengen und Erkenntnisfortschritt*, 70–90. Berlin: de Gruyter.

Munn, Alan, 1993. *Topics in the Syntax and Semantics of Coordinate Structures*. Ph.D. thesis, University of Maryland.

Nakashole, Ndapandula, Gerhard Weikum, and Fabian Suchanek, 2012. Patty: A taxonomy of relational patterns with semantic types. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 1135–1145. ACL.

Naseem, Tahira, Harr Chen, Regina Barzilay, and Mark Johnson, 2010. Using universal linguistic knowledge to guide grammar induction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 1234–1244. ACL.

Navigli, Roberto and Simone Paolo Ponzetto, 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence* 193:217–250.

Neal, Radford and Geoffrey Hinton, 1999. A view of the em algorithm that justifies incremental, sparse, and other variants. In Michael Jordan, ed., *Learning in Graphical Models*, 355–368. Cambridge, MA: MIT Press.

Neelakantan, Arvind, Benjamin Roth, and Andrew McCallum, 2015. Compositional vector space models for knowledge base completion. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 156–166. Beijing: ACL.

Neeleman, Ad and Hans van de Koot, 2010. A local encoding of syntactic dependencies. *Syntax* 13:331–372.

Nerbonne, John, 1986. “phantoms” and German fronting: Poltergeist constituents? *Linguistics* 24:857–870.

Newbolt, Henry, 1898. Vitaï lampada. In *Admirals All*. London: Elkin Matthews.

Newman, Elise, 2020. Facilitator effects in middles and more. *Glossa* 5(62):1–40.

Newmeyer, Frederick, 2005. *Possible and Probable Languages*. Oxford: Oxford University Press.

Ng, Dominick and James R Curran, 2012. Dependency hashing for n-best CCG parsing. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, 497–505.

Nishida, Chiyo, 1996. Second position clitics in Old Spanish and Categorial Grammar. In Aaron Halpern and Arnold Zwicky, eds., *Approaching Second: Second-Position Clitics and Related Phenomena*, 33–373. Stanford, CA: CSLI Publications.

Nishiguchi, Taisuke, 1990. *Quantificiation in the Theory of Grammar*. Dordrecht: Kluwer.

Nivre, Joakim, 2003. An efficient algorithm for projective dependency parsing. In *Proceedings of the 8th International Workshop on Parsing Technologies (IWPT*, 149–160. ACL.

Nivre, Joakim, 2004. Incrementality in deterministic dependency parsing. In *Proceedings of the Workshop on Incremental Parsing: Bringing Engineering and Cognition Together*, 50–57. ACL.

Nivre, Joakim, 2006. *Inductive Dependency Parsing*. Dordrecht: Springer.

Niyogi, Partha, 2006. *The Computational Nature of Language Learning and Evolution*. Cambridge, MA: MIT press.

Nordlinger, Rachel, 1997. Morphology building syntax: Constructive case in australian languages. In *Proceedings of the LFG97 Conference*. CSLI Publications, Stanford.

Nordlinger, Rachel, 1998. *Constructive Case: Dependent-Marking Nonconfigurationality in Australia*. Stanford, CA: CSLI Publications.

Nordlinger, Rachel, 2000. Australian case systems: Towards a constructive solution. In Miriam Butt and Tracy Holloway King, eds., *Argument Realization*, 41–72. Stanford, CA: CSLI Publications.

Nădejde, Maria, Siva Reddy, Rico Sennrich, Tomasz Dwojak, Marcin Junczys-Dowmunt, Philipp Koehn, and Alexandra Birch, 2017. Predicting target language ccg supertags improves neural machine translation. In *Proceedings of the 2nd Conference on Machine Translation*, 68–79. ACL.

Nunes, Jairo, 1995. *The Copy Theory of Movement and Linearization of Chains in the Minimalist Program*. Ph.D. thesis, University of Maryland.

Nunes, Jairo, 2001. Sideward movement. *Linguistic Inquiry* 32:303–344.

Nunes, Jairo, 2004. *Linearization of Chains and Sideward Movement*. Linguistic Inquiry Monograph 43. Cambridge, MA: MIT Press.

O'Day, Anita, 1962. Time for two. Verve. With Cal Tjader.

Oehrle, Richard, 1976. *The Grammatical Status of the English Dative Alternation*. Ph.D. thesis, MIT.

Oléron, Paul, 1953. Conceptual thinking of the deaf. *American Annals of the Deaf* 98:304–310.

Osborne, Timothy and Thomas Groß, 2016. Left node blocking. *Journal of Linguistics* 53:641–688.

Osgood, Charles, Charles Suci, and Percy Tannenbaum, 1957. *The Measurement of Meaning*. Champaign Urbana, IL: University of Illinois Press.

Otsuka, Yuko, 2005. Scrambling and information focus: VSO-VOS alternation in Tongan. In Joachim Sabel and Mamoru Saito, eds., *The Free Word Order Phenomenon: Its Syntactic Sources and Diversity*, 243–280. Berlin: Mouton de Gruyter.

Padó, Sebastian and Mirella Lapata, 2007. Dependency-based construction of semantic space models. *Computational Linguistics* 33:161–199.

Park, Jong and Hyung-Joon Cho, 2000. Informed parsing for coordination with combinatory categorial grammar. In *Proceedings of the 18th International Conference on Computational Linguistics*, 593–599. Saarbrücken: Coling/ACL.

Partee, Barbara, 1977. John is easy to please. In Antonio Zampoli, ed., *Linguistic Structures Processing*, 281–312. Amsterdam: North Holland.

Partee, Barbara and Mats Rooth, 1983. Generalised conjunction and type ambiguity. In Rainer Bäuerle, Christoph Schwarze, and Arnim von Stechow, eds., *Meaning, Use, and Interpretation of Language*, 361–383. Berlin: de Gruyter.

Perlmutter, David, 1971. *Deep and Surface Structure Constraints in Syntax*. New York: Holt, Rinehart& Winston.

Pesetsky, David, 1982. *Paths and Categories*. Ph.D. thesis, MIT.

Pesetsky, David, 1987. *Wh*–in situ: Movement and unselective binding. In Eric Reuland and Alice ter Meulen, eds., *The Representation of (In)definiteness*, 98–129. Cambridge, MA: MIT Press.

Pesetsky, David, 1995. *Zero Syntax*. Cambridge, MA: MIT Press.

Pesetsky, David, 2017. Complementizer-*trace* effects. In Martin Everaert and Henk van Riemsdijk, eds., *Blackwell Companion to Syntax*, to appear. New York: Wiley.

Petrick, Ronald and Fahiem Bacchus, 2002. A knowledge-based approach to planning with incomplete information and sensing. In *Proceedings of the Sixth International Conference on Artificial Intelligence Planning and Scheduling (AIPS-2002)*, 212–221. Menlo Park, CA: AAAI Press.

Petrick, Ronald and Fahiem Bacchus, 2004. Extending the knowledge-based approach to planning with incomplete information and sensing. In *Proceedings of the International Conference on Automated Planning and Scheduling (ICAPS-04)*, 2–11. Menlo Park, CA: AAAI Press.

Phillips, Colin, 1996. *Order and Structure*. Ph.D. thesis, MIT.

Phillips, Colin, 2003. Linear order and constituency. *Linguistic Inquiry* 34:37–90.

Pierrehumbert, Janet, 1980. *The Phonology and Phonetics of English Intonation*. Ph.D. thesis, MIT. Distributed by Indiana University Linguistics Club, Bloomington.

Pierrehumbert, Janet and Mary Beckman, 1988. *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Pierrehumbert, Janet and Julia Hirschberg, 1990. The meaning of intonational contours in the interpretation of discourse. In Philip Cohen, Jerry Morgan, and Martha Pollack, eds., *Intentions in Communication*, 271–312. Cambridge, MA: MIT Press.

Pinker, Steven, 1979. Formal models of language learning. *Cognition* 7:217–283.

Pinker, Steven, 1989. *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, MA: The MIT Press.

Pinker, Steven, 1994. *The Language Instinct*. New York: Harper Perennial.

Plank, Frans, 1995. *Double Case: Agreement by Suffixaufnahme*. Oxford University Press.

Polinsky, Maria, 2012. Tsez beginnings. In *Proceedings of the 25th Annual Meeting of the Berkeley Linguistics Society: Special Session on Caucasian, Dravidian, and Turkic Linguistics*, 14–29.

Polinsky, Maria, Carlos Gallo, Peter Graff, Ekaterina Kravtchenko, Adam Milton Morgan, and Anne Sturgeon, 2013. Subject islands are different. In Jon Sprouse and Norbert Hornstein, eds., *Experimental Syntax and Island Effects*, 286–309. Cambridge University Press.

Pollard, Carl, 1984. *Generalized Context Free Grammars, Head Grammars, and Natural Languages*. Ph.D. thesis, Stanford University.

Pollard, Carl and Ivan Sag, 1982. Reflexives and reciprocals in English: An alternative to the binding theory. In M. Barlow et al, ed., *Proceedings of the 2nd West Coast Conference on Formal Linguistics, Stanford, CA*, 189–203.

Pollard, Carl and Ivan Sag, 1992. Anaphors in English and the scope of binding theory. *Linguistic Inquiry* 23:261–303.

Pollard, Carl and Ivan Sag, 1994. *Head Driven Phrase Structure Grammar*. Stanford, CA: CSLI Publications.

Pollock, Jean-Yves, 1989. Verb movement, universal grammar, and the structure of ip. *Linguistic Inquiry* 20:365–424.

Ponvert, Elias, 2008. Two approaches to Mayan grammar development in CCG. In *Proceedings of the Conference of the Texas Linguistics Society*, 113. CSLI Publications.

Poon, Hoifung and Pedro Domingos, 2008. Joint unsupervised coreference resolution with markov logic. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 650–659. Honolulu: ACL.

Portner, Paul, 2003. The (temporal) semantics and (modal) pragmatics of the perfect. *Linguistics and Philosophy* 26:459–510.

Postal, Paul, 1974. *On Raising: One Rule of English Grammar and its Theoretical Implications*. MIT Press.

Postal, Paul, 1993. Parasitic gaps and the across-the-board phenomenon. *Linguistic Inquiry* 24:735–754.

Postal, Paul, 1994. Parasitic and pseudoparasitic gaps. *Linguistic inquiry* 25:63–117.

Postal, Paul, 1998. *Three Investigations of Extraction*. Cambridge, MA: MIT Press.

Postal, Paul and John Ross, 2009. Inverse reflexives. In William Lewis, Simin Karimi, Heidi Harley, and Scott Farrar, eds., *Time and Again: Theoretical Perspectives on Formal Linguistics: In Honor of D. Terrence Langendoen*. Amsterdam: John Benjamins.

Prevost, Scott, 1995. *A Semantics of Contrast and Information Structure for Specifying Intonation in Spoken Language Generation*. Ph.D. thesis, University of Pennsylvania.

Prince, Ellen, 1981. Towards a taxonomy of given-new information. In Peter Cole, ed., *Radical Pragmatics*, 223–256. New York: Academic Press.

Progovac, Ljiljana, 1998. Structure for coordination, I and II. *Glot International* 3(7,8):3–6,3–9.

Pullum, Geoffrey and Gerald Gazdar, 1982. Natural languages and context-free languages. *Linguistics and Philosophy* 4:471–504.

Pulman, Stephen, 2013. Second order inference in NL semantics. Talk given at the KCL Language and Cognition seminar, London.

Pustejovsky, James, 1991. The generative lexicon. *Computational Linguistics* 17:409–441.

Pustejovsky, James, 1998. *The Generative Lexicon*. MIT Press.

Quine, Willard Van Orman, 1960. *Word and Object*. Cambridge, MA: MIT Press.

Radford, Andrew, 2004. *Minimalist Syntax: Exploring the Structure of English*. Cambridge University Press.

Rajeh, Hamdi Ahmed, Zhiyong Li, and Abdullah Mohammed Ayedh, 2016. A novel approach by injecting ccg supertags into an arabic–english factored translation machine. *Arabian Journal for Science and Engineering* 1–10.

Ramchand, Gillian Catriona, 1996. Two subject positions in Scottish Gaelic: The syntax-semantics interface. *Natural Language Semantics* 4:165–191.

Rando, Emily and Donna Jo Napoli, 1978. Definites in *there*-sentences. *Language* 300–313.

Rappaport Hovav, Malka and Beth Levin, 1998. Building verb meanings. In Miriam Butt and Wilhelm Geuder, eds., *The Projection of Arguments: Lexical and Compositional Factors*, 97–134. Stanford: CSLI Publications.

Reape, Michael, 1994. Domain union and word order variation in German. In John Nerbonne, Klaus Netter, and Carl Pollard, eds., *German in Head-driven Phrase Structure Grammar*, 151–197. Stanford, CA: CSLI Publications.

Reichenbach, Hans, 1947. *Elements of Symbolic Logic*. Berkeley: University of California Press.

Reinhart, Tanya, 1981. Definite NP anaphora and c-command domains. *Linguistic inquiry* 12:605–635.

Reinhart, Tanya, 1983a. *Anaphora and Semantic Interpretation*. London: Croom Helm.

Reinhart, Tanya, 1983b. Coreference and bound anaphora: A restatement of the anaphora questions. *Linguistics and Philosophy* 6:47–88.

Reinhart, Tanya, 1987. Specifier and operator binding. In Eric Reuland and Alice ter Meulen, eds., *The Representation of (In)definiteness*, 130–167. Cambridge, MA: MIT Press.

Reinhart, Tanya, 2006. *Interface Strategies: Optimal and Costly Computations*. Linguistic Inquiry Monograph 45. Cambridge, MA: MIT Press.

Reinhart, Tanya and Eric Reuland, 1991. Anaphors and logophors: An argument structure perspective. In Jan Koster and Eric Reuland, eds., *Long Distance Anaphora*, 283–322. Cambridge: Cambridge University Press.

Richards, Norvin, 2000. An island effect in Japanese. *Journal of East Asian Linguistics* 9:187–205.

Richards, Norvin, 2001. *Movement in Language: Interactions and Architectures*. Oxford: Oxford University Press.

Richards, Norvin, 2002. Very local A′ movement in a root-first derivation. In Samuel Epstein and Daniel Seely, eds., *Derivation and Explanation in the Minimalist Program*, 227–248. Blackwell Publishing.

Richards, Norvin, 2010. *Uttering Trees*. No. 56 in Linguistic Inquiry Monograph. Cambridge, MA: MIT Press.

Richards, Norvin, 2016. *Contiguity Theory*. No. 73 in Linguistic Inquiry Monograph. Cambridge, MA: MIT Press.

Rizzi, Luigi, 1982. *Issues in Italian Syntax*. Dordrecht: Foris.

Rizzi, Luigi, 1990a. On the anaphor-agreement effect. *Rivista di Linguistica* 2:27–42.

Rizzi, Luigi, 1990b. *Relativized Minimality*. Cambridge, MA: MIT Press.

Roberts, Craige, 1991. Distributivity and reciprocal distributivity. In *Proceedings of the 1st Semantics and Linguistic Theory Conference*, 209–229. Ithaca: Cornell University Linguistics Publications.

Roberts, Ian, 2001. Head movement. In Mark Baltin and Chris Collins, eds., *Handbook of Contemporary Syntactic Theory*, 113–147. Oxford: Blackwell.

Roberts, Ian, 2005. *Principles and Parameters in a VSO Language: A Case Study in Welsh*. Oxford: Oxford University Press.

Romero, Maribel, 1998. *Focus and Reconstruction Effects in* wh-*phrases*. Ph.D. thesis, University of Massachusetts Amherst.

Rooth, Mats, 1985. *Association with Focus*. Ph.D. thesis, University of Massachusetts, Amherst.

Rooth, Mats, 1992. A theory of focus interpretation. *Natural Language Semantics* 1:75–116.

Rosch, Eleanor, 1974. Linguistic relativity. In A. Silverstein, ed., *Human communication: Theoretical explorations*, 95–121. Erlbaum Hillsdale, NJ.

Rosenbaum, Peter, 1967. *The Grammar of English Predicate Complement Constructions*. Cambridge, MA: MIT Press.

Rosenblatt, Frank, 1962. *Principles of Neurodynamics*. New York: Spartan Books.

Ross, John Robert, 1967. *Constraints on Variables in Syntax*. Ph.D. thesis, MIT. Published as Ross 1986.

Ross, John Robert, 1969. Guess who? In *Proceedings of the 5th Regional Meeting of the Chicago Linguistic Society*, 252–286. CLS.

Ross, John Robert, 1970. Gapping and the order of constituents. In Manfred Bierwisch and Karl Heidolph, eds., *Progress in Linguistics*, 249–259. The Hague: Mouton.

Ross, John Robert, 1986. *Infinite Syntax!* Norton, NJ: Ablex.

Rousseau, Jean Jacques, 1781. Essai sur l'origine des langues. In *Œuvres complètes de J.J. Rousseau*, vol. 8eme, 357–434. Geneva, 1782.

Roy, Deb, 2005. Semiotic schemas: A framework for grounding language in action and perception. *Artificial Intelligence* 167:170–205.

Ruangrajitpakorn, Taneth, Kanokorn Trakultaweekoon, and Thepchai Supnithi, 2009. A syntactic resource for Thai: CG treebank. In *Proceedings of the 7th Workshop on Asian Language Resources*, 96–102. Suntec, Singapore: ACL.

Ruppenhofer, Josef and Laura Michaelis, 2010. A constructional account of genre-based argument omissions. *Constructions and Frames* 2:158–184.

Sabbagh, Joseph, 2007. Ordering and linearizing rightward movement. *Natural Language & Linguistic Theory* 25:349–401.

Saffran, Jenny, Richard Aslin, and Elissa Newport, 1996. Statistical learning by 8-month-old infants. *Science* 274:1926–1928.

Sag, Ivan and Jorge Hankamer, 1984. Towards a theory of anaphoric processing. *Linguistics and Philosophy* 7:325–345.

Sagae, Kenji, Brian MacWhinney, and Alon Lavie, 2004. Adding syntactic annotations to transcripts of parent-child dialogs. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*. Lisbon, LREC.

Saito, Mamoru, 1992. Long distance scrambling in Japanese. *Journal of East Asian Linguistics* 1:69–118.

Sakas, William and Janet Dean Fodor, 2001. The structural triggers learner. In S. Bertolo, ed., *Language Acquisition and Learnability*, 172–233. Cambridge: Cambridge University Press.

Sarkar, Anoop and Aravind Joshi, 1996. Coordination in tree-adjoining grammars: Formalization and implementation. In *Proceeedings of the International Conference on Computational Linguistics*, 610–615. Copenhagen: ACL.

Sato, Maso-Aki, 2001. Online model selection based on the variational bayes. *Neural Computation* 13(7):1649–1681.

Satta, Giorgio, 1992. Recognition of linear context-free rewriting systems. In *Proceedings of the 30th annual meeting on Association for Computational Linguistics*, 89–95. Association for Computational Linguistics.

Sauerland, Uli, 1998. *The Meaning of Chains*. Ph.D. thesis, MIT, Cambridge, MA.

Savitch, Walter, 1987. Context-sensitive grammar and natural language syntax. In Walter Savitch, Emmon Bach, William Marsh, and Gila Safran-Naveh, eds., *The Formal Complexity of Natural Language*, 358–368. Dordrecht: Reidel.

Schank, Roger, 1972. Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology* 3:552–631.

Schein, Barry, 2017. *'And': Conjunction Reduction Redux*. Cambridge, MA: MIT Press.

Schlenker, Philippe, 2006. Scopal independence: A note on branching and wide scope readings of indefinites and disjunctions. *Journal of Semantics* 23:281–314.

Schlesinger, Izchak, 1971. Production of utterances and language acquisition. In Dan Slobin, ed., *The Ontogenesis of Grammar*, 63–101. New York: Academic Press.

Schönfinkel, Moses, 1924. On the building blocks of mathematical logic. In Jean van Heijenoort, ed., *From Frege to Gödel*, 355–366. Cambridge, MA: Harvard University Press.

Schreiber, Peter, 1978. *There*-insertion and number agreement. *Linguistic Inquiry* 9:318–325.

Schwarzschild, Roger, 1999. GIVENness, AvoidF, and other constraints on the placement of accent. *Natural Language Semantics* 7:141–177.

Schweiger, Fritz, 2000. Compound case markers in Australian languages. *Oceanic Linguistics* 39:256–284.

Seki, Hiroyuki, Takashi Matsumura, Mamoru Fujii, and Tadao Kasami, 1991. On multiple context-free grammars. *Theoretical Computer Science* 88:191–229.

Selkirk, Elisabeth, 1972. *The Phrase Phonology of English and French*. Ph.D. thesis, MIT. Published by Garland, New York, 1980.

Selkirk, Elisabeth, 1984. *Phonology and Syntax*. Cambridge, MA: MIT Press.

Selkirk, Elisabeth, 1986. Derived domains in sentence phonology. *Phonology* 3:371–405.

Selkirk, Elisabeth, 1990. On the nature of prosodic constituency. *Papers in Laboratory Phonology* 1:179–200.

Selkirk, Elisabeth, 1995. Sentence prosody: Intonation, stress, and phrasing. In John Goldsmith, ed., *The Handbook of Phonological Theory*, 550–569. Oxford: Blackwell.

Selkirk, Elisabeth, 2011. The syntax-phonology interface. In John Goldsmith, Jason Riggle, and Alan Yu, eds., *The Handbook of Phonological Theory*, 435–483. Blackwell, 2nd edn.

Seuren, Pieter, 1985. *Discourse Semantics*. Oxford: Blackwell.

Shakespeare, William, c.1592. The taming of the shrew. Act 1, Scene 2.

Shattuck-Hufnagel, Stefanie and Alice Turk, 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25:193–247.

Shieber, Stuart, 1985. Evidence against the context-freeness of natural language. *Linguistics and Philosophy* 8:333–343.

Simpson, Jane, 1991. *Warlpiri Morpho-syntax: A Lexicalist Approach*. Dordrecht: Kluwer.

Singhal, Amit, 2012. Introducing the knowledge graph: Things, not strings. *Official Google Blog* .

Siskind, Jeffrey, 1996. A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition* 61:39–91.

Smith, Neil and Annabel Cormack, 2015. Features from *aspects* via the minimalist program to combinatory categorial grammar. In Ángel Gallego and Dennis Ott, eds., *50 Years Later: Reflections on Chomsky's* Aspects, 233–247. MIT Working Papers in Linguistics.

Smolensky, Paul, 1990. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence* 46:159–216. Reprinted in Hinton (1990).

Smullyan, Raymond, 1985. *To Mock a Mockingbird*. New York: Knopf.

Smullyan, Raymond, 1994. *Diagonalization and Self-Reference*. Oxford: Clarendon Press.

Sobin, Nicholas, 1987. The variable status of comp-trace phenomena. *Natural Language & Linguistic Theory* 5:33–60.

Socher, Richard, John Bauer, Christopher Manning, and Andrew Ng, 2013. Parsing with compositional vector grammars. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 455–465. Sofia, Bulgaria: Association for Computational Linguistics.

Stabler, Edward, 2006. Sidewards without copying. In *Proceedings of the 11th conference on Formal Grammar*, 157–170. CSLI Publications.

Stabler, Edward, 2011. Computational perspectives on minimalism. In Cedric Boeckx, ed., *Oxford Handbook of Linguistic Minimalism*, 617–641. Oxford University Press.

Stalnaker, Robert, 1978. Assertion. In Peter Cole, ed., *Syntax and Semantics IX*, 315–332. New York: Academic Press.

Stalnaker, Robert, 2002. Common ground. *Linguistics and Philosophy* 25:701–721.

Steedman, Mark, 1985. Dependency and coordination in the grammar of Dutch and English. *Language* 61:523–568.

Steedman, Mark, 1985/1988. Combinators and grammars. In Richard Oehrle, Emmon Bach, and Deirdre Wheeler, eds., *Categorial Grammars and Natural Language Structures*, 417–442. Dordrecht: Reidel.

Steedman, Mark, 1987. Combinatory grammars and parasitic gaps. *Natural Language and Linguistic Theory* 5:403–439.

Steedman, Mark, 1990. Gapping as constituent coordination. *Linguistics and Philosophy* 13:207–263.

Steedman, Mark, 1991. Structure and intonation. *Language* 67:262–296.

Steedman, Mark, 1996. *Surface Structure and Interpretation*. Linguistic Inquiry Monograph 30. Cambridge, MA: MIT Press.

Steedman, Mark, 1997. Temporality. In Johan van Benthem and Alice ter Meulen, eds., *Handbook of Logic and Language*, 895–938. Amsterdam: North Holland/Elsevier.

Steedman, Mark, 1999. Connectionist sentence processing in perspective. *Cognitive Science* 23:615–634.

Steedman, Mark, 2000a. Information structure and the syntax-phonology interface. *Linguistic Inquiry* 34:649–689.

Steedman, Mark, 2000b. *The Syntactic Process*. Cambridge, MA: MIT Press.

Steedman, Mark, 2002. Plans, affordances, and combinatory grammar. *Linguistics and Philosophy* 25:723–753.

Steedman, Mark, 2005. Interfaces and the grammar. In *Proceedings of the 24th West Coast Conference on Formal Linguistics, Vancouver, March 2005*, 19–33. Somerville, MA: Cascadilla.

Steedman, Mark, 2006. Semantics in Categorial Grammar. In Keith Brown, ed., *Encyclopedia of Language and Linguistics*, vol. 11, 143–152. Oxford: Elsevier, 2nd ed. edn.

Steedman, Mark, 2009. Foundations of universal grammar in planned action. In Morten Christiansen, Chris Collins, and Shimon Edelman, eds., *Language Universals*, 174–199. Oxford: Oxford University Press.

Steedman, Mark, 2012. *Taking Scope: The Natural Semantics of Quantifiers*. Cambridge, MA: MIT Press.

Steedman, Mark, 2014. The surface-compositional semantics of English intonation. *Language* 90:2–57.

Steedman, Mark, 2017. Gapping and information structure. *submitted* 1–46.

Steedman, Mark, 2019. Combinatory Categorial Grammar. In Andraá Kertész, Edith Moravcsik, and Csilla Rákosi, eds., *Current Approaches to Syntax - A Comparative Handbook*, 389–420. Berlin: de Gruyter Mouton.

Steedman, Mark, 2020. A formal universal of natural language grammar. *Language* 96:618–660.

Steedman, Mark and Jason Baldridge, 2011. Combinatory Categorial Grammar. In Robert Borsley and Kersti Börjars, eds., *Non-Transformational Syntax: A Guide to Current Models*, 181–224. Oxford: Blackwell.

Steedman, Mark and Cem Bozşahin, 2016. Projecting from the lexicon. *Submitted* 1–56.

Steedman, Mark and Ron Petrick, 2007. Planning dialog actions. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, 265–272. Antwerp, Sept., ACL.

Steele, Susan, 1990. *Agreement and Anti-Agreement: A Syntax of Luiseño*. Dordrecht: Reidel.

Steels, Luc, 2004. Constructivist development of grounded construction grammars. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, 9–14. Barcelona.

Steels, Luc and Jean Christophe Baillie, 2003. Shared grounding of event descriptions by autonomous robots. *Robotics and Autonomous Systems* 43:163–173.

Stockwell, Robert, Paul Schacter, and Barbara Partee, 1973. *The Major Syntactic Structures of English*. Holt, Rinehart, and Winston.

Stroik, Thomas, 2009. *Locality in Minimalist Syntax*. MIT Press.

Stromswold, Karin, 1995. The acquisition of subject and object *wh*-questions. *Language Acquisition* 4:5–48.

Sussman, Gerald and Terry Winograd, 1970. *Micro-Planner Reference Manual*. Tech. Rep. AIM-203, MIT AI Lab.

Svenonius, Peter, 2007. 1…3–2. In Gillian Ramchand and Charles Reiss, eds., *The Oxford Handbook of Linguistic Interfaces*, 239–287. Oxford University Press.

Szabolcsi, Anna, 1983. ECP in categorial grammar. Ms., Max Planck Institute, Nijmegen.

Szabolcsi, Anna, 1983/1992. Combinatory grammar and projection from the lexicon. In Ivan Sag and Anna Szabolcsi, eds., *Lexical Matters*, 241–268. Stanford, CA: CSLI Publications.

Szabolcsi, Anna, 1989. Bound variables in syntax: Are there any? In Renate Bartsch, Johan van Benthem, and Peter van Emde Boas, eds., *Semantics and Contextual Expression*, 295–318. Dordrecht: Foris.

Szabolcsi, Anna, 2007. Strong vs. weak islands. In Martin Everaert and Henk van Riemsdijk, eds., *The Blackwell Companion to Syntax*, 479–531. Wiley-Blackwell.

Szabolcsi, Anna, 2010. *Quantification*. Cambridge: Cambridge University Press.

Szendrői, Kriszta, 2001. *Focus and the Syntax-Phonology Interface*. Ph.D. thesis, University College London.

Szendrői, Kriszta, 2004. Focus and the interaction between syntax and pragmatics. *Lingua* 114(3):229–254.

Szendrői, Kriszta, 2017. The syntax of information structure and the PF interface. *Glossa* 2.

Takahashi, Shoichi and Sarah Hulsey, 2009. Wholesale late merger: Beyond the A/Ā distinction. *Linguistic Inquiry* 40:387–426.

Talmy, Leonard, 2000. *Towards a Cognitive Semantics*, vol. 1 and 2. Cambridge, MA: MIT Press.

Thomas, I, 1969. Perceived pitch of whispered vowels. *Journal of the Acoustical Society of America* 46:468–470.

Thomason, Richmond, ed., 1974. *Formal Philosophy: Papers of Richard Montague*. New Haven, CT: Yale University Press.

Thomason, Richmond, 1990. Accomodation, meaning, and implicature. In Philip Cohen, Jerry Morgan, and Martha Pollack, eds., *Intentions in Communication*, 325–363. Cambridge, MA: MIT Press.

Thompson, Cynthia and Raymond Mooney, 1998. Semantic lexicon acquisition for learning natural language interfaces. In *Proceedings of the 6th Workshop on Very Large Corpora, Montreal, August*, 57–65.

Thornton, Rosalind, 1990. *Adventures in Long-Distance Moving: The Acquisition of Complex* Wh-*questions*. Ph.D. thesis, University of Connecticut, Storrs, CT.

Thornton, Rosalind and Stephen Crain, 1994. Successful cyclic movement. In Ken Wexler, Teun. Hoekstra, and Bonnie Schwartz, eds., *Language Acquisition Studies in Generative Grammar*, 215–253. Amsterdam: John Benjamins.

Thornton, Rosalind and Graciela Tesan, 2007. Categorical acquisition: Parameter setting in Universal Grammar. *Biolinguistics* 1:49–98.

Thráinsson, Höskuldur, 1990. A semantic reflexive in icelandic. In Joan Maling and Annie Zaenen, eds., *Modern Icelandic Syntax*, *Syntax and Semantics*, vol. 24, 289–307. New York: Academic Press.

Thráinsson, Höskuldur, 2007. *The Syntax of Icelandic*. Cambridge: Cambridge University Press.

Tolman, Edward, 1948. Cognitive maps in rats and men. *Psychological Review* 55:189–208.

Tomasello, Michael, 1999. *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

Tomasello, Michael, 2009. *Why We Cooperate*. Cambridge, MA: MIT Press.

Torr, John, 2019. *Broad-Coverage Parsing with Minimalist Grammars*. Ph.D. thesis, University of Edinburgh.

Torr, John and Edward Stabler, 2016. Coordination in minimalist grammars: Excorporation and across the board (head) movement. In *Proceedings of the 12th International Workshop on Tree Adjoining Grammars and Related Formalisms*, 1–17. ACL.

Torr, John, Miloč Stanojević, Mark Steedman, and Shay Cohen, 2019. Wide-coverage neural a* parsing for Minimalist Grammars. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2486–2505. ACL.

Travis, Lisa, 1984. *Parameters and Effects of Word Order Variation*. Ph.D. thesis, MIT.

Trechsel, Frank, 2000. A CCG account of Tzotzil Pied Piping. *Natural Language and Linguistic Theory* 18:611–663.

Truckenbrodt, Hubert, 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30:219–255.

Truckenbrodt, Hubert, 2012. Semantics of intonation. In Claudia Maienborn, Klaus von Heusinger, and Paul Portner, eds., *Semantics: An International Handbook of Natural Language Meaning*, vol. 3, 2039–2969. Berlin: de Gruyter.

Truswell, Robert, 2007a. Extraction from adjuncts and the structure of events. *Lingua* 117:1355–1377.

Truswell, Robert, 2007b. *Locality of Wh-Movement and the Individuation of Events*. Ph.D. thesis, University College London.

Tse, Daniel and James Curran, 2010. Chinese CCGbank: Extracting CCG derivations from the Penn Chinese Treebank. In *Proceedings of the 23rd International Conference on Computational Linguistics*, 1083–1091. Beijing: Coling/ACL.

Turian, Joseph, Lev-Arie Ratinov, and Yoshua Bengio, 2010. Word representations: A simple and general method for semi-supervised learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 384–394. ACL.

Uematsu, Sumire, Takuya Matsuzaki, Hiroki Hanaoka, Yusuke Miyao, and Hideki Mima, 2013. Integrating multiple dependency corpora for inducing wide-coverage Japanese CCG resources. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1042–1051. Sofia: ACL.

UzZaman, Naushad, Hector Llorens, Leon Derczynski, James Allen, Marc Verhagen, and James Pustejovsky, 2013. SemEval-2013 task 1: TempEval-3: Evaluating time expressions, events, and temporal relations. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, 1–9. Atlanta: ACL.

Vadas, David and James Curran, 2007. Adding noun phrase structure to the penn treebank. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 240–247. Prague: ACL.

Vadas, David and James Curran, 2008. Parsing noun phrase structure with CCG. In *Proceedings of the 46th Annual Meeting of the Association of Computational Linguistics*, 335–343. Columbus, OH: ACL.

Vallduví, Enric, 1990. *The Information Component*. Ph.D. thesis, University of Pennsylvania.

Vallduví, Enric and Maria Vilkuna, 1998. On rheme and kontrast. In Peter Culicover and Louise McNally, eds., *Syntax and Semantics, Vol. 29: The Limits of Syntax*, 79–108. San Diego: Academic Press.

Van Eynde, Frank, 2007. The big mess construction. In *The Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar*, 415–433. Stanford: CSLI.

van Geenhoven, Veerle, 1998. *Semantic Incorporation and Indefinite Descriptions: Semantic and Syntactic Aspects of Noun Incorporation in West Greenlandic*. Stanford, CA: CSLI publications.

van Riemsdijk, Henk, 1978. *A Case Study in Syntactic Markedness*. Dordrecht: Foris.

Van Valin, Robert, 1993. A synopsis of role and reference grammar. In Robert Van Valin, ed., *Advances in Role and Reference Grammar*, 1–164. Amsterdam: John Benjamins.

Van Valin, Robert, ed., 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.

Vendler, Zeno, 1957. Verbs and times. *The Philosophical Review* 66:143–160. Reprinted in Vendler (1967).

Vendler, Zeno, 1967. *Linguistics in Philosophy*. Ithaca, NY: Cornell University Press.

Vijay-Shanker, K. and David Weir, 1990. Polynomial time parsing of combinatory categorial grammars. In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, 1–8. Pittsburgh: ACL.

Vijay-Shanker, K. and David Weir, 1993. Parsing some constrained grammar formalisms. *Computational Linguistics* 19:591–636.

Vijay-Shanker, K. and David Weir, 1994. The equivalence of four extensions of context-free grammar. *Mathematical Systems Theory* 27:511–546.

Vijay-Shanker, K., David Weir, and Aravind Joshi, 1987. Characterizing structural descriptions produced by various grammatical formalisms. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, 104–11. ACL.

Villavicencio, Aline, 2002. *The Acquisition of a Unification-Based Generalised Categorial Grammar*. Ph.D. thesis, University of Cambridge.

Vinyals, Oriol, Łukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, and Geoffrey Hinton, 2015. Grammar as a foreign language. In *Advances in Neural Information Processing Systems*, 2755–2763.

von Stechow, Arnim, 1996. Against LF Pied-Piping. *Natural Language Semantics* 4:57–110.

Wagner, Michael, 2005. *Prosody and Recursion*. Ph.D. thesis, MIT, Cambridge, MA.

Wagner, Michael, 2010. Prosody and recursion in coordinate structures and beyond. *Natural Language & Linguistic Theory* 28:183–237.

Wagner, Michael, 2012a. Contrastive topics decomposed. *Semantics and Pragmatics* 5:212–233.

Wagner, Michael, 2012b. Focus and givenness: A unified approach. In Ivona Kučerová and Ad Neeleman, eds., *Contrasts and Positions in Information Structure*, 102–147. Cambridge: Cambridge University Press.

Wagner, Michael, 2012c. A givenness illusion. *Language and Cognitive Processes* 27:1433–1458.

Wallenberg, Joel, 2009. *Antisymmetry and the Conservation of c-Command: Scrambling and Phrase Structure in Synchronic and Diachronic Perspective*. Ph.D. thesis, University of Pennsylvania.

Wallenberg, Joel, 2013. Scrambling, LF, and phrase structure change in Yiddish. *Lingua* 133:289–318.

Wang, Adrienne, Tom Kwiatkowski, and Luke Zettlemoyer, 2014. Morpho-syntactic lexical generalization for ccg semantic parsing. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1284–1295. Doha: ACL.

Ward, Gregory and Julia Hirschberg, 1985. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* 61:747–776.

Watanabe, Akira, 1992. Subjacency and S-structure movement of *wh*-in-situ. *Journal of East Asian Linguistics* 1:255–291.

Watanabe, Akira, 2001. *Wh*-in-situ languages. In Mark Baltin and Chris Collins, eds., *The Handbook of Contemporary Syntactic Theory*, 203–225. Oxford: Blackwell.

Wechsler, Stephen, 2015. *Word Meaning and Syntax*. Oxford University Press.

Weeds, Julie and David Weir, 2003. A general framework for distributional similarity. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, 81–88. ACL.

Weinberg, Amy, 1988. *Locality Principles in Syntax and In parsing*. Ph.D. thesis, MIT.

Weinberg, Amy, 1998. Minimalist theory of human sentence processing. In Samuel Epstein and Norbert Hornstein, eds., *Working Minimalism*. Cambridge, MA: MIT Press.

Weir, David, 1988. *Characterizing Mildly Context-Sensitive Grammar Formalisms*. Ph.D. thesis, University of Pennsylvania, Philadelphia. Published as Technical Report CIS-88-74.

Weir, David, Julie Weeds, Jeremy Reffin, and Thomas Kober, 2016. Aligning packed dependency trees: A theory of composition for distributional semantics. *Computational Linguistics* 42:727–761.

Wells, Rulon, 1947. Immediate constituents. *Language* 23:81–117.

West, Mae, 1933. I'm no angel. Paramount.

Wexler, Kenneth and Peter Culicover, 1980. *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.

White, Michael, 2004. Reining in CCG chart realization. In *Proceedings of the 3rd International Conference on Natural Language Generation*. ACL.

White, Michael, 2006. Efficient realization of coordinate structures in combinatory categorial grammar. *Research on Language and Computation* 4:39–75.

White, Michael and Jason Baldridge, 2003. Adapting chart realization to CCG. In *Proceedings of the 9th European Workshop on Natural Language Generation, held in conjunction with the 11th Conference of the European Chapter of the Association for Computational Linguistics*, 119–126. Budapest: ACL.

Whitelock, Pete, 1991. What sort of trees do we speak?: A computational model of the syntax-prosody interface in Tokyo Japanese. In *Proceedings of the Fifth Conference of the European Chapter of the Association for Computational Linguistics*, 75–82. Berlin, Germany: ACL.

Whitman, Neal, 2009. Right-node wrapping: Multimodal categorial grammar ond the "friends in low places" coordination. In Erhard Hinrichs and John Nerbonne, eds., *Theory and Evidence in Semantics*, 235–256. Stanford, CA: CSLI Publications.

Whorf, Benjamin, 1946. Grammatical categories. *Language* 16:67–72. Written in 1937.

Whorf, Benjamin, 1950. An american indian model of the universe. *International Journal of American Linguistics* 21:1–11. Written prior to 1944.

Whorf, Benjamin and John Carroll, eds., 1956. *Language Thought and Reality*. Cambridge, MA: MIT Press.

Wilder, Chris, 1999. Right node raising and the LCA. In *Proceedings of 18th WCCFL*, 586–598. Cascadilla Press Somerville MA.

Wilder, Chris, 2008. Shared constituents and linearization. In Kyle Johnson, ed., *Topics in ellipsis*, 229. Cambridge University Press.

Williams, Alexander, 2015. *Arguments in Syntax and Semantics*. Cambridge University Press.

Williams, Edwin, 1978. Across-the-board rule application. *Linguistic Inquiry* 9:31–43.

Williams, Edwin, 1981. Transformationless grammar. *Linguistic Inquiry* 645–653.

Williams, Edwin, 1984. *There*-insertion. *Linguistic inquiry* 131–153.

Williams, Edwin, 1986. A reassignment of the functions of LF. *Linguistic Inquiry* 17:265–299.

Williams, Edwin, 1990. The atb theory of parasitic gaps. *The Linguistic Review* 6:265–279.

Williams, Edwin, 1992. Adjunct control. In Richard Larson, Sabine Iatridou, Utpal Lahiri, and James Higginbotham, eds., *Control and Grammar*, 297–322. Springer.

Williams, Edwin, 1994. *Thematic Structure in Syntax*. Cambridge, MA: MIT Press.

Williams, Edwin, 2003. *Representation Theory*. MIT Press.

Winter, Yoad, 2001. *Flexibility Principles in Boolean Semantics*. Cambridge, MA: MIT Press.

Winter, Yoad, 2007. Multiple coordination: Meaning composition vs. the syntax-semantics interface. *unpublished manuscript* .

Wittenburg, Kent, 1987. Predictive combinators: A method for efficient processing of combinatory grammars. In *Proceedings of the 25th Annual Conference of the Association for Computational Linguistics*, 73–80. Stanford: ACL.

Wittgenstein, Ludwig, 1953. *Philosophische Untersuchungen (Philosophical Investigations)*. Oxford: Basil Blackwell.

Woods, William, 1978. Semantics and quantification in natural language question answering. *Advances in Computers* 17:1–87.

Wu, Huijia, Jiajun Zhang, and Chengqing Zong, 2016. A dynamic window neural network for ccg supertagging. In *Archiv*.

Wurmbrand, Susi, 2004. West Germanic verb clusters: The empirical domain. In Katalin É. Kiss and Henk van Riemsdijk, eds., *Verb clusters: A study of Hungarian, German, and Dutch*, 43–85. Amsterdam: John Benjamins.

Xu, Wenduan, 2016. Lstm shift-reduce ccg parsing. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 1754–1764. Austin, TX: ACL.

Xu, Wenduan, Michael Auli, and Stephen Clark, 2015. Ccg supertagging with a recurrent neural network. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 250–255. Beijing: ACL.

Xu, Wenduan, Stephen Clark, and Yue Zhang, 2014. Shift-reduce ccg parsing with a dependency model. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 218–227. Baltimore, MD.

Yang, Charles, 2002. *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.

Yang, Charles, 2004. Universal grammar, statistics, or both? *Trends in Cognitive Science* 8.

Yang, Charles, 2006. *The Infinite Gift*. New York: Scribner.

Yang, Charles, 2016. *The Price of Productivity*. Cambridge, MA: MIT Press.

Yoshikawa, Masashi, Hiroshi Noji, and Yuji Matsumoto, 2017. A* CCG parsing with a supertag and dependency factored model. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*.

Young, Robert and William Morgan, 1980. *The Navajo Language: A Grammar and Colloquial Dictionary*. Albuquerque, NM: University of New Mexico Press.

Zaenen, Annie and Joan Maling, 1990. Unaccusative, passive, and quirky case. In Joan Maling and Annie Zaenen, eds., *Modern Icelandic*, *Syntax and Semantics*, vol. 24, 137–152. San Diego: Academic Press.

Zanuttini, Raffaella, 1997. *Negation and Clausal Structure: A Comparative Study of Romance Languages*. Oxford University Press.

Zettlemoyer, Luke and Michael Collins, 2005. Learning to map sentences to logical form: Structured classification with Probabilistic Categorial Grammars. In *Proceedings of the 21st Conference on Uncertainty in AI (UAI)*, 658–666. Edinburgh: AAAI.

Zhang, Congle and Daniel Weld, 2013. Harvesting parallel news streams to generate paraphrases of event relations. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 1776–1786. Seattle: ACL.

Zhang, Niina, 2010. *Coordination in Syntax*. Cambridge University Press.

Zhang, Niina Ning, 2007. On the categorial issue of coordination. *Lingua et Linguistica* 1:7–46.

Zhang, Niina Ning, 2008. Repetitive and correlative coordinators as focus particles parasitic on coordinators. *SKY Journal of Linguistics* 21:295–342.

Zribi-Hertz, Anne, 1989. Anaphor binding and narrative point of view: English reflexive pronouns in sentence and discourse. *Language* 65:695–727.

Zubizarreta, Maria Luisa, 1998. *Prosody, Focus, and Word Order*. Cambridge, MA: MIT Press.

Zubizarreta, Maria Luisa and Emily Nava, 2011. Encoding discourse-based meaning: Prosody vs. syntax. implications for second language acquisition. *Lingua* 121:652–669.

Zwart, Jan-Wouter, 2005. The coordinate structure constraint: A minimalist perspective. In *Workshop on Verb Clusters and Coordination*.

Zwicky, Arnold, 1984. Welsh soft mutation and the case of object NPs. In *Proceedings of the 20th Meeting of the Chicago Linguistic Society: Papers from the General Session*, 387–402. Chicago: University of Chicago Press.

Index