Mark Steedman

14 Combinatory Categorial Grammar

Abstract: Combinatory Categorial Grammar (CCG) is a radically lexicalized theory of grammar in which all language-specific information, including the linear order of heads, arguments, and adjuncts, is specified in the lexicon, from which it is projected onto sentences by language-independent universal type-dependent combinatory rules of low "slightly non-context-free" expressive power, applying to strictly adjacent phonologically-realised categories. Syntactic and phonological derivation are isomorphic, and are synchronously coupled with semantic composition in a purely type-dependent rule-to-rule relation.

1 Overview

1.1 Goals

The central problem addressed by Combinatory Categorial Grammar (CCG) is the nature of the mapping between sound and meaning. The goal is to achieve an explanatory theory of natural language grammar that is immediately applicable to psychological and computational models of syntactic and semantic processing of spoken and written natural language, and of language acquisition by children.

1.2 Data

The data which are drawn upon in order to *define* CCG are facts generally agreed among linguists concerning long-range dependency, coordination, and prosodic structure, all of which give the appearance of displacement, or non-contiguity of elements that belong together semantically, such as governors (heads) and their complements.

The data to which the theory has been *applied* is much more various, and includes corpus data, both labeled and unlabeled, that is used to train parsers, and the various test-sets that are used to evaluate them, including corpora of child-directed utterance, and psycholinguistic data.

Acknowledgement: Thanks to the editors and reviewers of this volume for their helpful comments. Some of these ideas were developed through a class "Introduction to Combinatory Categorial Grammar" which was presented by the author at the 2017 LSA Summer Institute in Lexington KY. Thanks also to the participants for their input. The writing was supported by ERC Advanced Fellowship 742137 SEMANTAX.

https://doi.org/10.1515/9783110540253-014

1.3 Tools

Crosslinguistic similarities and differences are represented in CCG solely at the level of the lexicon, which specifies all language-specific properties including the linear order and semantic dependency of governors and dependent or complement constituents. The lexicon is projected onto the sentences of the language by "combinatory" rules—that is, by strictly string-adjacent operations, combining contiguous categories without the involvement of any form of "action at a distance", such as movement, copying, or deletion under identity.

The only representational levels in CCG are phonological and logical form. Syntactic derivation is not itself a level of representation, and is dispensible. All syntactic rules are type-dependent, rather than structure-dependent, and assemble logical and phonological form in lockstep with syntactic derivation. The hypothesis is that the degrees of freedom in the type-system of the lexicon and the combinatory rules are both necessary and sufficient for the analysis of the languages of the world. The categories are those of categorial grammar. The relations between categories are combinatory in the sense defined above, and are fully formalized.

The theory outlined in this chapter, and developed in slightly different forms and at greater length in earlier publications, has been applied to the syntactic and semantic analysis of coordination and unbounded dependency in a wide range of languages. It has also been widely applied computationally in practical natural language processing (NLP) applications, particularly those requiring that the syntax support semantic interpretation. There is a CCG-based computational account of acquisition and development, based on semantic bootstrapping of the language-specific lexicon (Abend et al., 2017). There is a hypothesis concerning the origins of the categories and combinatory rules in terms of their use for planning complex actions in human and prehuman cognition (Steedman, 2002, 2017). Neither is discussed at any length here. Wide coverage parsers for CCG have been developed.[1]

1.4 Sample analysis

The following sentence, selected by the editors for comparison across the various approaches in this volume, is quite long:

(1) After Mary introduced herself to the audience, she turned to a man that she had met before.

1 The interested reader can try out the "Easy CCG" parser (Lewis et al., 2016) by typing or pasting sentences such as (1) into the input box at http://4.easy-ccg.appspot.com/, bearing in mind that this is a probabilistic parser, with a lexicon and parsing model primarily trained on the Penn WSJ treebank.

Accordingly, its derivation is presented in Figure 1 in three steps, with all discussion of semantics and logical form deferred until the detailed discussion of the constructions involved.

First, the preposed adjunct *After Mary introduced herself to the audience* is derived syntactically as in Figure 1a. CCG derivations like this are written in the acceptance direction, with the words at the top and the "start symbol" (usually, S) at the bottom, but are otherwise equivalent to standard derivational phrase-structure trees. Slashes / and \ define the English transitive verb as looking for its first NP (object) argument to the right and its second (subject) argument to the left. Underlines indicate combination, and the directional arrows > and < indicate that the rule involved is forward (rightward) or backward (leftward) application. The ↑ notation indicates that the category in question such as $NP^{\uparrow}$ has a type-raised or cased category such as the *nominative* category $S/(S\backslash NP)$, abbreviated here for readability. Since the derivation shown is entirely applicative, type-raising has no effect here other than to reverse the directionality of the rule that combines verb and argument, so can be temporarily ignored. The binding of the reflexive anaphor "herself" is also lexicalized, via the logical form (not shown), whose details are discussed in Section 3.5.

The main clause involves an unbounded relativized dependency, and is more complicated syntactically, making crucial use of composition and type-raising, as in Figure 1b.

This derivation crucially involves composition rules, indexed $>\mathbf{B}$ and $>\mathbf{B}_{\times}$. Their operation, whose details are discussed in Section 4, crucially depends on the arguments being type-raised. In particular, the subject "she" of the relative clause must bear the *nominative* raised category for the derivation to go through, although for the purposes of this overview, we continue to abbreviate it as $NP^{\uparrow}$. (Thus, English is highly ambiguous as to case, unlike morphologically cased languages like Latin and Japanese.)

To complete the derivation, the sentential adjunct derived in 1a combines with the sentence derived in 1b by simple forward application, yielding a sentence, as in Figure 1c.

Although the assembly of logical form is not shown in this introductory analysis, its derivation is entirely compositional and homomorphic to the surface syntactic derivations shown. In particular, the logical form corresponding to the complex noun "man that she had met before" is under the analysis of relativization developed in Section 4 itself a property of type N with the appropriate logical form $\lambda n \lambda x.past\ (perfect\ (meet\ x\ pro_{she})) \land n\ x$ (cf. (25)). Such details are discussed at length in the body of the chapter.

Anaphoric relations, including the binding of the pronoun "she" in the main clause to "Mary" in the adjunct, and the binding of the past tense of the main clause "turned" in an "after' relation to the antecedent reference time of the adjunct "introduced", among others, are not treated in CCG as falling in the domain of sentence grammar proper.

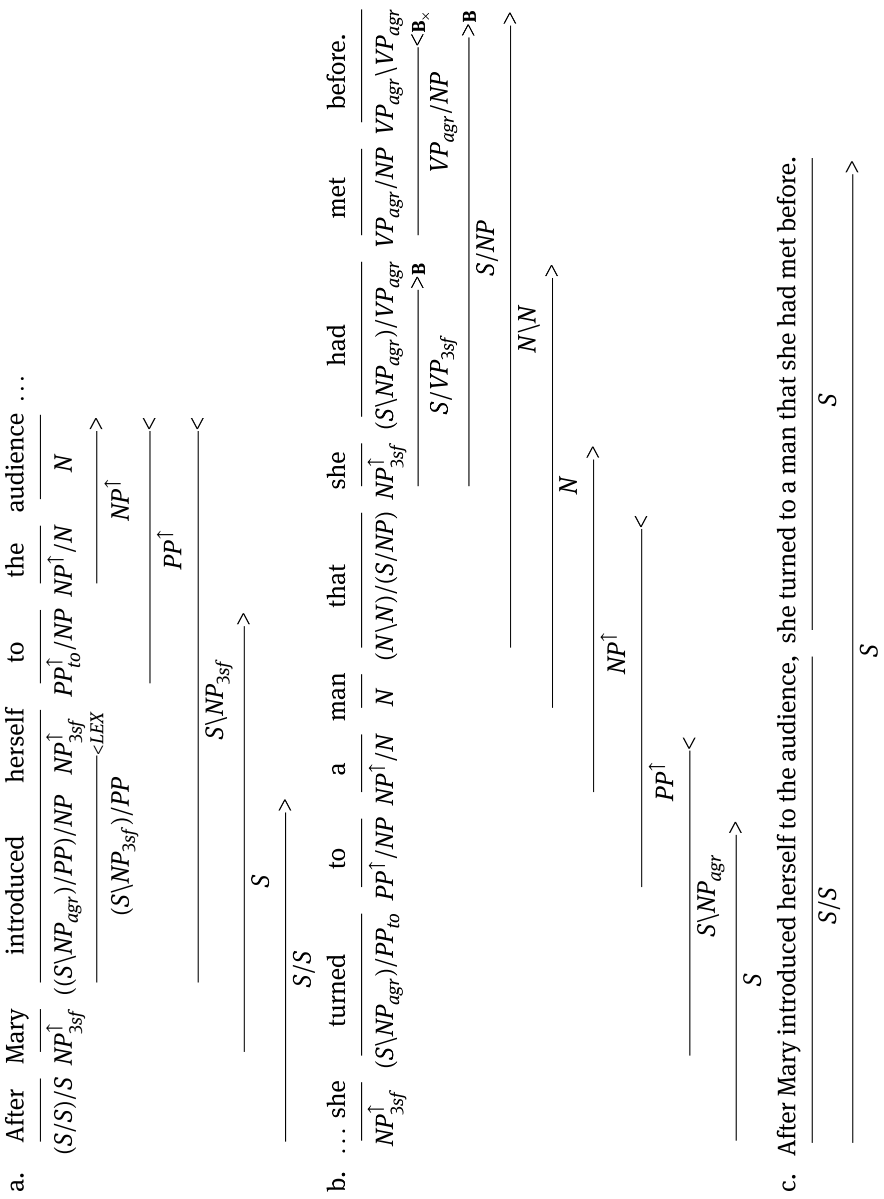

Figure 1: Sample analysis.

1.5 Evaluation criteria

The evaluation criteria for comparing CCG with alternative approaches are descriptive and explanatory adequacy, and applicability to practical computational natural language processing, including the building of logical form.

Descriptive adequacy is attained by capturing all the phenomena of a system. Explanatory adequacy is attained by capturing *only* those phenomena, and being unable to capture other comparable phenomena that are not exhibited by the system. A theory which can express phenomena that we believe can never occur is overly-expressive and less than explanatory (although such theories may be extremely useful in laying out the phenomena in ways that help us to find our way to more explanatory ones).[2]

Since in the case of the grammatical system we only have a fairly small sample of languages to work with, we don't have complete knowledge of the set of possible phenomena. It follows that any claim to explanatory adequacy in the theory of grammar is a hostage to fortune, and can be disproved by the discovery of languages that controvert its prediction of their non-existence.

Nevertheless, the available descriptive accounts make grammar seem relatively systematic. For example, we shall see in Section 7 that CCG predicts that two of the 24 possible ways of linearising the four elements corresponding to the English words comprising the noun-phrase "these five fat cats" are impossible, and will never be found in any natural language (cf. Greenberg, 1963; Cinque, 2005). A theory that is descriptively adequate in other respects, and also accurately predicts the same generalization concerning word orders in other constructions is empirically falsifiable, and therefore more explanatorily adequate than one that does not.[3]

2 Historical background to CCG

When syntactic theory as defined in Chomsky, 1965 (hereafter, *Aspects*) fragmented in the '70s and '80s, leading to the profusion of approaches assembled in the present volume, attempts to develop alternatives to *Aspects*-style transformational rules took two forms.

2 While initially acknowledging something like the above as a definition of explanatory adequacy, Chomsky (1965) proposes the provision of a theory of child language acquisition as a proxy for explanatory adequacy. However, once we have admitted that some amount of innate knowledge must be available for language acquisition to be possible at all, then *every* theory of language has a theory of acquisition if we assume that its key assumptions and constraints are by some evolutionary process innate. In order to avoid merely pushing the burden of explanation off onto the theory of evolution, an explanatory theory in the first sense is actually a prerequisite for an explanatory theory of child language acquisition.

3 For reasons that have nothing to do with the theory of grammar, all such variation is Zipfian in distribution, with the long tail of less common orders becoming double-exponentially rarer, so that it is in practice hard to know whether an unseen word order is truly impossible, or just so rare that we haven't seen it yet.

One group of constraint-based theories, usually expressed in unification-based formalisms, were as overly expressive as *Aspects* transformational grammars, but were easier to implement computationally (and therefore to automatically check for over- and under-generalization). They included Augmented Transition Network Grammar (ATNG. Woods, 1970), Functional Unification Grammar (FUG, Kay, 1984), Lexical Functional Grammar (LFG, Bresnan, 1982), Dependency/Word Grammar (DG/WG, Hays, 1964; Hudson, 1984), Head-driven Phrase Structure Grammar (HPSG, Pollard and Sag, 1987), Autolexical Grammar (ALG, Sadock, 1991), Role and Reference Grammar (R&RG, Van Valin, 1993), Sign-based Construction Grammar (SBCxG, Boas and Sag, 2012), Simpler Syntax (SS, Jackendoff, 1997, 2002; Culicover and Jackendoff, 2005), Type-Logical Grammar (TLG, Moortgat, 1988; Hepple, 1990; Dowty, 1993; Morrill, 1994; Jacobson, 1999), and some versions of Montague Grammar (MG, Montague, 1973; Bach, 1976; Cooper, 1983).

A second group sought for formalisms that were much less expressive in the first place, such as Generalized Phrase Structure Grammar (GPSG, Gazdar, 1981), Combinatory Categorial Grammar (CCG, Ades and Steedman, 1982; Szabolcsi, 1989), Head Grammar (HG, Pollard, 1984), and Tree Adjoining Grammar (TAG Joshi, 1988).

In the same period, the transformational theory itself evolved through the Extended Standard Theory (EST, Chomsky, 1972), the Revised Extended Standard Theory (REST, Chomsky and Lasnik, 1977), and Principles and Parameters/Government-Binding (P&P/GB, Chomsky, 1981), until the more radical reform of the current Minimalist Program (MP, Chomsky, 1995b,a, 2000), defined by the assumption that syntactic derivation is determined by its function of creating objects that are phonologically and semantically well-formed, an assumption that (at least in aspiration) makes it more akin to the latter approaches, and in particular, the present approach of Combinatory Categorial Grammar.

In particular, both Minimalism and CCG are committed to the view that syntactic derivation works by language-independent principles, from which it follows that all language-specific properties of constructions must derive from the lexicon of the language, not via language-specific rules or constraints.

In that sense, as Adger (2013) has pointed out, Chomskian Minimalism can be seen as a form of Categorial Grammar that adds Movement as the mechanism for handling discontinuous dependency, rather than the combinatory rules of CCG that are defined below. One of the purposes of the present chapter is to compare and contrast the movement theory of discontinuity with the alternative combinatory extension of Categorial Grammar proposed by CCG, as well as with the other alternative theories noted above.

In order to make these comparisons, the presentation of the formal specifics of CCG will be tied to the constructions in English and other languages that motivate their introduction. Our rules are empirically motivated, and it is only in the latter sections of the chapter that we turn to the question of why they take the form that they do.

The constructions in question fall into two groups. The first group comprises the "bounded" constructions like raising, control, passive, unaccusative, intransitivization, the binding of reflexive pronouns, etc., which concern relations or alternations in relations among the arguments of a single head, such as a verb, together with related matters like agreement and case. These are phenomena on which all theories more or less agree, differing only in the degree of lexicalization vs. syntacticalization that they assume, with CCG occupying the radically lexicalized end of the spectrum. These are dealt with fairly briefly in Section 3.[4]

The remaining sections of the paper concern a much more problematic range of constructions which we will loosely refer to as "unbounded", which constitute a much more difficult problem for the theory of grammar, and for which CCG presents a radically different analysis from other theories. They include relativization and its allies such as topicalization and *wh*-question formation, together with its subspecies such as "pied-piping" and "parasitic" extractions, all of which have been attributed to unbounded movement, and are the subject of Section 4, which introduces all the remaining syntactic operations, and shows that the *wh*-constructions can be analysed without movement. Section 5 then goes on to show that various forms of coordination reduction, which have elsewhere been attributed to deletion under identity, copying, or parallelism, can be eliminated under exactly the same assumptions as movement.

Section 6 then briefly reviews the notion of constituency that is implicit in CCG, and notes that English intonation structure and its semantics reflect exactly the same notion of derivational constituency as the earlier constructions, without the stipulation of extra-syntactic features such as "edges" and non-syntactic processes such as "Focus Projection".

Section 7 is more technical, and addresses the question of the degrees of freedom that have been exercised in achieving this account, and the explanatory adequacy of the theory that results. A brief conclusion then sums up.

3 Pure Categorial Grammar (CG)

The pure Combinatory Grammar of Ajdukiewicz and Bar-Hillel eschews language-specific syntactic production rules like (2) for English.

(2)
$$\begin{array}{lll} S & \rightarrow & NP\ \ VP \\ VP & \rightarrow & TV\ \ NP \\ TV & \rightarrow & \{proved,\ found,\ sees,\ \ldots\} \end{array}$$

4 Many of the constructions of central concern to Construction Grammarians arguably belong in this class of lexically-governed constructions.

Instead, the same language-specific syntactic information is lexicalized, via lexical entries like (3) for the English transitive verb:

(3) sees := $(S\backslash NP_{3s})/NP$

This "category" identifies the transitive verb as a function or governing category, specifying the type, directionality, and agreement of its NP arguments and the type of its result, S. Thus, it specifies "sees" as a transitive verb wanting an NP on the right (with unspecified agreement) as its first argument, to yield a function wanting an NP bearing third singular agreement on the left, to yield S.[5]

The lexical notation for Chomskean Minimalism is essentially categorial (Chomsky, 1995a, 2000; Stabler, 2011; Adger, 2013):

(4) sees :: { =D+case, =D V} ("yields V; selects two D (NPs); assigns case to the first")

(The above is Chomsky's notation, which omits directional alignment, like a categorial grammar with non-directional slashes $|\ X$, which loses CCG's transparency to language-specific linear order of governors and arguments. Stabler also discusses a Directional Minimalist Grammar (DMG) with =X and X= directionality, parallel to $/X$ and $\backslash X$.)

3.1 The categorial lexicon

When stated in full, categories also specify a semantics or logical form, as in (5a), as anatomized in (5b), in which the separator ":=" pairs a phonological/graphological form with a category, and the separator ":" pairs a syntactic type with a logical form:

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

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

The predicate-argument structure component of logical form is assumed to be cross-linguistically universal, although elements like *sees* are of course a proxy for more complex structures involving tense, aspect, etc. Predicate-argument structure is therefore essentially equivalent to the lexical component of thematic structure in Minimalism, f-structure in LFG, ARG-ST in HPSG, the grammatical function tier of SS, and dependency structures in DG/WG, although unlike some of these formalisms, CCG

5 3s agreement is of course specified by -s morpho-phonology by a process discussed later in this section, and in Section 7. We assume a standard mechanism of simple non-recursive feature-value unification.

does not include explicit rôle-labels, or define these structures as linearly ordered or aligned.

The use of an explicit representational level of linguistic form, distinct from syntactic derivation, is a point of difference from the TLG tradition of Lambek in Categorial Grammar. While TLG often presents interpretations as λ-terms, they are included purely for ease of reading, and are proclaimed to be dispensible. The semantics itself is defined by "direct surface composition" on the syntactic derivation itself (Jacobson, 1999—see Bozşahin, 2012, 87–106 for discussion).

While direct surface compositionality is technically possible for CCG, there is a good reason to include a representational level of logical form. Since the only plausible account of child language acquisition is that it is semantically bootstrapped from a prior representation of meaning, and since that meaning representation must be independent of the surface syntax of any specific language, it must be a language-independent logical form. It is therefore syntactic derivation, rather than LF, that is dispensible as a level of representation.

To take a slightly more complex lexical verb, the following is the category for a subject control verb for a sentence such as *He promises her to leave*, again anatomized as (b):

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

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

The control relation between the subject and the infinitival VP complement is entirely captured at the level of the logical form, via the variables y and p. Here, the minimalist notation would be slightly different, treating the infinitival complement as a "small" clause, and treating the relation to the surface syntactic subject as mediated by movement or an anaphoric element. To that extent, CCG can be seen as lexicalizing the A-movement analysis *statically*, via the use of bound variables at the level of logical form. Because such lexicalization is by definition limited to relations between co-arguments of the verb, via the logical form, it necessarily obeys minimality conditions variously identified as "subjacency", "relativized minimality", "the Minimal Link Condition", etc., as do GPSG/HPSG/LFG also.

In CCG, all bounded constructions, such as passive, reflexivization, raising, and control, are lexically governed in a similar way, so that all instances of so-called A-movement are specified statically in the lexicon via the logical form of the governor.[6]

6 It follows that the phenomenon of so-called Backward Control, in which an explicit subject in an infinitival complement clause appears to bind an implicit argument in the matrix clause, as has been proposed for Tsez by Polinsky and Potsdam (2000), cannot be handled as control in CCG.

As in any theory of grammar, rather than listing every single lexical category in its own right, we may want to capture "parametric" generalizations across the lexicon for any given language via "lexical redundancy rules" Jackendoff (1975), such as that not only "sees", but every transitive tensed main verb has an SVO syntactic type, or that some identifiable class of the same transitive verbs including "eat" but not "find" map systematically onto a corresponding class of intransitivized verbs. Some, like intransitivization itself, will be morphologically unmarked, others like passivization, marked. Such regularities will allow the language learner to infer the existence of other members of such paradigms when they first encounter a novel verb, and possibly to learn more rapidly. They also have the advantage of allowing the grammar to be represented more compactly, although the fact that all of these generalizations are liable to admit of exceptions or irregularities shows that this is not the only consideration, and it may well be the case that such paradigms are compiled out into multiple lexical entries in their own right, as soon as encountered. All such solutions are formally equivalent, and the present chapter remains agnostic as to which should be preferred.

The present chapter maintains an open mind about exactly how those lexical generalizations should be captured. Give or take the notational idiosyncrasies applying to the lexicon, and the use in addition in some of the other theories discussed in this volume of language-specific syntactic rules, all theories are pretty much equivalent in this respect, and can be applied to the categorial lexicon. Accordingly, we will pass them over in this chapter, to concentrate on the syntactic component, which is more distinctive to CCG.

3.2 Syntactic rules I: Pure application

Verb categories like (5) and (6) combine most simply by the following rules, which are universal and language-independent:

(7) a. Forward Application:

$$X/_{\star}Y \quad Y \quad \Rightarrow \quad X \qquad (>)$$

b. Backward Application:

$$Y \quad X\backslash_{\star}Y \quad \Rightarrow \quad X \qquad (<)$$

X and Y can be any syntactic CCG type, and may include simple feature-value pairs, such as agreement. $\Rightarrow$ reads as "the things on the left combine to yield the thing on the right", and is entirely analogous to the reverse of the rewrite arrow of the PS rules

Cormack and Smith (2002) show that the construction in Tsez is limited to just two verbs, meaning "begin" and "continue", and that the supposedly controlling infinitival complement subject cannot be referential, as with a universally quantified NP. Both restrictions are reminiscent of *there* insertion in English, which is limited to raising verbs and non-referential subjects, so it seems reasonable to assume that so-called backward control arises from some form of lexicalized equivalent of expletive insertion at the level of lexical logical form.

in (2). > and < are annotations in derivations indicating the application of the rule in question. The slash-type $\star$ on the functor categories in (8) means that any functor category of the form X/Y or $X\backslash Y$ can combine by these rules.

Such rules, like all rules in CCG, correspond to the Minimalist operation of (external) **merge**, with the effect of "canceling" Y term, as if they were rules of fractional multiplication. They are analogous to Minimalist "feature-checking" between argument and governor or head (Adger, 2003, 90–96).

Such rules are rules of semantic merger, as well as syntactic. Thus we can extend them as follows, with ":" again acting as an uninterpreted separator between syntactic type and semantic interpretation or logical form:

(8) a. Forward Application:

$$X/_{\star}Y : f \quad Y : a \quad \Rightarrow \quad X : f\,a \qquad (>)$$

b. Backward Application:

$$Y : a \quad X\backslash_{\star}Y : f \quad \Rightarrow \quad X : f\,a \qquad (<)$$

Such rules are both syntactically and semantically rules of functional application, or combination of functions with their arguments, as in the following derivation, in which syntactic derivation and the composition of logical form are synchronous and homomorphic.

(9)

$$
\begin{array}{cccl}
\text{Harry} & \text{sees} & \text{Sally} & \\
\cline{1-1}\cline{2-2}\cline{3-3}
NP_{3s} & (S\backslash NP_{3s})/NP & NP & \\
: harry & : \lambda x\lambda y.sees\,x\,y & : sally & \\
\cline{2-3}
 & \multicolumn{2}{c}{S\backslash NP_{3s}} & > \\
 & \multicolumn{2}{c}{: \lambda y.sees\,sally\,y} & \\
\cline{1-3}
\multicolumn{3}{c}{S} & < \\
\multicolumn{3}{c}{: sees\,sally\,harry} &
\end{array}
$$

(The absence of any explicit slash type on the categories in the derivation (9) means that those categories can combine by any rule, including some discussed below that are more restricted than the application rules.)

Since categories like (6) achieve the effect of "A-movement" via λ-binding at the level of logical form, the application rules are all that is needed to capture the phenomenon of control, as in the following derivation:

(10)

$$
\begin{array}{cccccl}
\text{Harry} & \text{promises} & \text{Sally} & \text{to} & \text{leave} & \\
\cline{1-1}\cline{2-2}\cline{3-3}\cline{4-4}\cline{5-5}
NP_{3sm} & ((S\backslash NP_{3s})/VP_{to})/NP & NP_{3sf} & VP_{to}/VP & VP & \\
: harry & : \lambda x\lambda p\lambda y.promises\,(p\,y)\,x\,y & : sally & : \lambda p.p & : \lambda y.leave\,y & \\
\cline{2-3}\cline{4-5}
 & \multicolumn{2}{c}{S\backslash NP_{3s} \quad >} & \multicolumn{2}{c}{VP_{to} \quad >} & \\
 & \multicolumn{2}{c}{: \lambda p\lambda y.promises\,(p\,y)\,sally\,y} & \multicolumn{2}{c}{: \lambda y.leave\,y} & \\
\cline{2-5}
 & \multicolumn{4}{c}{S\backslash NP : \lambda y.promises\,(leave\,y)\,sally\,y} & > \\
\cline{1-5}
\multicolumn{5}{c}{S : promises\,(leave\,harry)\,sally\,harry} & <
\end{array}
$$

Such purely applicative derivations will correctly form "chains" of raising and control relations in examples like the following, which are left as exercises:

(11) a. Harry promises Sally to persuade Alice to leave.
b. Harry seems to promise Sally to leave.
c. Harry wants to try to begin to write a play.

3.3 The Combinatory Projection Principle

The application rules in (8) constitute directionally specified forms of the simplest "external" form of the Chomskian Minimalist operation "Merge". They conform to a simple generalization which governs all rules in CCG:

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

This principle is defined more formally in Steedman, 2000 in terms of three more fundamental principles of *Adjacency* or Contiguity, Directional *Inheritance*, and Directional *Consistency*, and forbids rules like the following, which combine forward functions backward (a), combine inner arguments before outer (b), or switch directionality between input and output (c):

(13) a. $Y{:}a \quad X/Y{:}f \quad \not\Rightarrow \quad X{:}f\,a$
b. $(X/Y)/Z{:}f \quad Y{:}a \quad \not\Rightarrow \quad X/Z{:}f\,a$
c. $(X/Y)/Z{:}f \quad Z{:}a \quad \not\Rightarrow \quad X\backslash Y{:}f\,a$

All bounded constructions—that is, those defining relations over the arguments of a single head, such as raising, control, passive, unaccusatives, causatives, etc.—are defined morpholexically in CCG. Where there are systematic relations or alternations between subcategorizations by the same head, as in *the door opened*, *the door was opened*, *Harry opened the door*, *the door opened itself*, etc., these may be mediated by morphological markers, or by lexical rules, or by autonomous lexical entries, or by some mixture of the above, all of which may admit of phonological regularities and exceptions, as in many other theories mentioned above, such as LFG, HPSG, SS, etc. While such choices may be extremely important to efficient representation of the grammar for purposes of processing or acquisition, they are all formally equivalent, and will not be distinguished here.

3.4 Case and morpholexical type-raising

Case is assumed to be a universal primitive of grammar (cf. Vergnaud, 1977/2006). That is to say that all noun-phrases (NP) like "Harry" are (polymorphically) *type-raised* in the morpholexicon. Type raising is so-called because it assigns to predicate-argument-structural arguments the category of a higher-order function over predicates that take NPs like "Harry" as an argument. For example, in place of (9), we have

(14)

$$
\begin{array}{ccc}
\text{Harry} & \text{sees} & \text{Sally} \\
\overline{S/(S\backslash NP_{3s})} & \overline{(S\backslash NP_{3s})/NP} & \overline{(S\backslash NP_{3s})\backslash((S\backslash NP_{3s})/NP)} \\
:\lambda p.p\ harry & :\lambda x\lambda y.sees\ x\ y & :\lambda p.p\ sally \\
 & \multicolumn{2}{c}{\underline{\hphantom{xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx}}{<}} \\
 & \multicolumn{2}{c}{S\backslash NP_{3s}} \\
 & \multicolumn{2}{c}{:\lambda y.sees\ sally\ y} \\
\multicolumn{3}{c}{\underline{\hphantom{xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx}}{>}} \\
\multicolumn{3}{c}{S} \\
\multicolumn{3}{c}{:sees\ sally\ harry}
\end{array}
$$

The effect of type-raising is to swap the roles of function and argument between subject and predicate, so that the forward application rule (8a) applies, rather than the backward rule (8b), and vice versa. crucially, the logical form that results is identical to that in (9).

Type-raising is the job of case morphemes like the nominative suffixes *-ga* in Japanese and *-us* in Latin, as in Figure 2a, in which the \\ double-slash indicates a morphemic function that can only apply inside the lexicon. In contrast, English NPs are underspecified as to case, as in Figure 2b and c (the latter involves first-person subject pro-drop, represented in the logical form as anaphorically bound one_{1s}).

Thus, even in English, type-raising is an operation of the lexical component of the grammar, not a syntactic rule.[7]

From now on we will usually abbreviate English underspecified type-raised NP etc. as $NP^{\uparrow}$ etc., with the meaning "whatever type-raised NP category is required for the derivation". Determiners will accordingly be written as $NP^{\uparrow}/N$.

Type-raising (Case) makes arguments into function categories that are more like adjuncts or specifiers than like complements (Adger, 2013). Adger uses such type-raising to avoid problematic "roll-up" derivations under a minimalist approach. This use seems parallel to the use in CCG of lexicalized type-raising to capture pied-piping relatives and *in situ wh* in examples like (32) below.

3.5 Reflexive anaphora

We assume for present purposes that reflexive pronouns are clitic, like French *se*. The boundedness of reflexivization then arises from the fact that cliticization is a morpholexical process, despite the fact that in both languages the clitic in question is written as a separate word.

---

7 Of course, the processor might choose for reasons of efficiency to leave case under-specified, and apply type-raising dynamically in context, under the control of a parsing "oracle" such as a statistical parsing model. (This is in fact how all CCG parsers for English work.) However, that does not imply that it is a rule of syntax.

a. Balb +us mur +um ædificat

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

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

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

$\overline{\quad S\backslash NP_{nom,3s}\quad}^{>}$

$\overline{\quad S\quad}^{>}$

"Balbus is building a wall."

b. Harry sees Sally

$S/(S\backslash NP)$ $: \lambda p.p\ harry$ $\quad (S\backslash NP)/NP$ $: sees$ $\quad (S\backslash NP)\backslash((S\backslash NP)/NP)$ $: \lambda p.p\ sally$

$\overline{\quad S\backslash NP : sees\ sally\quad}^{<}$

$\overline{\quad S : sees\ sally\ harry\quad}^{>}$

c. missed the Saturday dance

S/NP $: \lambda x.missed\ x\ one_{1s}$ $\quad S\backslash(S/NP)$ $: \lambda p.p\ saturdaydance$

$\overline{\quad S : missed\ saturdaydance\ one_{1s}\quad}^{<}$

d. man who I think likes Lester

N_{3s} $: man$ $\quad (N_{agr}\backslash_{\diamond}N_{agr})/_{\diamond}(S/NP_{+WH,agr})$ $\lambda p\lambda n\lambda x.p\,x \wedge n\,x$ $\quad NP^{\uparrow}_{1s}$ $\lambda p.p\ me$ $\quad ((S\backslash NP)/NP_{+WH,agr})/(S\backslash NP_{agr})$ $\lambda p\lambda x\lambda y.think\ (p\,x)\,y$ $\quad (S\backslash NP_{3s})/NP$ $\lambda x\lambda y.likes\ x\,y$ $\quad NP^{\uparrow}$ $lester$

$\overline{\quad (S/NP_{+WH,agr})/(S\backslash NP_{agr}) : \lambda p\lambda x.think\ (p\,x)\ me\quad}^{>\mathbf{B}^2}$ $\quad \overline{\quad S\backslash NP_{agr} : \lambda y.likes\ lester\ y\quad}^{<}$

$\overline{\quad S/NP_{+WH,3s} : \lambda x.think\ (likes\ lester\ x)\ me\quad}^{>}$

$\overline{\quad N_{3s}\backslash N_{3s} : \lambda n\lambda x.think\ (likes\ lester\ x)\ me \wedge n\,x\quad}^{>}$

$\overline{\quad N_{3s} : \lambda x.think\ (likes\ lester\ x)\ me \wedge man\ x\quad}^{>}$

Figure 2

We have the following categories for clitic "himself", in which the morphological slash \\ restricts its application to lexical verbs:

(15) himself := $(S\backslash NP_{3sm})\backslash\backslash((S\backslash NP_{3sm})/NP) : \lambda p\lambda y.p\,(self\ y)\,y$
$((S\backslash NP_{3sm})/PP)\backslash\backslash(((S\backslash NP_{3sm})/PP)/NP) : \lambda p\lambda w\lambda y.p\,w\,(self\ y)\,y$
etc.

Syntactically, these categories are accusative instances of type-raised cased $NP^{\uparrow}$. However, their semantics is not **T**.[8]

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

(16)
$$\begin{array}{ccc}
\text{Harry} & \text{sees} & \text{himself.} \\
\hline
NP^{\uparrow}_{3sm} & (S\backslash NP_{3s})/NP & (S\backslash NP_{3sm})\backslash\backslash((S\backslash NP_{agr})/NP) \\
:\lambda p.p\ harry & :\lambda x\lambda y.sees\ x\ y & :\lambda p\lambda y.p\ (self\ y)\ y \\
\cline{2-3}
 & \multicolumn{2}{c}{S\backslash NP_{3sm} : \lambda y.sees\ (self\ y)\ y \quad {<}LEX} \\
\hline
\multicolumn{3}{c}{S : sees\ (self\ harry)\ harry \quad {>}}
\end{array}$$

For reflexive ditransitives of the kind we saw in (1), we have the following:

(17)
$$\begin{array}{cccc}
\text{Mary} & \text{introduced} & \text{herself} & \text{to the audience.} \\
\hline
NP^{\uparrow}_{3sf} & ((S\backslash NP_{agr})/PP_{to})/NP & ((S\backslash NP_{3sf})/PP)\backslash\backslash(((S\backslash NP_{3sf})/PP)/NP) & PP^{\uparrow} \\
:\lambda p.p\ mary & :\lambda x\lambda w\lambda y.introduced\ w\ x\ y & :\lambda p\lambda w\lambda y.p\ w\ (self\ y)\ y & :\lambda p.p\ audience \\
\cline{2-3}
 & \multicolumn{2}{c}{(S\backslash NP_{3sf})/NP : \lambda w\lambda y.introduced\ w\ (self\ y)\ y \quad {<}LEX} & \\
\cline{2-4}
 & \multicolumn{3}{c}{S\backslash NP_{3sf} : \lambda y.introduced\ audience\ (self\ y)\ y \quad {<}} \\
\hline
\multicolumn{4}{c}{S : introduced\ audience\ (self\ mary)\ mary \quad {>}}
\end{array}$$

It seems reasonable to assume that *Harry talks to himself* is also a true *se*-type reflexive arising from lexicalization of "talks to", as in the following derivation:[9]

(18)
$$\begin{array}{ccc}
\text{Harry} & \text{talks to} & \text{himself.} \\
\hline
NP^{\uparrow}_{3sm} & (S\backslash NP_{3s})/NP & ((S\backslash NP_{3sm}))\backslash\backslash((S\backslash NP_{agr})/NP) \\
:\lambda p.p\ harry & :\lambda x\lambda y.talks\ (to\ x)\ y & :\lambda p\lambda y.p\ (self\ y)\ y \\
\cline{2-3}
 & \multicolumn{2}{c}{S\backslash NP_{3sm} : \lambda y.talks\ (to\ (self\ y))\ y \quad {<}LEX} \\
\hline
\multicolumn{3}{c}{S : talks\ (to\ (self\ harry))\ harry \quad {>}}
\end{array}$$

Example (19a) can be analysed similarly to (17). However, the reflexives in (19b–e) cannot reasonably be analysed as clitic in the same way, and must be "exempt" or logophoric pronouns, of a kind to be discussed below:

(19) a. Harry showed himself a movie.
b. Harry showed a movie to himself.

---

8 The analysis is similar to that of Szabolcsi (1989), which is also lexicalized.
9 This possibility may be related to the cross-linguistically unusual possibility in English of "preposition-stranding" *wh*-extraction.

c. Harry talks to and about himself.
d. Harry talks to only himself.
e. Harry sees and admires himself.

The following further “subject reflexive” instance of the type-raised reflexive for the non-existent “*heself”, (20) is excluded for English because it is not a possible English cased category, since $(S\backslash NP)/NP_{3sm}$ is not an English transitive verb category:[10]

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

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

The presence in English 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:

(21) 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, rather than from true reflexives.

(22) 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$_i$.

4 The unbounded *wh*-constructions

Relativization and all unbounded dependencies crucially involve type-raising and the syntactic combinatory rules of function composition, considered next.

10 $(S\backslash NP)/NP_{3sm}$ is in fact the category of an *ergative* transitive verb with absolutive agreement. Woolford (1999, 267–8) shows that ergative languages with absolutive agreement such as Inuit and Nez Perce disallow the equivalent of “He$_{erg}$ sees himself$_{abs}$”, which would require raising over the English transitive type, while ergative languages without absolutive agreement, such, as Enga, do allow them.

4.1 Syntactic rules II: Composition

We will need the following rules of composition, which constitute all and only those allowed by the CPP (12) for these (first-order) categorial types:

(23) 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:

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

Like the application rules (8), these rules have the effect of "canceling" Y, as if this were fractional multiplication. The types $\diamond$ and $\times$ on the slashes in these rules mean that only categories whose own slash type is compatible under a type-hierarchy of slash types (Baldridge, 2002) can combine by these rules. The simplified convention used in the present paper is that only categories with $\diamond$ slashes or unrestricted slashes can combine by $\diamond$ rules, and $\diamond$ categories cannot combine by the crossing $\times$ rules. Similarly, categories with a $\times$ or unrestricted slash-type can combine by $\times$ rules, but $\times$ categories cannot combine by $\diamond$ rules.[11]

The rules in (23) obey the CPP (12), including the Principle of Adjacency or contiguity. The absence of slash-typing on the secondary function $Y|Z$ means that it can apply to any type, but the CPP requires that that type will be passed to the result $X|Z$. Thus, as with the application rules (8), in Minimalist terms, the composition rules (23) constitute additional cases of External Merge, except for allowing the equivalent of some "feature-checking" that is not allowed under Minimalist merger (Adger, 2003, 93–94). However, they thereby achieve the same result as Minimalist *Internal* Merge, or Move.

To see this in the context of the relative clause, we will assume the following category for the English relative pronoun:

(24) that := $(N\backslash_{\diamond}N)/_{\diamond}(S/NP)$

We can then derive a relativized noun modifier "that she had met" of type $N\backslash_{\diamond}N$ (from the introductory example (1) and Figure 1b, slight simplified), as follows, using the first of the composition rules (23a) to form a constituent of type $S\backslash NP$ adjacent to the relative pronoun by adjacent merger of the elements of the residue of relativization.

11 However, another less restrictive convention is possible, in which these two slash types are explicitly conjunctive, written $\diamond *$ and $\times *$, while $\diamond$ and $\times$ types can *only* compose.

(25) (The man) that she had met

$$\begin{array}{ccccc}
\text{(The man)} & \text{that} & \text{she} & \text{had} & \text{met} \\
\hline
 & (N\backslash_{\diamond}N)/_{\diamond}(S/NP) & S/(S\backslash NP) & (S\backslash NP)/VP_{pptcpl} & VP_{pptcpl}/NP \\
 & :\lambda p\lambda n\lambda x.p\,x \wedge n\,x & :\lambda p.p\,pro_{she} & :\lambda p\lambda y.past\,(perfect\,(p\,y)) & :\lambda x\lambda y.meet\,x\,y
\end{array}$$

$$\xrightarrow{\;>\mathbf{B}\;} S/VP : \lambda p.past\,(perfect\,(p\,pro_{she}))$$

$$\xrightarrow{\;>\mathbf{B}\;} S/NP : \lambda x.past\,(perfect(meet\,x\,pro_{she}))$$

$$\xrightarrow{\;>\;} N\backslash_{\diamond}N : \lambda n\lambda x.past\,(perfect(meet\,x\,pro_{she})) \wedge n\,x$$

4.1.1 Unbounded relativization

The same combinatory rule (23a) can apply recursively, to build a constituent of the same type S/NP by multiple adjacent mergers, to which the relative pronoun can apply as before to yield a noun modifier (semantics omitted):

(26)

$$\begin{array}{ccccccc}
\text{(The man)} & \text{that} & \text{she} & \text{says} & \text{that} & \text{she} & \text{met} \\
\hline
 & (N\backslash_{\diamond}N)/_{\diamond}(S/NP) & NP^{\uparrow} & (S\backslash NP)/_{\diamond}S' & S'/_{\diamond}S & NP^{\uparrow} & (S\backslash NP)/NP
\end{array}$$

$$\text{she says}: \xrightarrow{\;>\mathbf{B}\;} S/_{\diamond}S' \qquad \text{she met}: \xrightarrow{\;>\mathbf{B}\;} S/NP$$

$$\text{that she met}: \xrightarrow{\;>\mathbf{B}\;} S'/NP$$

$$\text{she says that she met}: \xrightarrow{\;>\mathbf{B}\;} S/NP$$

$$\text{that she says that she met}: \xrightarrow{\;>\;} N\backslash_{\diamond}N$$

4.1.2 Embedded subject extraction

However, the $\diamond$ slash-types of the complement of *think* correctly forbids extraction of the subject of a *that* complement:

(27) *men$_N$ that [she says that]$_{S/_{\diamond}S}$ [met her]$_{S\backslash NP}$

This is not a stipulation that could be otherwise: if a language like English allowed *says* or *says that* to compose with *sees her* by crossed composition it would immediately allow such non-English orders as the following with the meaning "she says the men met her":

(28) *She the men says met her.

Such considerations lead us to expect that a general prohibition against extraction of complement subjects should be a specific characteristic of rigidly SVO languages and constructions, in which the directionality of S and O arguments differ, but not of SOV and VSO, in which they are the same, a generalization that appears to be correct (Pesetsky, 2017).

It also follows that in order to overcome the general prohibition on subject extraction out of bare complements for the class of English verbs that allow them, we need to give them an extra lexical category, like the following:

(29) think, believe, ... $:= ((S\backslash NP)/NP_{+WH,agr})/(S\backslash NP_{agr}) : \lambda p \lambda x \lambda y.think(px)y$

This category combines with a tensed predicate to yield something requiring an NP marked as $+WH$ which can only reduce with a relativized category, as in the derivation of "[a] man who I think likes Lester" in Figure 2d.[12]

4.1.3 Parasitic extraction

For completeness, we note in passing that a further class of rule including the one here indexed $<\mathbf{S}_\times$, constituting a "duplicating" generalization of composition is needed to capture "parasitic gapping" cases of relativization, again building a constituent of type S/NP by a succession of adjacent mergers:

(30)
$$\begin{array}{lccccccl}
\text{(The articles)} & \text{that} & \text{Harry} & \text{rejected} & \text{without} & \text{reading} & \\
\cline{2-2}\cline{3-3}\cline{4-4}\cline{5-5}\cline{6-6}
 & (N\backslash_{\diamond}N)/_{\diamond}(S/NP) & NP^{\uparrow} & (S\backslash NP)/NP & ((S\backslash NP)\backslash(S\backslash NP))/VP & VP/NP & \\
\cline{5-6}
 & & & & \multicolumn{2}{c}{((S\backslash NP)\backslash(S\backslash NP))/NP} & >\mathbf{B} \\
\cline{4-6}
 & & & \multicolumn{3}{c}{(S\backslash NP)/NP} & <\mathbf{S}_{\times} \\
\cline{3-6}
 & & \multicolumn{4}{c}{S/NP} & >\mathbf{B} \\
\cline{2-6}
 & \multicolumn{5}{c}{N\backslash_{\diamond}N} & >
\end{array}$$

We pass over the details of the **S** rules here, referring to Szabolcsi (1983, 1989) and Steedman (1987, 1996), noting not only that they are constrained by the Combinatory Projection Principle (12), but also that they exploit all degrees of freedom allowed under that principle, which correctly allows parasitism to be supported by complement subject extraction, forcing the choice of *rightward* subcategorization for the extracting NP in the subject-extracting category (29).

(31) A man that Harry will [tell you is a crook]$_{VP/NP_{+WH}}$ [while pretending to admire]$_{(VP\backslash VP)/NP}$

4.1.4 Pied-piping extraction

The phenomenon of "pied-piping" in relativizations like the following is captured by giving relative pronouns like *which* (but not *that*) the further category shown in the fol-

---

12 The reason for the extracting NP being a *rightward* argument will become clear below, when we consider the "Across-the-Board" constraint on extraction under coordination.

lowing example, the details of whose derivation we pass over (Steedman, 1987; Morrill, 1994; Steedman, 2012):

(32) [[Reports]$_N$ [[[the height of the lettering on the covers of]$_{(S\backslash(S/NP))/NP}$
[which]$_{((N\backslash N)/(S/NP))\backslash(((S\backslash(S/NP))/NP))}$ [the government prescribes]$_{S/NP}$

4.2 Crossing dependencies

The inclusion of crossing composition rules, together with the following generalization of the composition rules to "higher order" cases with second-order secondary functors of the form $(Y/Z)/W$ allows the set of possible non-terminal categories to grow unboundedly, showing CCG grammars to be trans-context free.

(33) a. Forward 2nd-order 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 2nd-order Composition:
$$(Y\backslash Z)\backslash W : g \quad X\backslash_{\diamond}Y : f \quad \Rightarrow \quad (X\backslash Z)\backslash W : \lambda w\lambda z.f\,(g\,w\,z) \qquad (<\mathbf{B}^2)$$
c. Forward Crossing 2nd-order Composition:
$$X/_{\times}Y : f \quad (Y\backslash Z)\backslash W : g \quad \Rightarrow \quad (X\backslash Z)\backslash W : \lambda w\lambda z.f\,(g\,w\,z) \qquad (>\mathbf{B}^2_{\times})$$
d. Backward Crossing 2nd-order Composition:
$$(Y/Z)/W : g \quad X\backslash_{\times}Y : f \quad \Rightarrow \quad (X/Z)/W : \lambda w\lambda z.f\,(g\,w\,z) \qquad (<\mathbf{B}^2_{\times})$$

(Again, the effect of these rules is to "cancel" Y.) We assume following *SP* that these rules are the only higher-order composition rules. (In particular, there are no such rules with mixed directionality in the secondary function.)

This feature of the theory allows elegant capture of a Germanic control construction that allows unboundedly many arguments to cross dependencies with their governing verbs, as in Figure 3a,b. This was a phenomenon which allowed the first formal proof that natural languages were not even weakly context-free (Huybregts, 1984; Shieber, 1985, data for Zurich German from Shieber).

Some alternative orders to those in Figure 3a,b including the following are correctly also allowed in CCG (Shieber, 1985):

(34) a. Das mer em Hans hälfed es huus aastriiche.
b. Das mer d'chind lönd em Hans hälfe es huus aastriiche.

5 Coordination

5.1 Right node raising

Because the category S/NP of the domain of relativization does not distinguish the status of the argument $/NP$ as extracted or lexical, we immediately predict the possibility

a. das mer em Hans es huus hälfed aastriiche

that we–NOM Hans–DAT the house–ACC helped paint

$NP^{\uparrow}_{nom}$ $NP^{\uparrow}_{dat}$ $NP^{\uparrow}_{acc}$ $((S_{+SUB}\backslash NP_{nom})\backslash NP_{dat})/_{\times}VP$ $VP\backslash NP_{acc}$

$$\xrightarrow{>\mathbf{B}_{\times}} ((S_{+SUB}\backslash NP_{nom})\backslash NP_{dat})\backslash NP_{acc}$$

$$\xrightarrow{>} (S_{+SUB}\backslash NP_{nom})\backslash NP_{dat}$$

$$\xrightarrow{>} S_{+SUB}\backslash NP_{nom}$$

$$\xrightarrow{>} S_{+SUB}$$

"that we helped Hans paint the house"

b. 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

$NP^{\uparrow}_{nom}$ $NP^{\uparrow}_{acc}$ $NP^{\uparrow}_{dat}$ $NP^{\uparrow}_{acc}$ $((S_{+SUB}\backslash NP_{nom})\backslash NP_{acc})/_{\times}VP$ $(VP\backslash NP_{dat})/_{\times}VP$ $VP\backslash NP_{acc}$

$$\xrightarrow{>\mathbf{B}^{2}_{\times}} (((S_{+SUB}\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat})/_{\times}VP$$

$$\xrightarrow{>\mathbf{B}_{\times}} (((S_{+SUB}\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat})\backslash NP_{acc}$$

$$\xrightarrow{>} ((S_{+SUB}\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat}$$

$$\xrightarrow{>} (S_{+SUB}\backslash NP_{nom})\backslash NP_{acc}$$

$$\xrightarrow{>} S_{+SUB}\backslash NP_{nom}$$

$$\xrightarrow{>} S_{+SUB}$$

"that we let the children help Hans paint the house"

Figure 3: Crossing dependencies in Zurich German.

of unbounded *rightward* movement, as in the following derivation:

(35)
$$\begin{array}{llll}
[\text{Harry sees}] & \text{and} & [\text{Fred says he likes}] & \text{Sally}\\
\overline{}^{>\mathbf{B}} & \overline{} & \overline{}^{>\mathbf{B}} & \overline{}\\
S/NP & (X\backslash_{\star}X)/_{\star}X & S/NP & NP^{\uparrow}\\
 & \multicolumn{2}{l}{\overline{}^{>}} & \\
 & \multicolumn{2}{c}{(S/NP)\backslash(S/NP)} & \\
\multicolumn{3}{l}{\overline{}^{<}} & \\
\multicolumn{3}{c}{(S/NP)} & \\
 & \multicolumn{3}{l}{\overline{}^{<}}\\
 & \multicolumn{3}{c}{S}
\end{array}$$

The category $(X\backslash_{\star}X)/_{\star}X\ :\ \lambda p\lambda q.p \sqcap q$ of the conjunction is restricted by $\star$ slash-typing to only combining by the application rules (8) rather than the more restricted composition rules (23). This condition imposes Ross' 1967 "Across the Board" (ATB) constraint (1967) on both rightward and leftward extraction out of coordinate structures including the "same case" condition, as in the following examples.

(36) a. the woman that [sees Harry]$_{S/NP}$ and [likes Fred]$_{S\backslash NP}$
b. the woman that [Harry sees]$_{S/NP}$ and [Fred likes]$_{S/NP}$
c. *the woman that[Harry sees]$_{S/NP}$ and [Fred likes her]$_{S}$
d. *the woman that[Harry sees her]$_{S}$ and [Fred likes]$_{S/NP}$
e. *the woman that[Harry sees]$_{S/NP}$ and [likes Fred]$_{S\backslash NP}$

The restriction on the conjunction category prevents application to S in one conjunct and composition into S on the other. This restriction should be seen as a consequence of the semantics, which is defined as Partee and Rooth's transitive closure $p \sqcap q$ over function types (1983), and must therefore apply to p and q of the same type.

Constituents including extracted complement subjects can coordinate with those containing extracted objects:

(37) a man that we had invited, and believed would come

Under the CCG account of coordination proposed here, this fact again forces the assumption that they have the same category—in this case VP_{pstp}/NP—with the subject-extracted ones differing only in being restricted to *wh* "antecedent government" via the subject-extracting category (29).

The same restriction means that right-node raising corresponding to (37) is blocked, in a rare exception to the symmetry of right- and left-extraction:

(38) *We had invited, and believed would come, the man who broke the bank at Monte Carlo.

5.2 Argument/Adjunct cluster coordination

Less obviously, the assumption that all arguments are mopholexically *cased*, or type-raised, including accusatives and datives as well as nominatives predicts the following "Argument/Adjunct Cluster" coordination (Dowty, 1988):

(39)

$$
\begin{array}{cccccc}
\text{Give} & \text{Harry} & \text{books} & \text{and} & \text{Sally} & \text{records} \\
\hline
(VP/NP)/NP & (VP/NP)\backslash((VP/NP)/NP) & VP\backslash(VP/NP) & (X\backslash X)/X & (VP/NP)\backslash((VP/NP)/NP) & VP\backslash(VP/NP) \\
\mathit{give} & :\lambda p.p\ \mathit{harry} & :\lambda p.p\ \mathit{books} & :\lambda q\lambda p.p \sqcap q & :\lambda p.p\ \mathit{sally} & :\lambda p.p\ \mathit{records}
\end{array}
$$

$$\underline{\qquad\qquad\qquad\qquad}_{<\mathbf{B}} \qquad\qquad\qquad \underline{\qquad\qquad\qquad\qquad}_{<\mathbf{B}}$$

$$VP\backslash((VP/NP)/NP) : \lambda p.p\ \mathit{harry\ books} \qquad VP\backslash((VP/NP)/NP) : \lambda p.p\ \mathit{sally\ records}$$

$$\underline{\qquad\qquad\qquad\qquad}_{>}$$

$$(VP\backslash((VP/NP)/NP))\backslash(VP\backslash((VP/NP)/NP)) : \lambda p\lambda r.r\ \mathit{harry\ books} \wedge p\ r$$

$$\underline{\qquad\qquad\qquad\qquad}_{<}$$

$$VP\backslash((VP/NP)/NP) : \lambda r.r\ \mathit{harry\ books} \wedge r\ \mathit{sally\ records}$$

$$\underline{\qquad\qquad\qquad\qquad}_{<}$$

$$VP : \mathit{give\ harry\ books} \wedge \mathit{give\ sally\ records}$$

The argument cluster coordination construction (39) is an example of a universal tendency for “deletion under coordination” to respect basic word order: in all languages, if arguments are on the left of the verb then argument clusters coordinate on the left, if arguments are to the right of the verb then argument clusters coordinate to the right of the verb, while SVO languages pattern with verb-initial (Ross, 1970):

(40)

SVO:	*SO and SVO	SVO and SO
VSO:	*SO and VSO	VSO and SO
SOV:	SO and SOV	*SOV and SO

For example, all contiguous substrings of Shieber’s Zurich German examples in Figure 3a,b are correctly predicted to coordinate with sequences of the same type (Steedman, 1985). (However, there is more to say concerning the precise mechanism that allows verb-medial gapping in the SVO case, and why it patterns with VSO—see Steedman, 1990, 2000.)

Such cluster coordinations were the motivation for Pesetsky’s (1995) postulation of a level of “cascade structure” as the domain of binding and coordination, distinct from “layered structure”, the domain of movement. In CCG, layered and cascade structure correspond to the same single level of derivation structure.

6 On intonation structure and the notion “surface constituent” in CCG

It is important to be clear at this point that CCG categories like S/NP are not equivalent to SLASH notations in GPSG/HPSG. In particular, in CCG, a category of the form X/NP does not denote a constituent of type X including a trace or gap of type NP.

On the contrary, S/NP is in CCG a constituent type in its own right, free to either combine with a preposed or *in situ* relativized element, or to combine with a full NP—in particular, one that is right node-raised across-the-board, as in (35).[13]

13 The accounts of right node-raising in Gazdar (1981) and Gazdar et al. (1985) both require a separate metarule to license RNR.

As a consequence, CCG generalizes the notion of constituency beyond the traditional notion to include *any sequence that is typable using combinatory rules*, including *Harry sees*, of type S/NP, and even *Harry books*, which we saw in (39) can be typed as $VP\backslash((VP/NP)/NP)$. As a consequence, CCG also necessarily allows (many) alternative derivations or constituent structures for canonical sentences. For example, as well as the earlier standard derivation (14), it allows the following:

(41)
$$\begin{array}{ccc}
\text{Harry} & \text{sees} & \text{Sally} \\
\hline
NP^{\uparrow} & (S\backslash NP)/NP & NP^{\uparrow} \\
:\lambda p.p\ harry & :sees & :\lambda p.p\ sally \\
\hline
\multicolumn{2}{c}{S/NP : \lambda x.sees\ x\ harry} {}_{>\mathbf{B}} & \\
\hline
\multicolumn{3}{c}{S : sees\ sally\ harry} {}_{<}
\end{array}$$

For longer sentences of length n there will be a number of alternative analysis up to the nth number in the more than exponentially rapidly-growing Catalan series, all yielding the same logical form.

This proliferation of constituent structures is sometimes referred to as "spurious ambiguity". However, it should not be regarded as a weakness in the combinatory theory of grammar. Many languages have freer word order than English, and do not support any clear notion of surface constituency. Even for English, there is no clear consensus on whether the surface structure of the ditransitive VP or the subject auxilliary-inverted question is flat and ternary-branching, left-branching, or right-branching (Adger, 2003, 122–131, and cf. Barss and Lasnik, 1986; Larson, 1988; Pesetsky, 1995; Jackendoff, 1990; Larson, 1990, *passim*). Nor is structural ambiguity a problem for performance or processing. It is just a fact of life (many other constructions, such as noun-noun compounding, yield Catalan-serial numbers of analyses). Parsers are good at dealing with other sources of ambiguity by the use of an "oracle" such as a statistical model, and they can do the same with this one. (It is worth remembering that ambiguity is endemic in all natural languages, and that none of them shows the slightest sign of evolving in the direction of reducing their overall level of ambiguity—Labov, 1994, 40–42, chs. 19,20; Croft, 2000, 68, 102–4; Newmeyer, 2003, 694; *passim*.)

CCG's unorthodox notion of constituency is transparently reflected in prosodic structure. Thus the following intonation contours appropriately mark the two alternative derivations for the transitive clause appropriate to the two context setting questions Q:

(42) Q: I know BARRY sees ALICE. But who sees SALLY?

A: (HARRY) (sees SALLY) .

$$\begin{array}{ccc}
\text{H*} & \text{L+H*} & \text{LH\%} \\
\hline
NP^{\uparrow} & \multicolumn{2}{c}{S\backslash NP}
\end{array}$$

(43) Q: I know BARRY sees ALICE. But who does HARRY see?

A: (HARRY sees) (SALLY) .
L+H* LH % H* LL %

$\underline{}$ $\underline{}$

S/NP $NP^{\uparrow}$

The notation for intonation-phrasal tunes is from Pierrehumbert and Hirschberg (1990). Here, L+H* LH % marks topic or Theme, H* LL % marks comment or Rheme. Exchanging the A(nswer)s to the Q(uestions) is highly unacceptable. Steedman (2014) develops a theory of intonation structure and its meaning using the Alternative-Semantic framework of Rooth (2016).

As in the Match Theory of Intonation Structure of Selkirk (2011) and its earlier Edge-marking incarnation (1990), Theme/Rheme marking is projected onto phrasal constituents directly, by syntactic derivation alone, bounded by combination of the phrase with an edge-based boundary tone. However, no independent extra-syntactic mechanism of "Focus Projection" is needed to achieve the semantics of "broad focus". Nor are any violable constraints needed to explain departures of intonation structure from syntactic derivational structure, for there are no such departures. In CCG, surface syntactic structure is simply identical to phonological form.

Thus, the domain of the prosodic phrase ϕ is the same as the domain of *wh*-movement, a state which is aspired to in Minimalist Contiguity Theory (Richards, 2016, Ch.3; Richards, 2016, 9). Prosodic structure is thereby defined as part of "Narrow Syntax" in the sense of Chomsky (2001).[14]

7 Explanatory adequacy

CCG's combination of type-raising and composition subject to the CPP (12) yields a permuting and rebracketing calculus closely tuned to the needs of and constraints on natural grammar.

CCG thereby reduces Minimalisms' MOVE/INTERNAL MERGE and COPY/DELETE, together with intonational phrasing, to contiguous EXTERNAL MERGE.

Constraints on dependency projection, such as the **that-t* and across-the-board conditions, arise from the nature of the lexicon and combinatorics of CCG rather than from additional constraints on syntactic derivation. Part of this explanatory force arises from the low expressive power of the combinatory rules under the CPP, to which we now turn.

14 However, there are some important differences. For Richards, the possibility of *in situ wh*-elements depends on everything between COMP and *wh* forming a single prosodic phrase. Otherwise, *wh*-movement is forced. In the present terms, *wh*-movement *also* is only possible if everything can be composed and thereby also become a contiguous intonational phrase.

7.1 Expressive power of CCG

CCG and TAG are provably weakly equivalent to Linear Indexed Grammar (LIG) (Vijay-Shanker and Weir, 1994, Kuhlmann et al., 2015).[15] Both are therefore "Mildly Context Sensitive" under the definition of Joshi (1988) and Joshi et al. (1991), which (among other properties), in the latter case calls for non-permutation-completeness.

In particular, of the $n!$ possible permutations on n functional heads, CCG only allow a proportion defined by the nth Large Schröder Number $S(n)$. For example, for a "cartographic" right-branching spine of only 8 functional heads, nearly 80 % of the 8! permutations are excluded.

This property was first noted by Williams (2003, 125) for his categorial system CAT. Williams (2003, 209) wrongly claimed that the inclusion of type-raising in CCG would allow all orders. However, Williams failed to note that, as we have seen, type-raising in CCG is morpho-lexical and defined over lexical functions over the original types, rather than a free syntactic operation. Type-raising changes the set of types involved, and therefore changes the "Basic Order of Merger", defined by purely applicative derivation. Nevertheless, any fixed set of types, including raised types, is as a consequence non-permutation-complete.

Specifically, for a set of four categories of the form $A|B$, $B|C$, $C|D$, D, determining a basic order of merger $\{1,2,3,4\}$, 22 out of the 24 possible permutations are allowed. The two that CCG excludes are the following:

(44) a. 3 1 4 2
b. 2 4 1 3

An example of a construction of this form is the nounphrase construction investigated by Cinque (2005) and Abels and Neeleman (2012), and mentioned in Section 1.5, for which the categories in English are the following:

(45) 1: $\text{these}_{NP^{\uparrow}/N_{num}}$ 2: $\text{five}_{N_{num}/N}$ 3: $\text{fat}_{N/N}$ 4: cats_{N}

The prediction is that no language will require or allow orders corresponding to either of the following glosses:[16]

(46) a. *fat these cats five
b. *five cats these fat

These two orders are indeed not listed among the fourteen orders that Cinque identifies as attested for the languages of the world, nor are they included among the nine-

15 Weak equivalence means that they admit the same stringsets, though not via the same derivations. Kuhlmann et al. (2015) show that the specific slash-typing version of CCG presented in this chapter is actually slightly less expressive than TAG.

16 In testing such predictions, Cinque (2005) points out that it is important to be sure in particular that adjectives like "fat" are functioning as such, rather than being extraposed.

teen orders that Nchare (2012) identifies as occurring in the free word-order language Shupamem.

If we renumber the original set 1, 2, 3, 4 as X, 1, 2, 3, then (44b) is 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 "roll-up" derivations such as pied-piping in Germanic. Thus, it seems likely that the orders in (44) are indeed universally excluded. In CCG this restriction is of the kind identified in *Aspects* as a Formal Universal, stemming without stipulation from the theory of grammar itself.

This property of non-permutation-completeness puts CCG at a different level of the extended language hierarchy of "abstract families of languages" (Ginsburg and Greibach, 1967) than standard Minimalist theories. Michaelis (1998, 2001) showed that Minimalist Grammars (MG) under the definition of Stabler (2011) and including the "Shortest Move" or Minimal Link Condition (MLC) on movement, are weakly equivalent to Linear Context-Free Rewriting Systems (LCFRS), or, equivalently, to Multiple Context-Free Grammars (MCFG). At the time, it was conjectured that LCFRS/MCFG were Mildly Context Sensitive (MCS), from which it seemed to follow that MG+MLC was also MCS. However, it has subsequently been shown by Salvati (2011/2015) that the artificial language MIX, consisting of all permutations on the string $a^n b^n c^n$ is an MCFL.

Under the (informal) definition of Joshi et al. (1991), mild context-sensitivity explicitly excludes permutation-completeness. So, since MCFL includes MIX, a permutation-complete language by definition, MCFG is not MCS under that definition, and so neither is MG. On the other hand, under the formal definition of MCS given in Kallmeyer, 2010, 23–24, the MCS languages include MCFL. Salvati suggests that the languages characterized by the well-nested subset of MCFG, MCFG_{WN} might formally correspond to the set of MCS languages, and shows that TAG is weakly in MCFG_{WN}, while Kanazawa and Salvati (2012) show that MIX is *not* in MCFL_{WN}. However, the MCFGs to which Stabler's MG corresponds are known to be non-well-nested (Boston et al., 2010), (although Kanazawa et al., 2011 show that the addition of a further Specifier Island Contraint to MG restricts them to a subset of MCFG_{WN}).

By contrast, Kanazawa and Salvati (2012) also show that MIX is not a Tree Adjoining Language (TAL), and therefore not a Combinatory Categorial Language (CCL), so TAG and CCG remain mildly context sensitive in a stronger sense, without constraints, under all definitions.

In this rather confusing state of affairs, it therefore seems helpful to distinguish the latter, not merely as mildly context sensitive, but more narrowly as "Slightly Non-Context-Free" (SNCF).

7.2 Child language acquisition

The restriction of the CCG lexicon and combinatory rules to strict type-transparency between constituents and their logical forms means that it supports a practically computable model of child language acquisition via “semantic bootstrapping” (Pinker, 1979), or, more properly, learning given access to contextually accessible universal logical form.

The problem of child language acquisition then reduces to the problem of learning (a) a lexicon, and (b) a parsing model, for all rules consistent with the (noisy) language-specific data and the (ambiguously) contextually available meaning representations, of which those sound-meaning pairs belonging to the actual target grammar will be vastly more frequent than the spurious ones (Abend et al., 2017).

Interestingly, the learner of Abend et al. gives a superficial appearance of learning parameters. (For example, in the later stages of learning English, the probability mass assigned to the SVO category for an unseen transitive verb will be near 1, and all other categories will be near 0.) However, there is no learned parametric prior directly associating this semantic type with this category. Instead, the information is implicit in the probabilities assigned to the various instantiated syntactic rules in the grammar as a whole that are used in calculating the prior probabilities of alternative derivations.

8 Conclusion: Towards a combinatory minimalism

CCG is a theory that embodies in a very direct form the Minimalist insight that syntactic derivation is determined by need to create objects that are phonologically and semantically well-formed, and nothing else, reducing unbounded movement to contiguous composition and case to morpholexical type-raising.

Other bounded types of movement, including “raising” and “control” relations; “head movement” (Roberts, 2001), “scrambling” (Ross, 1967), and “sideward movement” (Nunes, 2001) are defined statically, in the lexicon, at the level of logical form, from which they are projected by syntactic derivation.

In view of recent invocations among Minimalists of both prosodic contiguity and type-raising, there seems to be a possibility of extending the Minimalist Program to cover the full range of movement, coordination, and prosodic structural phenomena under the following equation, subject to the combinatory projection principle (CPP), (12):

Minimalism = Categorial Grammar + Case + Composition

References

Abels, Klaus & Ad Neeleman. 2012. Linear asymmetries and the LCA. *Syntax* 15. 25–74.
Abend, Omri, Tom Kwiatkowski, Nathaniel Smith, Sharon Goldwater & Mark Steedman. 2017. Bootstrapping language acquisition. *Cognition* 164. 116–143.
Ades, Anthony & 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. 2013. *A Syntax of Substance*. MIT Press.
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.
Bach, Emmon. 1976. An extension of classical Transformational Grammar. *Problems in Linguistic Metatheory: Proceedings of the 1976 Conference at Michigan State University*, 183–224. Lansing, MI: Michigan State University.
Baldridge, Jason. 2002. Lexically Specified Derivational Control in Combinatory Categorial Grammar. Ph.D. thesis, University of Edinburgh.
Bar-Hillel, Yehoshua. 1953. A quasi-arithmetical notation for syntactic description. *Language* 29. 47–58.
Barss, Andrew & Howard Lasnik. 1986. A note on anaphora and double objects. *Linguistic Inquiry* 17. 347–354.
Boas, Hans & Ivan A. Sag. 2012. *Sign-based Construction Grammar*. Stanford, CA: CSLI Publications.
Boston, Marisa Ferrara, John Hale & Marco Kuhlmann. 2010. Dependency structures derived from Minimalist Grammars. *Proceedings of the Conference on the Mathematics of Language*, 1–12. Springer.
Bozşahin, Cem. 2012. *Combinatory Linguistics*. Berlin: de Gruyter.
Bresnan, Joan (ed.). 1982. *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press.
Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
Chomsky, Noam. 1972. *Studies on Semantics in Generative Grammar*. Mouton: The Hague.
Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
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 & 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 & Howard Lasnik. 1977. Filters and control. *Linguistic Inquiry* 8. 425–504.
Cinque, Guglielmo. 2005. Deriving Greenberg's universal 20 and its exceptions. *Linguistic Inquiry* 36. 315–332.
Cooper, Robin. 1983. *Quantification and Syntactic Theory*. Dordrecht: Reidel.
Cormack, Annabel & Neil Smith. 2002. Compositionality, copy theory, and control. *Working Papers in Linguistics* 14. 355–373.
Croft, William. 2000. *Explaining Language Change: An Evolutionary Approach*. Longman.
Culicover, Peter W. & Ray Jackendoff. 2005. *Simpler Syntax*. Oxford: Oxford University Press.
Dowty, David R. 1988. Type-raising, functional composition, and nonconstituent coordination. In Richard Oehrle, Bach Emmon & Deirdre Wheeler (eds.), *Categorial Grammars and Natural Language Structures*, 153–198. Dordrecht: Reidel.
Dowty, David R. 1993. variable-free' syntax, variable-binding syntax, the natural deduction Lambek calculus, and the crossover constraint. *Proceedings of the Eleventh West Coast Conference on Formal Linguistics, 1992*, 161–176., Stanford, CA: SLA.
Gazdar, Gerald. 1981. Unbounded dependencies and coordinate structure. *Linguistic Inquiry* 12. 155–184.
Gazdar, Gerald, Ewan Klein, Geoffrey Pullum & Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar*. Oxford: Blackwell.

Ginsburg, Seymour & Sheila Greibach. 1967. Abstract families of languages. *Proceedings of Eighth Annual Symposium on Switching and Automata Theory, 1967, SWAT 1967*, 128–139. IEEE.

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.

Hays, David. 1964. Dependency theory: a formalism and some observations. *Language* 40. 511–525.

Hepple, Mark. 1990. The Grammar and Processing of Order and Dependency: A Categorial Approach. Ph.D. thesis, University of Edinburgh.

Higgins, F. Roger. 1973. The Pseudocleft Construction in English. Ph.D. thesis, MIT, Cambridge, MA.

Hudson, Richard. 1984. *Word Grammar*. Oxford: Blackwell.

Huybregts, Riny. 1984. The weak inadequacy of context-free phrase-structure grammars. In Ger de Haan, Mieke Trommelen & Wim Zonneveld (eds.), *Van Periferie naar Kern*, 81–99. Dordrecht: Foris.

Jackendoff, Ray. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 1975. Morphological and semantic regularities in the lexicon. *Language* 50. 639–671.

Jackendoff, Ray. 1990. On Larson's treatment of the double object construction. *Linguistic inquiry* 21. 427–456.

Jackendoff, Ray. 1997. *The Architecture of the Language Faculty* (Linguistic Inquiry Monograph 28). Cambridge, MA: MIT Press.

Jackendoff, Ray. 2002. *Foundations of Language*. Oxford: Oxford University Press.

Jacobson, Pauline. 1999. Towards a variable-free semantics. *Linguistics and Philosophy* 22. 117–184.

Joshi, Aravind, K. 1988. Tree-adjoining grammars. In David R. Dowty, Lauri Karttunen & Arnold Zwicky (eds.), *Natural Language Parsing*, 206–250. Cambridge: Cambridge University Press.

Joshi, Aravind, K., K. Vijay-Shanker & David Weir. 1991. The convergence of mildly context-sensitive formalisms. In Peter Sells, Stuart Shieber & Thomas Wasow (eds.), *Processing of Linguistic Structure*, 31–81. Cambridge, MA: MIT Press.

Kallmeyer, Laura. 2010. *Parsing Beyond Context-Free Grammars*. Springer.

Kanazawa, Makoto, Jens Michaelis, Sylvain Salvati & Ryo Yoshinaka. 2011. Well-nestedness properly subsumes strict derivational Minimalism. *International Conference on Logical Aspects of Computational Linguistics*, 112–128. Springer.

Kanazawa, Makoto & Sylvain Salvati. 2012. MIX is not a Tree-Adjoining Language. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 666–674. Jeju Island, Korea: ACL.

Kay, Martin. 1984. Functional unification grammar: A framework for machine translation. *Proceedings of the 22nd Annual Conference of the Association for Computational Linguistics, Stanford, CA*, 75–78. ACL Anthology.

Kuhlmann, Marco, Alexander Koller & Giorgio Satta. 2015. Lexicalization and generative power in CCG. *Computational Linguistics* 41. 187–219.

Labov, William. 1994. *Principles of Linguistic Change. Vol. 1: Internal Features*. Oxford: Blackwell.

Lambek, Joachim. 1958. The mathematics of sentence structure. *American Mathematical Monthly* 65. 154–170.

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.

Lewis, Mike, Kenton Lee & Luke Zettlemoyer. 2016. LSTM CCG parsing. *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 221–231. San Diego CA: ACL.

Michaelis, Jens. 1998. Derivational minimalism is mildly context-sensitive. In Michael Moortgat (ed.), *Logical Aspects of Computational Linguistics (LACL'98)* (Lecture Notes in Artificial Intelligence 2014), 179–198. Berlin: Springer.

Michaelis, Jens. 2001. Transforming linear context-free rewriting systems into minimalist grammars. In Phillipe de Groote, Glyn Morrill & Christian Retoré (eds.), *Logical Aspects of Computational Linguistics (LACL'01)*, (Lecture Notes in Computer Science 2099), 228–244. Berlin: Springer.

Montague, Richard. 1973. The proper treatment of quantification in ordinary English. In Jaakko Hintikka, Julius Moravcsik & 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.

Morrill, Glyn. 1994. *Type-Logical Grammar*. Dordrecht: Kluwer.

Nchare, Abdoulaye Laziz. 2012. The Grammar of Shupamem. Ph.D. thesis, New York University.

Newmeyer, Frederick J. 2003. Grammar is grammar and usage is usage. *Language* 79. 682–707.

Nunes, Jairo. 2001. Sideward movement. *Linguistic Inquiry* 32. 303–344.

Partee, Barbara & Mats Rooth. 1983. Generalised conjunction and type ambiguity. In Rainer Bäuerle, Christoph Schwarze & Arnim von Stechow (eds.), *Meaning, Use, and Interpretation of Language*, 361–383. Berlin: de Gruyter.

Pesetsky, David. 1995. *Zero Syntax*. Cambridge, MA: MIT Press.

Pesetsky, David. 2017. Complementizer-*trace* effects. In Martin Everaert & Henk van Riemsdijk (eds.), *Blackwell Companion to Syntax*. New York: Wiley.

Pierrehumbert, Janet & Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip Cohen, Jerry Morgan & Martha Pollack (eds.), *Intentions in Communication*, 271–312. Cambridge, MA: MIT Press.

Pinker, Steven. 1979. Formal models of language learning. *Cognition* 7. 217–283.

Polinsky, Maria & Eric Potsdam. 2000. Backward control in Tsez. *19th West Coast Conference on Formal Linguistics*, 357–370.

Pollard, Carl. 1984. Generalized Context Free Grammars, Head Grammars, and Natural Languages. Ph.D. thesis, Stanford University.

Pollard, Carl & Ivan A. Sag. 1987. *Information-Based Syntax and Semantics, Vol. 1*, Stanford, CA: CSLI Publications.

Pollard, Carl & Ivan A. Sag. 1992. Anaphors in English and the scope of binding theory. *Linguistic Inquiry* 23. 261–303.

Richards, Norvin. 2016. *Contiguity Theory* (Linguistic Inquiry Monograph 73). Cambridge, MA: MIT Press.

Rizzi, Luigi. 1990. On the anaphor-agreement effect. *Rivista di Linguistica* 2. 27–42.

Roberts, Ian. 2001. Head movement. In Mark Baltin & Chris Collins (eds.), *Handbook of Contemporary Syntactic Theory*, 113–147. Oxford: Blackwell.

Rooth, Mats. 2016. Alternative semantics. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford Handbook of Information Structure*, 19–40. Oxford: Oxford University Press.

Ross, John Robert. 1967. Constraints on Variables in Syntax. Ph.D. thesis, MIT, published as Ross, 1986.

Ross, John Robert. 1970. Gapping and the order of constituents. In Manfred Bierwisch & Karl Heidolph (eds.), *Progress in Linguistics*, 249–259. The Hague: Mouton.

Ross, John Robert. 1986. *Infinite Syntax!*. Norton, NJ: Ablex.

Sadock, Jerrold. 1991. *Autolexical Syntax: A Theory of Parallel Grammatical Representations*. University of Chicago Press.

Salvati, Sylvain. 2011/2015. Mix is a 2-mcfl and the word problem in $\mathbb{Z}^2$ is captured by the io and the oi hierarchies. *Journal of Computer and System Sciences* 81. 1252–1277, circulated in 2011.
Selkirk, Elisabeth. 1990. On the nature of prosodic constituency. *Papers in Laboratory Phonology* 1. 179–200.
Selkirk, Elisabeth. 2011. The syntax-phonology interface. In John Goldsmith, Jason Riggle & Alan Yu (eds.), *The Handbook of Phonological Theory*, 2nd edn., 435–483, Blackwell.
Shieber, Stuart. 1985. Evidence against the context-freeness of natural language. *Linguistics and Philosophy* 8. 333–343.
Stabler, Edward. 2011. Computational perspectives on minimalism. In Cedric Boeckx (ed.), *Oxford Handbook of Linguistic Minimalism*, 617–641. Oxford University Press.
Steedman, Mark. 1985. Dependency and coordination in the grammar of Dutch and English. *Language* 61. 523–568.
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. 1996. *Surface Structure and Interpretation* (Linguistic Inquiry Monograph 30). Cambridge, MA: MIT Press.
Steedman, Mark. 2000. *The Syntactic Process*. Cambridge, MA: MIT Press.
Steedman, Mark. 2002. Plans, affordances, and combinatory grammar. *Linguistics and Philosophy* 25. 723–753.
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. The emergence of language. *Mind and Language* 32. 579–590.
Svenonius, Peter. 2007. 1…3–2. In Gillian Ramchand & 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. 1989. Bound variables in syntax: Are there any? In Renate Bartsch, Johan van Benthem & Peter van Emde Boas (eds.), *Semantics and Contextual Expression*, 295–318. Dordrecht: Foris.
Thomason, Richmond (ed.). 1974. *Formal Philosophy: Papers of Richard Montague*. New Haven, CT: Yale University Press.
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.
Vergnaud, Jean Roger. 1977/2006. Letter to Noam Chomsky and Howard Lasnik on "Filters and Control, " April 17, 1977. In Robert Freidin & Howard Lasnik (eds.), *Syntax: Critical Concepts in Linguistics*, 21–34. Oxford: Blackwell.
Vijay-Shanker, K. & David Weir. 1994. The equivalence of four extensions of context-free grammar. *Mathematical Systems Theory* 27. 511–546.
Williams, Edwin. 2003. *Representation Theory*. MIT Press.
Woods, William. 1970. Transition network grammars for natural language analysis. *Communications of the Association for Computing Machinery* 18. 264–274.
Woolford, Ellen. 1999. More on the anaphor agreement effect. *Linguistic Inquiry* 30. 257–287.
Zribi-Hertz, Anne. 1989. Anaphor binding and narrative point of view: English reflexive pronouns in sentence and discourse. *Language* 65. 695–727.