Lexically Specified Derivational Control in Combinatory Categorial Grammar

Jason Baldridge

Doctor of Philosophy

Institute for Communicating and Collaborative Systems
Division of Informatics
University of Edinburgh

2002

Abstract

This dissertation elaborates several refinements to the Combinatory Categorial Grammar (CCG) framework which are motivated by phenomena in parametrically diverse languages such as English, Dutch, Tagalog, Toba Batak, and Turkish. I present Multi-Modal Combinatory Categorial Grammar, a formulation of CCG which incorporates devices and category constructors from related categorial frameworks and demonstrate the effectiveness of these modifications both for providing parsimonious linguistic analyses and for improving the representation of the lexicon and computational processing.

I begin by introducing the various grammar frameworks which set the background for this dissertation and then discuss aspects of providing substantive universals for the theory of CCG. Most importantly, I lay out some of the components necessary for providing a theory of the lexicon, outlining previous approaches and suggesting directions forward. I then turn to a description of the syntactic extraction asymmetries found in English, Tagalog, and Toba Batak and the word order flexibility of Tagalog and Turkish, and discuss previous approaches to handling the data.

Having explicated the foundations and the linguistic motivations for the dissertation, I show how the multi-modal perspective on grammatical composition provided by the logical tradition of categorial grammar can be incorporated into CCG's rule-based approach. The enhanced resource-sensitivity of this perspective allows me to utilize an invariant rule component, controlling the applicability of the combinatory rules via lexical specification rather than with constraints on the rules themselves. This control is shown to be necessary for many aspects of English syntax, and I furthermore demonstrate that the multi-modal approach can improve upon existing CCG analyses for English and Dutch.

The second major development is a redefinition of categories and combinatory rules which relaxes the strict ordering inherent in categories that is normally assumed in categorial grammars. The manner in which this is done permits an intuitive account of local scrambling behavior without increasing the generative power of the system. Bounded long-distance scrambling is handled with the same mechanisms as CCG – type-raising and crossed composition rules. I furthermore show how the resource-sensitivity of the system effectively limits the permutative possibilities for some constructions in the otherwise quite free grammar of Turkish.

Having thus motivated and developed the MULTI-MODAL CCG system, I present

an account of syntactic extraction asymmetries in Tagalog and Toba Batak, showing how the categories of each language license only a subset of the resource-sensitive combinatory rules and thereby give rise to the observed asymmetries. This leads to a cross-linguistic perspective on the appearance of extraction asymmetries triangulated between English, Tagalog, and Toba Batak. We see that rigid languages like English and Toba Batak are forced to restrict permutativity and this leads naturally to certain arguments being inaccessible for extraction. Tagalog, with its more flexible word order, restricts associativity rather than permutativity, leading to robust asymmetries in a different manner.

Finally, I discuss the implementation of MULTI-MODAL CCG provided in Grok, highlighting the ways in which the properties of MULTI-MODAL CCG can be exploited to improve the use of CCG for parsing. In particular, the invariant rule component and modalities of MULTI-MODAL CCG make it possible to write hard-coded procedures that perform the work of the combinatory rules more efficiently than naïve implementations of the rules. Multiset categories also provide a more compact encoding of several rigid categories, and it is demonstrated that a few simple assumptions about unification significantly reduce the potential for them to induce indeterminacy in parsing. Finally, I discuss how the linguistic analyses presented in this dissertation have been encoded as grammars for Grok and improved in the process.

Altogether, this dissertation provides many formal, linguistic, and computational justifications for the central thesis that this dissertation puts forth — that an explanatory theory of natural language grammar can be based on a categorial grammar formalism which allows cross-linguistic variation only in the lexicon and has computationally attractive properties.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Jason Baldridge*)

Acknowledgements

My first and foremost thanks goes to my supervisor, Mark Steedman. From the time that I first started working with him in Philadelphia until today in Edinburgh, Mark has been a constant source of inspiration, both professionally and personally. The depth and breadth of his knowledge is truly staggering, and yet is equally matched by his personal warmth, care, and concern. Our often long meetings were absolutely essential for picking apart the threads of my research and sewing them back together into a coherent story. Mark was always patient with my meanderings and made sure to keep me on track with the issues that really mattered.

A big thanks also goes to Alex Lascarides, my second supervisor. Though she took on this role at quite a late stage, Alex quickly stepped up to the task of getting her head into my work. Her swift responses and detailed comments on drafts of this dissertation improved its content and argumentation immensely, and her enthusiasm was a major boost that helped me push through the final stages of the write-up.

I owe an immense debt of gratitude to my good friend and colleague Geert-Jan Kruijff. His appearance in Edinburgh almost three years ago greatly reinvigorated my enthusiasm for categorial grammars and grammatical theory at a time when I had begun to consider other research topics. Many of the ideas in this dissertation came from or grew out of our collaborations and many discussions. Both I and my work have benefitted greatly from Geert-Jan's input and insight, and I hope that this will continue to be the case for as long as we can both still stand in front of a whiteboard.

Gann Bierner, Julia Hockenmaier, and Mark McConville have been my friends, office-mates, and fellow categorial grammarians during my time at Edinburgh. It has been a true blessing to have such friendly and clever people to work with. Also, all three have given me extensive feedback on various drafts of this dissertation, for which I am very grateful. During the last several months, Mark has been particularly helpful in hashing out many of the details and high-level issues that have arisen during the write-up, and he has saved me from numerous errors.

Thanks are also due to my committee, Caroline Heycock, Ewan Klein and Dick Oehrle, as well as Cem Bozsahin, Robin Clark, Nobo Komagata, Frank Trechsel, and Aline Villavicencio, who all provided many helpful comments on various drafts and parts of the dissertation.

Many friends in Edinburgh have all helped make my stay here a wonderful and enriching experience — special thanks go to John, Laleh, Leslie, and Marcelo for extra support in the final phases of this dissertation. Also, despite the physical distance between the USA and Scotland, the members of my family – Justin, Fred, Lee, Lisa, Mary, Scott, and Sue — have managed to stay right by my side through both good times and difficult times. I am extremely fortunate to be surrounded by so many caring and loving people.

Finally, thanks to Fernanda, the love of my life, for everything. She has supported me throughout the course of my Ph.D. studies while never letting me lose sight of all the other important things in life.

Table of Contents

List of Figures

List of Acronyms and Abbreviations

AB	Ajdukiewicz/Bar-Hillel Calculus (Basic Categorial Grammar)
ACC	Accusative case
ANT	Antecedent-government
AV	Active Voice
CCG	Combinatory Categorial Grammar
CTL	Categorial Type Logic
DAT	Dative case
DGL	Dependency Grammar Logic
DV	Dative Voice
FGD	Functional Generative Description
GB	Government-Binding Theory
GEN	Genitive case
HLDS	Hybrid Logic Dependency Semantics
HPSG	Head-Driven Phrase Structure Grammar
HRC	Headless Relative Clause
IV	Instrumental Voice
LFG	Lexical-Functional Grammar
LNK	Linker

MRS	Minimal Recursion Semantics
NOM	Nominative case
OV	Objective Voice
PNM	Proper noun marker
SPEC	Specifier
SVO	Subject-Verb-Object
TAG	Tree-Adjoining Grammar
TDFS	Typed Default Feature Structure
TDL	Type Definition Language
UB-GCG	Unification-Based Generalized Categorial Grammar
VSO	Verb-Subject-Object
VOS	Verb-Object-Subject

Chapter 1

Introduction

The central thesis of this dissertation is that an explanatory theory of natural language grammar can be based on a categorial grammar formalism which allows cross-linguistic variation only in the lexicon and has computationally attractive properties. To back up this thesis, I present Multi-Modal Combinatory Categorial Grammar, a refinement of the Combinatory Categorial Grammar (CCG) framework (Steedman, 2000b), and apply it to several phenomena from typologically diverse languages.

There are three primary goals of this work: first, to boost the predictive power and explanatory force of the CCG theory by enhancing its sensitivity to the resources it manipulates and consumes; second, to provide new accounts for linguistic phenomena, improved adaptations of existing analyses, and cross-linguistic comparisons; and finally, to demonstrate some of the advantages of the resulting formulation of CCG for computational implementations. In this chapter, I outline and discuss the theses behind these goals and provide an overview of the dissertation.

1.1 Theses Proposed

The primary linguistic focus of this dissertation is a detailed examination of two core types of behavior in natural language grammar: syntactic extraction asymmetries and scrambling behavior. The former is characterized by situations in which particular arguments in a sentence are unsuitable targets for extraction in certain contexts; that is, it is not possible to use these arguments in forming questions, relative clauses, topicalized sentences and the like. For example, the well-known subject/object asymmetryof English appears in embedded clauses such as the following:

(1) a. *Brazil is the team$_i$ that John knew that t_i would beat Germany.

b. Germany is the team$_i$ that John knew that Brazil would beat t_i.

We see in (1a) that extraction of the subject from the embedded clause to form the relative clause is ungrammatical, whereas the object is accessible for extraction, as shown in (1b).[1] Unlike the situation with many island violations, there is nothing semantically incoherent about a relative clause such as that in (1a), and the grammar apparently disallows it for entirely syntactic reasons.

Perhaps the majority of languages exhibit greater flexibility in word order than English, some to a greater extent than others. For example, languages like Czech, Modern Greek, Russian, Turkish, Korean, and Tagalog all permit the same propositional content to be conveyed with multiple word orders in which the arguments of verbs can permute with respect to one another. Word order freedom is even greater in a language such as Warlpiri, which even permits parts of a single noun phrase to permute with other elements in a sentence.

Paying attention to these two core phenomena, we thus observe a basic tension in natural language grammar: sometimes it blocks perfectly sensible meanings from being expressed in certain ways which at first glance appear to be arbitrary, and sometimes it allows a single meaning – modulo information structure – to be expressed in multiple ways. As we will see with Tagalog, these restrictions and freedoms can co-exist in the grammar for a single language. This leads naturally to the question of how we can define a theory of grammar which is able to rule out examples such as (1a) whilst having the flexibility to permit multiple word orders in other contexts. The following thesis addresses this question.

Thesis 1. *A resource-sensitive approach which distinguishes multiple modes of grammatical composition is necessary to adequately characterize both the restrictiveness and the freedom exhibited by natural language grammars.*

The task of almost any formal system is to apply some group of operations to collections of structured objects, or resources, in order to determine some global properties about each collection. Resource-sensitivity is a notion that governs the *manner* in which a system's operations can utilize its resources: how often they may be used, how they can be assembled together to create larger structures, and how they can be reconfigured into other equivalent structures.

[1] It should be noted that I use the term *extraction* metaphorically, and that the traces shown in (1) are only used descriptively to highlight the canonical position of the extracted element.

Resource-sensitivity is nicely exemplified in Linear Logic (Girard, 1987), which counters the lack of discrimination inherent in many logical systems. As an example of a rather insensitive system, consider classical propositional logic, in which a single proposition may be used multiple times or wasted in proving a result. This can be seen in the following proof, where the proposition p is used twice (steps 6 and 7) and the proposition s is not used at all in proving t from the assumptions $p, p \rightarrow q, p \rightarrow r, (q \wedge r) \rightarrow t$, and s.

1.	p	*assumption*
2.	$p \rightarrow q$	*assumption*
3.	$p \rightarrow r$	*assumption*
4.	$(q \wedge r) \rightarrow t$	*assumption*
5.	s	*assumption*
6.	q	$1, 2, \rightarrow$
7.	r	$1, 3, \rightarrow$
8.	t	$4, 6, 7, \rightarrow$

Linear Logic gets a grip on resource consumption by employing a new implicational connective, $\multimap$. The rule associated with its use consumes the resource that is needed to prove the consequent; thus, the formula $p \multimap q$ is read as 'consume p yielding q'. After the rule for $\multimap$ is applied, the resource p is no longer available for further inferential steps, and therefore the above proof would not be valid (if $\rightarrow$ is replaced with $\multimap$). The proof also would not close since s is wasted.

These logical concerns have direct parallels in natural language grammar. Clearly, the multiplicity of linguistic material is important, since linguistic elements must generally be used once and only once during an analysis. Thus, we cannot ignore or waste linguistic material (2), nor can we indiscriminately duplicate it (3).

(2) a. *The coach smiled the ball. $\neq$ The coach smiled.

b. *The fans the coach cheered. $\neq$ The fans cheered.

(3) *Ronaldo passed the ball to. $\neq$ Ronaldo passed the ball to himself.

Actually, Linear Logic does permit a single resource to be used in multiple proof steps through the rule of Contraction and resources to be wasted via the rule of Weakening. However, there is an important difference in that these rules are restricted to apply only to resources of the appropriate types and are thus not globally available, unlike the case in Classical and Intuitionistic Logics. Resources themselves are designed so as to invoke only a subset of the available rules, making it possible to capture the distinction between resources such as money and love: money gets used up when it is given but love can be spread around infinitely. Linear Logic thus shifts to a perspective in which logics with different behaviors can co-exist and operate over the same set of premises without stepping on each other's feet.

The issue of resource consumption is a fundamental basis of resource-sensitivity, and most formal systems for natural language grammar do indeed respect the dictum that resources may not be indiscriminately wasted or used multiple times. However, a resource-sensitive system must also pay attention to the arrangement of its resources — how they are ordered linearly and hierarchically and the means by which they may have been combined. It is hardly surprising that linear order matters for natural language since it is the only aspect of syntax to which we have *direct* access. Formal grammar systems thus typically respect the importance of order and thereby ensure that sentences with the same lexical material but different word orders do not *necessarily* have the same analytical properties. Indeed, most syntacticians would be rather suspect of any system that could not differentiate the strings *Brazil defeated Germany* and *defeated Brazil Germany* in English.

To continue with the theme of viewing the properties of natural language grammar through a logical lens, we can consider building a logical system that uses *directional* implications and treats lexical items as proof terms, as is done in the Categorial Type Logic (CTL) tradition of categorial grammar (Morrill, 1994; Moortgat, 1997; Oehrle, to appear). The task of the grammar is then to find a proof that some set of axioms (in the form of items retrieved from the lexicon based on a given sentence or string) can be arranged in a manner that gives rise to the correct order and has the appropriate resultant properties (syntactic category, semantics, etc.).

CTL provides sensitivity to much more than linear order — it also permits the definition of multiple modes of grammatical composition which each have their own associated connectives. Operations which restructure the premises are keyed to particular modes so that they are not globally applicable. This is similar to the restricted use of Contraction and Weakening in Linear Logic, except that CTL allows a far wider range of rules to be defined in this manner. This allows one to use different kinds of implicational operators, each exhibiting its own unique behavior. Some might permit associative or permutative restructuring of the premises, while others might be much more limited in their capabilities. The premises themselves are constructed using these keyed connectives, and thereby endow the system with what Oehrle (to appear) calls *self-contained inferential control*. This means that instead of acting as absolute and global choices, parametric options regarding the way in which a set of premises can be restructured are selectively invoked via the appropriate type-declarations in the premises.

It is precisely this aspect of resource-sensitivity which is least obvious and which is the crux of **Thesis 1**. It allows us to get a precise handle on the ability of different parts of the syntax to have access to rules which induce associativity and permutativity, and this plays a major role in this dissertation's account of why syntactic extraction asymmetries arise. Having an explicit resource-management regime is also crucial for defining a system that is liberal enough to permit word order variation without needing *ad hoc* constraints to ensure that it does not fall into word order collapse. To exploit sensitivity to structural arrangement in CCG, I use the category constructors of CTL and redefine the rules of CCG so that they respect the modes of grammatical composition licensed by the categories they attempt to combine. It this manner, CCG gains the ability to utilize *lexically specified derivational control*, the implications of which are explored extensively in this dissertation.

Thesis 1 regards general architectural considerations that must be supplemented with investigations into specific patterns of natural language. One phenomenon which I consider in this dissertation is the kind of word order variation known as scrambling, which can occur in both local and long distance contexts. Locally scrambled arguments are those which are dependents of a single head that can permute with respect to one another, whilst an argument which has scrambled long distance is found not in the domain of its own head, but in that of another head. For example, the following Turkish sentences, in which the subject and object arguments can permute, convey the same basic propositional content:

(4) a. Ayse kitabi okuyor
Ayse book read

b. Kitabi Ayse okuyor
book Ayse read
Ayse reads the book.

With long distance scrambling, we find an argument of a lower clause appearing higher up, as in the following example:

(5) Esra'nın$_i$ Fatma [t_i gittiğini] biliyor.
Esra Fatma left know
As for Esra, Fatma knows that she left.

The question thus arises as to whether local and long distance scrambling should be accounted for with the same or different kinds of grammatical mechanisms. The position I take is that these are fundamentally different processes, as stated by the following thesis.

Thesis 2. *Local scrambling behavior results when heads subcategorize for their arguments in a manner which does not specify an explicit linear order of combination. Long-distance scrambling arises instead as a reflex of the interaction between lexical subcategorization and the rules made available by the resource-sensitive system.*

The generative power of grammatical formalisms is of interest in many traditions. The grammar of natural languages is recognized to require at least mildly context-sensitive power (Huybregts, 1984; Shieber, 1985), and it has been argued that long distance scrambling requires more power than this out of the competence grammar (Rambow, 1994; Hoffman, 1995). The multi-modal formulation of CCG provided in this dissertation remains mildly context-sensitive (like CCG) and is nonetheless able to handle long distance scrambling to the level that appears to correspond with the amount of scrambling which native speakers tolerate.

Thesis 3. *Mildly context-sensitive generative power is sufficient for handling long distance scrambling.*

By using a system with limited generative power, many linguistic predictions come for free since the system simply cannot perform a wide range of potential operations. Nonetheless, we of course should not be absolutely *stuck* with a mildly context-sensitive formalism if we do eventually need more power. The multi-modal formulation provides the means to increase the power of the system in a highly controlled fashion such that more powerful operations are used only by grammars that need them, only when they need them, and without precipitating a collapse in word order. Having said this, I strongly contend that we should result to more powerful formulations only with great skepticism in the face of overwhelming evidence for their necessity.

Categorial grammar in general is an extremely lexicalist tradition, but it is nonetheless standardly assumed in most categorial formalisms that variation between the grammars of different languages can occur not only in the lexicon, but also in the rules of combination. The multi-modal approach I adopt in this dissertation facilitates the creation of an invariant rule component that permits me to take a *fully* lexicalist position. This leads to the following thesis:

Thesis 4. *It is possible and desirable to define a framework in which all variation between grammars is specified in the lexicon.*

While I do not wish to claim that a parametric view on grammatical rules is inherently flawed, this thesis acts as a handcuff that leads to interesting observations about how a given lexicon will exploit a universal set of rules and simplifies the task of the grammar developer over formulations that assume that rules have a parametric nature. This dissertation demonstrates that a great deal of mileage can be obtained from a relatively small, invariant rule component that is sensitive to the grammatical resources that it combines. One of the most important effects of this perspective is that it permits a straightforward characterization of how syntactic extraction asymmetries arise cross-linguistically, as summarized in the following thesis:

Thesis 5. *Syntactic extraction asymmetries emerge in grammars which enforce limits on local and/or long distance scrambling by utilizing lexical categories that are inaccessible to syntactic operations which induce associativity and/or permutativity.*

The strategy of removing all variation from the rule-base places increased demands on the lexicon. It is thus important that generalizations can be expressed so that redundant information can be shared between categories. I therefore adopt the approach put forth by Villavicencio (2002) for permitting the categorial lexicon to be structured via an inheritance hierarchy of typed default feature structures (Pollard and Sag, 1987). Even without an invariant rule component, such a view of the lexicon is needed in CCG.

Parsing in CCG is generally construed as the application of a finite set of combinatory rules to the categories licensed by the input and created from previous applications of the rules. Lexical ambiguity is a major factor in reducing the speed of parsing. CCG has traditionally permitted its rules to be restricted in their applicability to only apply or not apply to certain categories. When using such rules in parsing, computational overhead is incurred as the input categories must be checked for compatibility with the restrictions.

Thesis 6. *The multi-modal formulation can be exploited to improve implementations of* CCG.

MULTI-MODAL CCG helps in two ways. First, it is possible to use one category in situations where otherwise several categories would be required (e.g. in languages with scrambling). Second, by providing modally aware formulations of the combinatory rules and disallowing restrictions on those rules, specialized implementations of the rules can be created which scan the input categories and fail much more quickly than is possible with restrictable CCG rules.

1.2 Overview

In Chapter 2, **Formal Foundations**, I begin by describing categorial grammars. Categorial grammars provide a type-driven perspective on natural language grammar that maintains a tight connection between syntactic and semantic composition. They are precisely defined, permit flexible surface constituency, are semantically transparent, and are at the center of a growing body of linguistic work. This chapter introduces the basic concepts behind categorial approaches, such as syntactic categories, semantic interpretation and rules of category combination, and it then gives greater detail for formalisms and traditions that the core categorial perspective has given rise to. Specifically, we consider Combinatory Categorial Grammar (CCG) (Steedman, 2000b), Multiset Combinatory Categorial Grammar (MULTISET-CCG) (Hoffman, 1995), and Categorial Type Logic (CTL) (Morrill, 1994; Moortgat, 1997; Oehrle, to appear), all of which play a major role in the approach developed in this dissertation. CCG provides the most important backdrop, whilst MULTISET-CCG and CTL point toward ways of relaxing and fine-tuning, respectively, grammatical composition in categorial grammar. The generative power of the various frameworks is then discussed with respect to the linguistic significance they attach to restricted generative capacity. The chapter finishes with a brief look at the dependency grammar tradition of Functional Generative Description (FGD) (Sgall *et al.*, 1986). Throughout the primarily linguistic parts of this dissertation, I make use of FGD's dependency relations for different kinds of arguments as a descriptive device to obviate the need to explicitly show logical forms for categorial derivations whilst demonstrating that the correct dependencies are obtained by the linguistic analysis.

I then turn to issues regarding the creation of a linguistic theory within a categorial approach in Chapter 3, **Substantive Universals**. I outline an initial approach to a theory of lexical categories based on distinctions made in the Government & Binding (Chomsky, 1981) and Minimalist traditions (Chomsky, 1994) and the typed feature structures of Head-driven Phrase Structure Grammar (Pollard and Sag, 1994). I also consider approaches for providing structure to the lexicon and expressing relationships between the objects stored within it. The final section of this chapter explicates the principles which guide the form of combinatory rules in CCG and discusses how restrictions can be placed on any given rule in standard CCG analyses.

In Chapter 4, **Linguistic Motivation**, I explicate the linguistic data which motivates the formal developments made in the dissertation. I begin with the English subject/object asymmetry in extraction from embedded clauses and discuss the CCG explanation of the asymmetry due to Steedman (1996). After this, I turn to the striking extraction asymmetries found in the Austronesian languages Tagalog and Toba Batak and describe some of the proposals that have been put forth to explain their distribution. Then, I discuss local and long distance scrambling in Turkish and different manners of handling such variability in categorial approaches, especially that of MULTISET-CCG (Hoffman, 1995). Finally, I show that there is a need for limitations on permutativity even in languages with a great deal of word order freedom like Turkish.

Chapter 5, **Modal Control in CCG**, explicates how CCG's resource-sensitivity can be boosted by incorporating the multi-modal perspective on grammatical composition familiar from CTL. I show how this provides fine-grained lexical control over the use of CCG's combinatory rules and permits me to dispense with restrictions on those rules. As such, I can claim a universal rule component for CCG and bring back the use of rules which were previously excluded from some grammars. Several aspects of English syntax are dealt with under this formulation and it is shown that many improvements can be made over the prior CCG analyses by using resource-sensitive rules. I then demonstrate that Steedman's analysis of Dutch (Steedman, 2000b) can be significantly improved by recasting it in MULTI-MODAL CCG. Following that, the next section develops the argument that the CCG rule set should be universally available by showing that certain combinatory rules are interconnected and cannot be arbitrarily activated or inactivated. Finally, I show that the multi-modal formulation of CCG has the same generative power as the original formalism.

In Chapter 6, **A Restricted Approach for Argument Scrambling**, the definition of MULTI-MODAL CCG is completed by including multisets in the category constructors and rules, based in part on developments by Hoffman (1995) for MULTISET-CCG. To motivate the use of multisets in categories, I demonstrate their use for local argument scrambling in Turkish. The need for resource-sensitivity in an approach which uses multisets is demonstrated with respect to limits on permutativity for some constructions in Turkish and for English phrasal verbs and adverb placement. Thereafter, I show how the system deals with long distance argument scrambling without suffering from some of the overgeneration that the less discrimating and more powerful MULTISET-CCG produces. Finally, I show that MULTI-MODAL CCG as defined in this chapter is mildly context-sensitive, like CCG.

Having thus motivated and developed MULTI-MODAL CCG, Chapter 7, **Syntactic Extraction Asymmetries in Tagalog and Toba Batak**, demonstrates how the modal control available in the grammar combined with the proposed categories conspire to explain the observed asymmetries. It is also shown how MULTI-MODAL CCG permits a simple account of local scrambling in these languages – not only without confounding the account of asymmetries, but at times even supporting it, in contrast with some previous approaches. The analysis given in Chapter 7 provides the most extensive account of Tagalog's asymmetries to date. In combination with the analysis of the English subject/object asymmetry by Steedman (1996) and a further analysis in Chapter 7 of asymmetries in Toba Batak, I explicate a cross-linguistic characterization of the appearance of asymmetries.

Chapter 8, **Implementation of Multi-Modal CCG**, begins by reviewing previous work in creating grammars and parsers for CCG, followed by a discussion of how I have adapted the Grok system (Hockenmaier *et al.*, 2001; Bierner, 2001) to support the data structures and properties of MULTI-MODAL CCG. Grammars have been implemented based on the linguistic analyses given for English, Dutch, Turkish, and Tagalog, in this dissertation, and I highlight some of the properties of these grammars and discuss how the the process of developing them not only ensured the correctness of the analyses but also led to interesting linguistic observations in some cases. Finally, an appendix to the chapter is provided to show some example interactions with Grok's text interface.

Chapter 9 concludes by reiterating the contributions of this dissertation and emphasizing how its different parts coordinate to support the theses laid out in this introduction.

In summary, this dissertation provides:

- Suggestions for and discussions of **substantive universals** from the categorial perspective.

- **Multi-Modal CCG**, a formalism which has a strict resource-management regime and permits variation only in the lexicon.

- **Linguistic application** of this formalism to English, Dutch, Turkish, Tagalog, and Toba Batak, including a cross-linguistic analysis of the appearance of syntactic extraction asymmetries in English, Tagalog, and Toba Batak.

- **Computational implementation** of the developed formalism.

Chapter 2

Formal Foundations

The task of any theory of natural language grammar is to provide the means to mediate the relation between the properties of sub-parts of a linguistic expression and the global properties associated with it. How this is achieved of course varies widely between different theories depending on the kinds of mechanisms, data structures, and constraints of which they avail themselves. This dissertation draws heavily on a number of *compositional* approaches to natural language syntax — formalisms for which the global properties associated with a linguistic expression are entirely determined by the correlative properties of its component parts. In such systems, linguistic expressions are generally considered to be multi-dimensional structured *signs* that contain phonological/orthographic, syntactic and semantic specifications for the expression (Pollard and Sag, 1994; Morrill, 1994; Oehrle, to appear; Steedman, 2000b). Each formalism provides its own way of defining basic signs and makes some set of operations available in order to combine them and thereby create more complex signs which monotonically incorporate the content of their subparts.

The most important compositional tradition for this dissertation is the family of categorial grammar formalisms. In these formalisms, grammatical composition takes the form of an inference problem in which the sub-signs of a linguistic expression act as premises in a process which deduces the concluding sign that contains the global properties associated with the expression (Oehrle, to appear). Different categorial approaches take this connection with logical deduction to varying degrees, with some being only partial systems of implicational reasoning and others providing a fully logical system.

Categorial grammar formalisms typically adopt an extremely lexicalist position in which nearly all grammatical information is contained within the entries of the lexicon. Words are assigned categories which may combine with others in a semantically transparent manner through a small set of rules. This lexicalist view on the objects of grammatical inquiry and adherence to semantic transparency is a starting point which leads to a variety of categorial perspectives on natural language syntax. It furthermore ties categorial grammar to the tradition of dependency grammar, which also focuses on the way in which patterns of semantic linkage hold a sentence together, rather than segmenting sentences according to analytic patterns such as phrase structure (Wood, 1993). Kruijff (2001) carries these connections to their logical conclusion by providing a formalization of dependency grammar which is driven by a resource-sensitive categorial system.

This chapter provides introductions to several categorial grammar formalisms which all play crucial roles in the developments of this dissertation. Beginning in §2.1 with pure categorial grammar and discussing some of its strengths and weaknesses, I then discuss Combinatory Categorial Grammar (CCG), which addresses some of the limitations of pure categorial grammar by defining rules of combination based on a small set of combinators from combinatory logic, and Multiset Combinatory Categorial Grammar, an extension of CCG designed to allow greater flexibility in word order. Then I discuss Categorial Type Logic (CTL), which departs from pure categorial grammar's rule-based approach by providing a powerful system of grammatical inference. Finally, the issue of the generative capacity of these various frameworks is then explored.

The chapter finishes with a brief description of notions from dependency grammar that are utilized thoughout the dissertation. Readers familiar with the material covered in this chapter can of course skip through much of it, but are nonetheless encouraged to observe some of the notational conventions that I assume for reducing the size of linguistic descriptions.

2.1 The AB Calculus

Pure categorial grammar is the common starting point of all categorial grammar formalisms. It is the product of the directional adaptation by Bar-Hillel (1953) of Ajdukiewicz's calculus of syntactic connection (Ajdukiewicz, 1935) and is thus generally referred to as the AB calculus. The discussion is brief, and the reader is referred to Wood (1993) and Steedman (2000b) for more detailed discussions of the material covered here.

2.1.1 Categories

The grammatical objects of the AB system are categories which may be either atomic elements or (curried) functions which specify the canonical linear direction in which they seek their arguments.

Definition 1 (AB categories). *Given $\mathcal{A}$, a finite set of atomic categories, the set of categories $\mathcal{C}$ is the smallest set such that:*

- $\mathcal{A} \subseteq \mathcal{C}$
- $(\mathsf{X}\backslash\mathsf{Y}), (\mathsf{X}/\mathsf{Y}) \in \mathcal{C}$ *if* $\mathsf{X}, \mathsf{Y} \in \mathcal{C}$

Lexical entries are specified by pairing words with categories via the $\vdash$ operator. Some simplified example entries are given below:

(6) a. *Brazil* $\vdash \mathsf{np}$

b. *Germany* $\vdash \mathsf{np}$

c. *defeated* $\vdash (\mathsf{s}\backslash\mathsf{np})/\mathsf{np}$

I use Steedman's argument rightmost notation for categories. It is common in the Lambek tradition to use an alternative notation in which arguments sought to the left are placed on the left of the functor. Under this notation, the entry for *defeated* becomes $(\mathsf{np}\backslash\mathsf{s})/\mathsf{np}$.

A flat representation of the lexicon such as that depicted in (6) is of course a simplification. While it is concise and useful when presenting linguistic analyses of particular phenomena, it is inadequate for specifying large grammars such as those used in computational settings. See §3.2 for a discussion about proposals for using inheritance to efficiently and concisely define the lexicon and reduce redundancy in lexical entries.

2.1.2 Rules

In the AB system, categories may combine through two directionally sensitive rules of functional application, presented below. The application rules are simply two instantiations of basic functional application, particularized to respect the direction in which the functor category expects its argument.

(7) *Functional application*

a. $\mathsf{X/Y} \quad \mathsf{Y} \;\Rightarrow\; \mathsf{X}$ $\qquad (>)$

b. $\mathsf{Y} \quad \mathsf{X\backslash Y} \;\Rightarrow\; \mathsf{X}$ $\qquad (<)$

In binary rules such as this, I will refer to the functor category (e.g. $\mathsf{X/Y}$) as the *primary* input to the rule and the other argument as the *secondary* input.

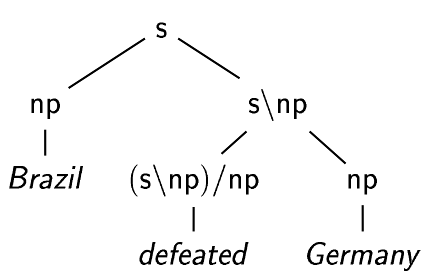
These rules can also be viewed as rules of a context-free grammar written in the accepting, rather than producing, direction (Steedman, 2000b), a relationship which forms the basis for the proof that AB is context-free (Bar-Hillel *et al.*, 1964).

In giving a derivation for a string ω, we line up the lexical items of ω, introduce their associated categories from the lexicon, and proceed by applying the rules to adjacent pairs of categories. Once a pair of categories have been combined, only the category resulting from the combination is visible for further derivational steps. Each step is annotated by underlining the combining categories and labeling the underline with the symbol for the rule used. Unlabeled underlines indicate introduction of a category from the lexicon. An example derivation is given below for the string *Brazil defeated Germany* using the lexical entries given in (6).

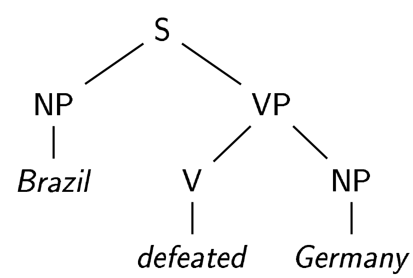
(8)
$$\begin{array}{ccc}
\textit{Brazil} & \textit{defeated} & \textit{Germany} \\
\hline
\mathsf{np} & \mathsf{(s\backslash np)/np} & \mathsf{np} \\
\cline{2-3}
 & \multicolumn{2}{c}{\mathsf{s\backslash np}} \quad > \\
\hline
\multicolumn{3}{c}{\mathsf{s}} \quad <
\end{array}$$

This derivation can be seen as a phrase structure tree turned upside-down, as shown in (9), in which the categorial analysis (9a) is clearly isomorphic to a standard phrase structure analysis (9b).

(9) a.



b.



Despite such a correspondence, the categories labeling the nodes of the tree in (9a) are much more informative than the atomic symbols of a tree produced by a context-free grammar such as (9b). Subcategorization is directly encoded in functor categories rather than through the use of new symbols such as $\mathsf{V}_{intrans}$, V_{trans} and $\mathsf{V}_{ditrans}$. Furthermore, there is a systematic correspondence between notions such as *intransitive* and *transitive* — after the transitive category $\mathsf{(s\backslash np)/np}$ consumes its object argument, the resulting category $\mathsf{s\backslash np}$ is exactly the same as that of an intransitive verb. This comes as a result of the way lexical categories are defined in combination with the universal rules of functional application. See Oehrle (to appear) for a deductive explanation within the Lambek framework of the relationship between categories and phrase structure labels.

The type-driven nature of categorial grammar provides a natural explanation for basic coordination phenomena. By assuming that the categories for *and* are functions over two categories of the same type into another category of that type, which can be represented schematically as in (10), we immediately predict the distribution in (11).[1]

(10) $\textit{and} \vdash \mathsf{(X\backslash X)/X}$

(11) a. Brazil played and defeated Germany.
b. *Brazil played Germany and defeated.
c. Brazil defeated Germany and won the World Cup.
d. Brazil defeated Germany and celebrated.
e. John bought and gave his family tickets to the World Cup.
f. Brazil defeated Turkey and Germany.

For example, (11a) receives the following derivation:

(12)
$$\dfrac{\textit{Brazil}}{\mathsf{np}} \quad \dfrac{\textit{played}}{\mathsf{(s\backslash np)/np}} \quad \dfrac{\textit{and}}{\mathsf{(((s\backslash np)/np)\backslash((s\backslash np)/np))/((s\backslash np)/np)}} \quad \dfrac{\textit{defeated}}{\mathsf{(s\backslash np)/np}} \quad \dfrac{\textit{Germany}}{\mathsf{np}}$$

$$\dfrac{\mathsf{(((s\backslash np)/np)\backslash((s\backslash np)/np))/((s\backslash np)/np)} \quad \mathsf{(s\backslash np)/np}}{\mathsf{((s\backslash np)/np)\backslash((s\backslash np)/np)}}>$$

$$\dfrac{\mathsf{(s\backslash np)/np} \quad \mathsf{((s\backslash np)/np)\backslash((s\backslash np)/np)}}{\mathsf{(s\backslash np)/np}}<$$

$$\dfrac{\mathsf{(s\backslash np)/np} \quad \mathsf{np}}{\mathsf{s\backslash np}}>$$

$$\dfrac{\mathsf{np} \quad \mathsf{s\backslash np}}{\mathsf{s}}<$$

[1]It is also common to use the following syncategorematic ternary rule for coordination rather than a category such as (10):

a. $\mathsf{X} \quad \mathsf{CONJ} \quad \mathsf{X} \Rightarrow_{\Phi} \quad \mathsf{X}$ $\quad (\mathit{Coordination}(\Phi))$

Using such a rule avoids certain undesirable combinations available with the category in (10) when categorial systems more powerful than AB are employed. The approach I develop in Chapter 5 enables coordination to be dealt with on a category rather than rule basis without the undesirable consequences suffered previously. See §5.2 for more details.

On the other hand, (11b) is blocked automatically because *played Germany* has the category of an intransitive verb and thus cannot coordinate with the transitive verb *defeated*:

(13)
$$\begin{array}{ccccc}
\textit{*Brazil} & \textit{played} & \textit{Germany} & \textit{and} & \textit{defeated} \\
\mathsf{np} & (\mathsf{s}\backslash\mathsf{np})/\mathsf{np} & \mathsf{np} & (((\mathsf{s}\backslash\mathsf{np})/\mathsf{np})\backslash((\mathsf{s}\backslash\mathsf{np})/\mathsf{np}))/((\mathsf{s}\backslash\mathsf{np})/\mathsf{np}) & (\mathsf{s}\backslash\mathsf{np})/\mathsf{np} \\
 & \multicolumn{2}{c}{\overline{\mathsf{s}\backslash\mathsf{np}}^{\,>}} & \multicolumn{2}{c}{\overline{((\mathsf{s}\backslash\mathsf{np})/\mathsf{np})\backslash(\mathsf{s}\backslash\mathsf{np}/\mathsf{np})}^{\,>}} \\
 & \multicolumn{4}{c}{\overline{\qquad\qquad\qquad\qquad\qquad\qquad}^{\,*}}
\end{array}$$

Convention 1 (Abbreviated derivations). *To avoid writing down derivations (such as (13)) in which most of the steps are obvious and only one critical step needs to be highlighted, it is common to use a one line representation that groups relevant words together and shows the categories associated with those substrings. For example, derivation (13) can be equivalently shown as follows:*

(14) *Brazil [played Germany]$_{\mathsf{s}\backslash\mathsf{np}}$ and [defeated]$_{(\mathsf{s}\backslash\mathsf{np})/\mathsf{np}}$.

As a further abbreviation strategy, some steps in full derivations are omitted when they are either entirely obvious or irrelevant for the purpose of the particular derivation.

(11e) demonstrates the coordination of a ditransitive verb which has consumed its indirect object with a transitive verb:

(15) John [bought]$_{(\mathsf{s}\backslash\mathsf{np})/\mathsf{np}}$ and [gave his family]$_{(\mathsf{s}\backslash\mathsf{np})/\mathsf{np}}$ tickets to the World Cup.

With its structured categories, AB thus provides a more naturally consistent picture of the kinds of constituents which are available for coordination than context-free grammar, which uses only atomic labels. The potential of an AB grammar is packed into its categories, which are combined in a logical manner via the rules of function application. Notice the similarity of the rules in (7) with the rule of Modus Ponens from propositional logic:

(16) $Y \rightarrow X,\ Y \ \vdash\ X$

As we will see later, the Lambek tradition of categorial grammar takes this connection to its extreme and provides a *fully* logical formulation of categorial grammar that permits hypothetical reasoning.

2.1.3 Associating Interpretations with Categories

One of the defining characteristics of all categorial frameworks is that syntactic and semantic combination are assumed to proceed in parallel. Every syntactic step thus has a semantic reflex, a quality which has endeared categorial grammars to logicians for decades (e.g. Montague (1973)). This section provides a very brief explication of one way in which this linkage can be implemented.

Since Montague, a common choice for representing semantic expressions with categorial grammars is the λ-calculus. To incorporate such terms into lexical entries, we extend the notation used in (6) with the ':' operator, giving entries in the format [*word* $\vdash$ category : **semantics**]. Our simple sample lexicon now appears as follows:

(17) a. *Brazil* $\vdash \mathsf{np} : \mathbf{Brazil}$

b. *Germany* $\vdash \mathsf{np} : \mathbf{Germany}$

c. *defeated* $\vdash (\mathsf{s}\backslash\mathsf{np})/\mathsf{np} : \lambda x.\lambda y.\mathbf{defeated}(y, x)$

The λ-operators λx and λy in $\lambda x.\lambda y.\mathbf{defeated}(y, x)$ are paired with the outermost np and innermost np, respectively, of the category $(\mathsf{s}\backslash\mathsf{np})/\mathsf{np}$. The semantics given for these categories (and the categories themselves) are of course vastly simplified, but they are nonetheless sufficient to convey the manner in which semantics keeps pace with syntax in categorial grammar.

The application rules must be expanded to include specifications for the interpretations associated with each input category and their combination in the result.

(18) *Functional application*

a. $\mathsf{X}/\mathsf{Y} : f \quad \mathsf{Y} : a \quad \Rightarrow \quad \mathsf{X} : fa$ (>)

b. $\mathsf{Y} : a \quad \mathsf{X}\backslash\mathsf{Y} : f \quad \Rightarrow \quad \mathsf{X} : fa$ (<)

With the entries in (17), we can now produce a derivation with a simple interpretation as a result. The result of applying the semantic functors to their arguments is shown in (19).

(19)
$$\dfrac{\dfrac{\textit{Brazil}}{\mathsf{np} : \mathbf{Brazil}} \quad \dfrac{\dfrac{\textit{defeated}}{(\mathsf{s}\backslash\mathsf{np})/\mathsf{np} : \lambda x.\lambda y.\mathbf{defeat}(y, x)} \quad \dfrac{\textit{Germany}}{\mathsf{np} : \mathbf{Germany}}}{\mathsf{s}\backslash\mathsf{np} : \lambda y.\mathbf{defeat}(y, \mathbf{Germany})}>}{\mathsf{s} : \mathbf{defeat}(\mathbf{Brazil}, \mathbf{Germany})}<$$

Note that the β-reductions have been performed implicitly in this derivation.

It is possible to bind arguments in semantic representations using mechanisms other than those of the λ-calculus. For example, Zeevat *et al.* (1987) and Zeevat (1988) employed unification for this purpose. The use of the λ-calculus as the representation framework is also optional since interpretations can instead be encoded with other representation languages such as Indexed Languages (Zeevat *et al.*, 1987; Zeevat, 1988), Hybrid Logic Dependency Semantics (Kruijff, 2001) or Minimal Recursion Semantics (Copestake *et al.*, 1999, 2001; Copestake, 2002). See Baldridge and Kruijff (2002) for an approach which integrates CCG with Hybrid Logic Dependency Semantics and Villavicencio (2002) for one which uses Minimal Recursion Semantics within the context of Unification-Based Generalized Categorial Grammar.

Due to the primarily syntactic nature of the dissertation, I will suppress semantics in most of the discussion. Even so, that does not mean that interpretations cannot be provided for the syntactic analyses; to the contrary, every category is designed so that the grammar will permit interpretations to be built in parallel.

2.1.4 Limitations of the AB Calculus

Though a great deal of linguistic phenomena can be handled with the AB apparatus, constructions such as relativization demonstrate an unacceptable deficiency in the system as it stands. Consider the following relative clauses:

(20) a. team$_i$ that t_i defeated Germany

b. team$_i$ that Brazil defeated t_i

Under almost all categorial accounts, categories for lexical items such as relativizers and *wh*-items range over functions missing one of their arguments to produce post-nominal modifiers. Thus, the subject and object extracting categories for *that* are given as in (21a) and (21b), respectively.

(21) a. *that* $\vdash$ $(\mathsf{n}\backslash\mathsf{n})/(\mathsf{s}\backslash\mathsf{np})$

b. *that* $\vdash$ $(\mathsf{n}\backslash\mathsf{n})/(\mathsf{s}/\mathsf{np})$

(20a) can be dealt with straightforwardly, because the string *defeated Germany* has the category $\mathsf{s}\backslash\mathsf{np}$ of an intransitive verb, which is precisely the category required by (21a).

(22) team [that]$_{(\mathsf{n}\backslash\mathsf{n})/(\mathsf{s}\backslash\mathsf{np})}$ [defeated Germany]$_{\mathsf{s}\backslash\mathsf{np}}$

On the other hand, (20b) cannot be derived even given the object extracting category (21b) because there is no way in AB to combine the subject with the verb until the verb has consumed its object. One possible solution is to assign the extra category $\mathsf{(s/np)\backslash np}$ to transitive verbs and thereby provide the following derivation for (20b):

(23)
$$\dfrac{\textit{team}}{\mathsf{n}} \qquad \cfrac{\dfrac{\textit{that}}{\mathsf{(n\backslash n)/(s/np)}} \qquad \cfrac{\dfrac{\textit{Brazil}}{\mathsf{np}} \quad \dfrac{\textit{defeated}}{\mathsf{(s/np)\backslash np}}}{\mathsf{s/np}}{<}}{\mathsf{n\backslash n}}{>}$$

While this strategy permits an analysis for this case, it requires categorial ambiguity of a kind that is generally avoided by categorial grammarians because it relies on stipulations rather than formal universals provided by the system. Also, the strategy quickly breaks down when we consider unbounded relativization. Any amount of intervening material can come between the relativizer and the function it seeks, as shown in (24).

(24)
a. team that I thought that Brazil defeated.
b. team that I thought that you said that Brazil defeated.
c. team that I thought that you said that John knew ... that Brazil defeated.

The category of *that* as a complementizer is generally assumed to be $\mathsf{s'/s}$. With the AB system there is no way of deriving the category $\mathsf{s/np}$ (required by the relativizer) for the string *I thought that Brazil defeated*, even if we use the alternative category for *defeated* that was utilized in (23).

(25) team that I $[\text{thought}]_{\mathsf{(s\backslash np)/s'}}$ $[\text{that}]_{\mathsf{s'/s}}$ $[\text{Brazil defeated}]_{\mathsf{s/np}}$

To capture such a relative clause, we would need further categories for both *thought* and *that* which explicitly pass up the extracted element, as shown in the following derivation:

(26)
$$\textit{team} \qquad \cfrac{\dfrac{\textit{that}}{\mathsf{(n\backslash n)/(s/np)}} \qquad \cfrac{\cfrac{\dfrac{\textit{I}}{\mathsf{np}} \quad \dfrac{\textit{thought}}{\mathsf{((s/np)/(s'/np))\backslash np}}}{\mathsf{(s/np)/(s'/np)}}{<} \qquad \cfrac{\dfrac{\textit{that}}{\mathsf{(s'/np)/(s/np)}} \quad \dfrac{\textit{Brazil defeated}}{\mathsf{s/np}}}{\mathsf{s'/np}}{>}}{\mathsf{s/np}}{>}}{\mathsf{n\backslash n}}{>}$$

Another problematic example comes from so-called *non-peripheral* extractions, such as the following:

(27) team$_i$ that Brazil defeated t_i yesterday

Here, the extraction site is between the verb and the adverb, so we must find some means to combine these two elements. With the standardly assumed categories, no derivation is available:

(28) team that Brazil [defeated]$_{(\mathsf{s}\backslash\mathsf{np})/\mathsf{np}}$ [yesterday]$_{(\mathsf{s}\backslash\mathsf{np})\backslash(\mathsf{s}\backslash\mathsf{np})}$

Categorial ambiguity could of course be used to overcome this deficiency — simply give all adverbs another category that takes the transitive category as input and produces the transitive category as output. However, here we see how the kind of categorial ambiguity suggested above begins to cause greater havoc with the grammar — we need not only the category $((\mathsf{s}\backslash\mathsf{np})/\mathsf{np})\backslash((\mathsf{s}\backslash\mathsf{np})/\mathsf{np})$, but also $((\mathsf{s}/\mathsf{np})\backslash\mathsf{np})\backslash((\mathsf{s}/\mathsf{np})\backslash\mathsf{np})$, and this would be necessary for other verbal categories as well (e.g. ditransitives). The situation gets even worse if we consider the many categories that would be necessary for pre-verbal adverbs to permit relative clauses like *that Brazil easily defeated yesterday.*

Another problem comes with phrasal verbs, in which a particle can shift with respect to a direct object, as in the following examples:

(29) a. Marcos picked the ball up.

b. Marcos picked up the ball.

With AB, it would be necessary to provide two categories to handle this variation. Also, with respect to extraction, yet another category for object extraction is needed.

These are just a few demonstrations of how the inherent limitations of AB induce considerable categorial ambiguity if we are to provide adequate coverage. The essential reason that AB cannot handle object extraction and unbounded dependencies with single category assignments is that it is completely non-associative and non-permutative. It could be argued that throwing a large number of categories at the problem is actually a fine strategy and that we should be content with the utter simplicity of the system itself. However, the inadequacy of AB is not simply an aesthetic issue. For one thing, there is no systematic relationship predicated between the multiple categories despite the fact that they are able to provide alternative derivations for the same strings. This is bound to create problems for a learning theory that uses categories as a data structure. A more damaging point is that AB's generative power is simply too weak. As shown by Huybregts (1984) and Shieber (1985), systems with only context-free generative power are unable to capture crossing dependencies in languages such as Dutch and Swiss German. As we will see later, these crossing dependencies require a grammar to license a limited degree of permutativity which comes into play for constructions in a wide variety of languages, including English.

The AB calculus has been extended in two primary ways to create systems that can provide the flexibility needed to capture both the associativity and permutativity that natural language apparently demands. The first strategy is to augment the rules of AB with further rule schemas that have greater combinatorial potential, whilst the other recasts AB as a logical system with full implicational reasoning capabilities. The rule-based strategy is exemplified by early work such as Geach (1972) and Ades and Steedman (1982), the main manifestations of which in the present day are CCG and Unification-Based Generalized Categorial Grammar. The logical tradition initiated by Lambek (1958) has evolved in a number of directions which I shall generally refer to as Categorial Type Logic (CTL), following Moortgat (1997). I begin with CCG.

2.2 Combinatory Categorial Grammar

Just as AB's syntactic rules of functional application correspond to the semantic operation of functional application as shown in (18), we might consider creating syntactic correlates for other semantic operations. In particular, several of the combinators which Curry and Feys (1958) use to define the λ-calculus and applicative systems in general are of considerable syntactic interest (Steedman, 1988). As the basis for defining new rule schemas to extend the AB base, CCG employs three combinators: composition ($\mathbf{B}$), type-raising ($\mathbf{T}$), and substitution ($\mathbf{S}$). The relationships of these combinators to terms of the λ-calculus are defined by the following equivalences (Steedman, 2000b):

(30) a. $\mathbf{B} f g \equiv \lambda x. f(g\,x)$

b. $\mathbf{T} x \equiv \lambda f. f\,x$

c. $\mathbf{S} f g \equiv \lambda x. f x (g\,x)$

The combinator **B** composes a function f with its argument g before g has applied to its own argument. The result is a new function that applies its argument to the embedded function g. The combinator **T** turns an argument x into a function whose argument f is a function that x applies itself to. Finally, the combinator **S** is quite similar to **B**, except that the function it creates applies its argument x to both f and g.

These combinators give rise to new classes of rules, the forms of which are governed by the following principle (Steedman, 2000b):

(31) *The Principle of Combinatory Type Transparency*
All syntactic combinatory rules are type-transparent versions of one of a small number of simple semantic operations over functions.

Each rule type is comprised of several directionally specific instantiations that allows CCG to systematically provide derivational capabilities that would either require excessive stipulations or be uncapturable under AB. See §2.2.5 for a discussion on the kinds of rules which the principle does not allow.

In following sections, I introduce several rules which are formed from the combinators **B**, **T**, and **S** and the Principle of Combinatory Type Transparency. I begin with the composition combinator **B** and its order-preserving syntactic rule correlates.

2.2.1 Harmonic Composition

There are a number of reasons to believe that the grammar of English is associative at some levels. In particular, evidence from extraction and coordination showed that we must assume that a grammatical system must be capable of deriving constituents for word sequences that do not correspond directly to standard notions of constituency. A further example that cannot be handled elegantly with the limited apparatus of AB involves the coordination of verbal complexes such as that in (32).

(32) Brazil will meet and should defeat China.

A standard analysis of modal verbs is that they are functions from intransitive verb phrases into intransitive verb phrases. However, for this coordination to proceed, the modal verbs *will* and *should* must combine with *meet* and *defeat* respectively before the latter two may consume their object argument. Because the only rule available in AB is functional application (7), there is no way of performing this combination.

(33) $[\text{will}]_{\mathsf{(s\backslash np)/(s\backslash np)}}\ [\text{meet}]_{\mathsf{(s\backslash np)/np}}$

Given these categories, there is a simple rule corresponding to the combinator **B** which allows modals to combine with transitive verbs directly by composing the two functions:

(34) *Forward composition* (>**B**)

$$\mathsf{X/Y}\quad \mathsf{Y/Z}\quad \Rightarrow_{\mathbf{B}}\quad \mathsf{X/Z}$$

It should be pointed out that *functional* composition should not be confused with the notion of *grammatical* composition employed in much work in the Lambek tradition, such as Oehrle (to appear).

The derivation of (32) then proceeds as in (35), where applications of the composition rule are indexed with >**B**. For now, I will also use the schematic category $\mathsf{(X\backslash X)/X}$ for the coordinator and collapse the two steps for it applying to its arguments into one step indexed by <Φ>.

(35)

$$
\begin{array}{ccccccc}
\textit{Brazil} & \textit{will} & \textit{meet} & \textit{and} & \textit{should} & \textit{defeat} & \textit{China} \\
\overline{\mathsf{np}} & \overline{\mathsf{(s\backslash np)/(s\backslash np)}} & \overline{\mathsf{(s\backslash np)/np}} & \overline{\mathsf{(X\backslash X)/X}} & \overline{\mathsf{(s\backslash np)/(s\backslash np)}} & \overline{\mathsf{(s\backslash np)/np}} & \overline{\mathsf{np}} \\
 & \multicolumn{2}{c}{\overline{\qquad\qquad\qquad\qquad}{>}\mathbf{B}} & & \multicolumn{2}{c}{\overline{\qquad\qquad\qquad\qquad}{>}\mathbf{B}} & \\
 & \multicolumn{2}{c}{\mathsf{(s\backslash np)/np}} & & \multicolumn{2}{c}{\mathsf{(s\backslash np)/np}} & \\
 & \multicolumn{5}{c}{\overline{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}{<}\Phi{>}} & \\
 & \multicolumn{5}{c}{\mathsf{(s\backslash np)/np}} & \\
 & \multicolumn{6}{c}{\overline{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}{>}} \\
 & \multicolumn{6}{c}{\mathsf{s\backslash np}} \\
\multicolumn{7}{c}{\overline{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}{<}} \\
\multicolumn{7}{c}{\mathsf{s}}
\end{array}
$$

The semantic component of a rule such as (34) is determined by the combinator **B** and the Principle of Combinatory Type Transparency (31) to be the following:

(36) *Forward composition* (>**B**)

$$\mathsf{X/Y} : f\quad \mathsf{Y/Z} : g\quad \Rightarrow_{\mathbf{B}}\quad \mathsf{X/Z} : \lambda x.f(gx)$$

The variable introduced by the rule must of course be fresh so that the λ-operator does not bind a variable occuring in the functions f and g. With these semantics, this rule guarantees that the interpretation built for a sentence like *Brazil should defeat China* is the same for both the derivation that uses only the application rules, shown in (37) and the one which composes *should* with *defeat* before applying to the object, shown in (38). Note that the interpretations shown here are not intended as serious proposals for the meaning of the various lexical items; rather, they are only simple interpretations that suffice to demonstrate the semantic invariance of the combinatory rules.

(37)
$$
\begin{array}{cccc}
\textit{Brazil} & \textit{should} & \textit{defeat} & \textit{China} \\
\hline
\mathsf{np}:\mathbf{Brazil} & \begin{array}{c}(\mathsf{s\backslash np})/(\mathsf{s\backslash np}) \\ :\lambda P.\lambda x_1.\mathbf{should}(Px_1)\end{array} & \begin{array}{c}(\mathsf{s\backslash np})/\mathsf{np} \\ :\lambda x_2.\lambda x_3.\mathbf{defeat}(x_3,x_2)\end{array} & \mathsf{np}:\mathbf{China} \\
\cline{3-4}
 & & \multicolumn{2}{c}{(\mathsf{s\backslash np})/\mathsf{np}:\lambda x_3.\mathbf{defeat}(x_3,\mathbf{China})} \;> \\
\cline{2-4}
 & \multicolumn{3}{c}{\mathsf{s\backslash np}:\lambda x_1.\mathbf{should}(\mathbf{defeat}(x_1,\mathbf{Brazil}))} \;> \\
\hline
\multicolumn{4}{c}{\mathsf{s}:\mathbf{should}(\mathbf{defeat}(\mathbf{Brazil},\mathbf{China}))} \;<
\end{array}
$$

(38)
$$
\begin{array}{cccc}
\textit{Brazil} & \textit{should} & \textit{defeat} & \textit{China} \\
\hline
\mathsf{np}:\mathbf{Brazil} & \begin{array}{c}(\mathsf{s\backslash np})/(\mathsf{s\backslash np}) \\ :\lambda P.\lambda x_1.\mathbf{should}(Px_1)\end{array} & \begin{array}{c}(\mathsf{s\backslash np})/\mathsf{np} \\ :\lambda x_2.\lambda x_3.\mathbf{defeat}(x_3,x_2)\end{array} & \mathsf{np}:\mathbf{China} \\
\cline{2-3}
 & \multicolumn{2}{c}{(\mathsf{s\backslash np})/\mathsf{np}:\lambda x_4.\lambda x_1.\mathbf{should}(\mathbf{defeat}(x_1,x_4))} \;>\mathbf{B} & \\
\cline{2-4}
 & \multicolumn{3}{c}{\mathsf{s\backslash np}:\lambda x_1.\mathbf{should}(\mathbf{defeat}(x_1,\mathbf{Brazil}))} \;> \\
\hline
\multicolumn{4}{c}{\mathsf{s}:\mathbf{should}(\mathbf{defeat}(\mathbf{Brazil},\mathbf{China}))} \;<
\end{array}
$$

Note that the variable x_4 in (38) is introduced by the composition rule. β-reductions have been performed implicitly in these derivations.

The rule (34) must be generalized to allow modal verbs to compose with ditransitives:

(39)
$$
\begin{array}{ccccccc}
\textit{I} & \textit{offered,} & \textit{and} & \textit{may} & \textit{give,} & \textit{my friend} & \textit{a ticket} \\
\mathsf{np} & ((\mathsf{s\backslash np})/\mathsf{np})/\mathsf{np} & (\mathsf{X\backslash X})/\mathsf{X} & (\mathsf{s\backslash np})/(\mathsf{s\backslash np}) & ((\mathsf{s\backslash np})/\mathsf{np})/\mathsf{np} & \mathsf{np} & \mathsf{np} \\
\cline{4-5}
 & & & \multicolumn{2}{c}{((\mathsf{s\backslash np})/\mathsf{np})/\mathsf{np}} \;>\mathbf{B}^2 & &
\end{array}
$$

Rather than listing out all of the possible versions of forward composition, it can be stated in generalized format using Steedman's "\$ convention" to schematize over functions of varying numbers of arguments:

(40) *Generalized forward composition* ($>\mathbf{B}^n$)

$$\mathsf{X/Y} \quad (\mathsf{Y/Z})/\$_1 \quad \Rightarrow_{\mathbf{B}^n} \quad (\mathsf{X/Z})/\$_1$$

In essence, the \$ acts as a stack of arguments that allows the rule to eat into a category. Steedman (2000b) gives the following recursive definition of the convention:

Convention 2 (\$ schematization). *For a category* X*,* $\{\mathsf{X}\$\}$ *denotes the smallest set containing* X *and all functions into a category in* $\{\mathsf{X}\$\}$*.*

Placing a slash in front of \$ constrains the set to contain only leftward or rightward functions for $\{\mathsf{X}\backslash\$\}$ and $\{\mathsf{X}/\$\}$ respectively. Unbracketed $\mathsf{X}\$$, $\mathsf{X}\backslash\$$, and $\mathsf{X}/\$$ are used to schematize over the members of the respective sets $\{\mathsf{X}\$\}$, $\{\mathsf{X}\backslash\$\}$ and $\{\mathsf{X}/\$\}$. Furthermore, subscripts are employed to differentiate distinct schematizations. As an example, $\{\mathsf{s}/\$\}$ is the set $\{\mathsf{s},\ \mathsf{s/np},\ \mathsf{(s/np)/np},\ldots\}$, and $\mathsf{s}/\$$, $\mathsf{s}/\$_1$, etc. act as schemas over that set. The subscript used on the two appearances of \$ in the rule (40) ensures that the argument stack matched by the schema on the left hand side of the rule is transferred to the result on the right side.

The rule of forward composition has a directional converse which composes two leftward looking categories. It appears as follows in its generalized form:

(41) *Generalized backward composition* ($<\mathbf{B}^n$)

$$\mathsf{(Y\backslash Z)\backslash\$_1}\quad \mathsf{X\backslash Y}\quad \Rightarrow_{\mathbf{B}^n}\quad \mathsf{(X\backslash Z)\backslash\$_1}$$

Because the categories of English generally contain forward slashes, it is difficult to provide a simple example of the use of backward composition. However, the next section shows how this rule is nonetheless an important component in providing the CCG analysis of argument cluster coordination (also known as *non-constituent coordination*). Before considering the permutation-inducing rules based on $\mathbf{B}$, I thus turn next to type-raising rules, which conspire with the harmonic composition rules to generalize CCG's associativity even further and permit the flexible constituency for which CCG is perhaps best known while using exactly the same categories used thus far.

2.2.2 Type-raising

CCG employs a further class of rules that allows a syntactic argument to become a function that seeks a function that seeks the original argument type, mirroring the effect of the combinator $\mathbf{T}$ (30b). For example, the following rule permits a subject noun phrase in English to become a function seeking an intransitive verb phrase:

(42) *Forward type-raising* ($>\mathbf{T}$)

$$\mathsf{X}\quad \Rightarrow_{\mathbf{T}}\quad \mathsf{Y/(Y\backslash X)}$$

In order to reduce the size of derivations and improve their readability, I will use the following conventions for type-raised categories.

Convention 3 (Instantiation of variables in type-raised categories). *When presenting derivations, the relevant instantiation of the variable in the type-raised category necessary for the derivation to succeed will be used in place of the variable. For example, a type-raised subject in English will be written as* $\mathsf{s/(s\backslash np)}$ *instead of* $\mathsf{Y/(Y\backslash np)}$.

Convention 4 (Abbreviated type-raised categories). *When a type-raised category is not the focal aspect of a derivation, the shortened representation* $\mathsf{np}^{\uparrow}$ *will be used in place of the full category.*

An immediate implication for the inclusion of forward type-raising is that it works in concert with the rule of forward composition (34) to induce associativity in English transitive sentences. Thus, we have the following alternative to the derivation (8) provided by AB.

(43)

```
 Brazil     defeated     Germany
 ------   -----------   --------
   np     (s\np)/np        np
 ------>T
s/(s\np)
 ---------------------->B
        s/np
 ---------------------------------->
               s
```

Because of the Principle of Combinatory Type Transparency(31), the logical form produced by this derivation is the same as that given in derivation (19).

By including just the two additional rules of forward harmonic composition and forward type-raising, CCG gains associativity in an entirely type-driven fashion in which the same core linguistic signs give rise to associativity by interaction with the combinatory rules. AB could only do this by using categorial ambiguity.

The ability of CCG to combine subject noun phrases with a verb before the verb has consumed its object noun phrases allows it to straightforwardly account for phenomena such as right-node raising, object extraction, and topicalization in English. For example, object extraction now proceeds as follows, using the categories assumed for AB:

(44)

```
team        that         Brazil    defeated
 --    -------------    -------   ----------
 n      (n\n)/(s/np)       np     (s\np)/np
                        ------>T
                        s/(s\np)
                        ------------------->B
                                s/np
       --------------------------->
                  n\n
 ---------------------------------<
                 n
```

The fact that object extraction is unbounded is immediately captured because forward harmonic composition allows the extracted argument to be successively passed up until it is revealed to the relativizer:

(45)
$$\begin{array}{ccccccc}
\textit{team} & \textit{that} & \textit{I} & \textit{thought} & \textit{that} & \textit{Brazil} & \textit{defeated}\\
\overline{\mathsf{n}} & \overline{\mathsf{(n\backslash n)/(s/np)}} & \overline{\mathsf{np}} & \overline{\mathsf{(s\backslash np)/s}} & \overline{\mathsf{s/s}} & \overline{\mathsf{np}} & \overline{\mathsf{(s\backslash np)/np}}\\
 & & \overline{\mathsf{s/(s\backslash np)}}^{>\mathbf{T}} & & & \overline{\mathsf{s/(s\backslash np)}}^{>\mathbf{T}} & \\
 & & \multicolumn{2}{c}{\overline{\mathsf{s/s}}^{>\mathbf{B}}} & & \multicolumn{2}{c}{\overline{\mathsf{s/np}}^{>\mathbf{B}}}\\
 & & & & \multicolumn{3}{c}{\overline{\mathsf{s/np}}^{>\mathbf{B}}}\\
 & & \multicolumn{5}{c}{\overline{\mathsf{s/np}}^{>\mathbf{B}}}\\
 & \multicolumn{6}{c}{\overline{\mathsf{n\backslash n}}^{>}}\\
\multicolumn{7}{c}{\overline{\mathsf{n}}^{<}}
\end{array}$$

The composition of the subject and the verb is also implicated in right node raising:

(46)
$$\begin{array}{cccccc}
\textit{Marcos} & \textit{threw} & \textit{and} & \textit{Ronaldo} & \textit{kicked} & \textit{the ball}\\
\overline{\mathsf{np}} & \overline{\mathsf{(s\backslash np)/np}} & \overline{\mathsf{(X\backslash X)/X}} & \overline{\mathsf{np}} & \overline{\mathsf{(s\backslash np)/np}} & \overline{\mathsf{np}}\\
\overline{\mathsf{s/(s\backslash np)}}^{>\mathbf{T}} & & & \overline{\mathsf{s/(s\backslash np)}}^{>\mathbf{T}} & & \\
\multicolumn{2}{c}{\overline{\mathsf{s/np}}^{>\mathbf{B}}} & & \multicolumn{2}{c}{\overline{\mathsf{s/np}}^{>\mathbf{B}}} & \\
\multicolumn{5}{c}{\overline{\mathsf{s/np}}^{<\Phi>}} & \\
\multicolumn{6}{c}{\overline{\mathsf{s}}^{>}}
\end{array}$$

The rule of forward type-raising of course has a directional converse:

(47) *Backward type-raising* ($<$**T**)

$$\mathsf{X} \Rightarrow_{\mathbf{T}} \mathsf{Y\backslash(Y/X)}$$

In conjunction with the rule of backward composition, this rule plays a crucial role in the analyses of Steedman (1985) and Dowty (1988) for argument cluster coordinations such as the following:

(48) The referee gave Unsal a card and Rivaldo the ball.

This phenomenon has also been called non-constituent coordination, reflecting the difficulty in assigning a sensible phrase structure constituency that groups indirect objects with direct objects. In categorial grammar, on the other hand, such coordinations receive a straightforward analysis through the use of type-raising and composition. In order to derive a sentence such as (48), we first type-raise each of the two objects and then compose them together, resulting in a function which is looking for a function that is missing its indirect object and direct object arguments. This function can then be coordinated with other functions of the same type, as illustrated in derivation (49).

(49)

The referee gave	*Unsal*	*a card*	*and*	*Rivaldo*	*the ball*
(s/np)/np	np	np	(X\X)/X	np	np
	(s/np)\((s/np)/np) <T	s\(s/np) <T		(s/np)\((s/np)/np) <T	s\(s/np) <T
	s\((s/np)/np) <B			s\((s/np)/np) <B	
	s\((s/np)/np) <Φ>				
s <					

Due to the semantic consistency maintained by the combinatory rules, this derivation of course builds a logical form with the correct dependencies between the verbal predicate and its arguments.

2.2.3 Crossed Composition

The rules discussed so far are all *order-preserving*. Thus, they are unhelpful when we consider sentences in which certain elements have permuted with respect to their canonical positions. An example of this in English is heavy-NP shift, where an adverb comes between a verb and its direct object:

(50) Kahn blocked skillfully a powerful shot by Rivaldo.

To handle such a sentence without relying on categorial ambiguity, we need the permutation-inducing rule of backward crossed composition (51), which allows the adverb to combine with the verb before the verb has consumed its direct object.

(51) *Backward crossed composition* ($<\mathbf{B}_\times$)

$$\mathsf{Y/Z} \quad \mathsf{X\backslash Y} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X/Z}$$

With this rule, the derivation of (50) is straightforward.

(52)

Kahn	*blocked*	*skillfully*	*a powerful shot by Rivaldo*
np	(s\np)/np	(s\np)\(s\np)	np
	(s\np)/np $<\mathbf{B}_\times$		
	s\np >		
s <			

Of course, this apparatus also makes non-peripheral extraction possible without categorial ambiguity.

(53)
$$\begin{array}{lllll}
\textit{shot} & \textit{that} & \textit{Kahn} & \textit{blocked} & \textit{skillfully} \\
 & \overline{\mathsf{(n\backslash n)/(s/np)}} & \overline{\mathsf{np}} & \overline{\mathsf{(s\backslash np)/np}} & \overline{\mathsf{(s\backslash np)\backslash(s\backslash np)}} \\
 & & \xrightarrow{}{>}\mathbf{T} & \multicolumn{2}{c}{\xrightarrow{}{<}\mathbf{B}_{\times}} \\
 & & \mathsf{s/(s\backslash np)} & \multicolumn{2}{c}{\mathsf{(s\backslash np)/np}} \\
 & & \multicolumn{3}{c}{\xrightarrow{}{>}\mathbf{B}} \\
 & & \multicolumn{3}{c}{\mathsf{s/np}} \\
 & \multicolumn{3}{c}{\xrightarrow{}{>}} & \\
 & \multicolumn{3}{c}{\mathsf{n\backslash n}} &
\end{array}$$

This example succinctly shows type-raising, harmonic composition, and crossed composition rules all working in concert to induce the associativity and permutativity required of the grammar.

As might be expected from the pattern with previous rules, there is of course a forward version of crossed composition.

(54) *Forward crossed composition* ($>\mathbf{B}_{\times}$)

$$\mathsf{X/Y} \quad \mathsf{Y\backslash Z} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X\backslash Z}$$

Forward crossed composition is generally considered to be inactive in the grammar of English because it can induce some highly ungrammatical scrambled orders. A detailed discussion of forward crossed composition and its implications for English is provided in §4.1.1. Nonetheless, it must be part of the universal grammar because it is crucial for permitting scrambled orders in languages like Turkish which have greater word order freedom than English. Furthermore, the developments in this dissertation will ultimately allow us to pin down and exploit the power of permutative rules such as (51) and (54) even in languages like English without engendering undesirable collapses in word order.

CCG makes available only one further rule class, based on the substitution combinator **S**.

2.2.4 Substitution

The combinator **S** is different from **B** and **T** in that it allows a single resource to be utilized by two different functors. We find the need for such a combinator in parasitic gap constructions (Ross, 1967) such as the following:

(55) a. John watched without enjoying the game between Germany and Paraguay.

b. game that John watched without enjoying

In both of these strings, a single dependent acts as the argument of both *watched* and *without enjoying*. If we consider the categories of the constituents in (55b), it is clear that the rules defined thus far will not allow the derivation to proceed.

(56) game that John [watched]$_{(\mathsf{s}\backslash\mathsf{np})/\mathsf{np}}$ [without enjoying]$_{((\mathsf{s}\backslash\mathsf{np})\backslash(\mathsf{s}\backslash\mathsf{np}))/\mathsf{np}}$

As Steedman (2000b) notes, the fact that sequences like *watched without enjoying* can coordinate with transitive verbs as in *watched without enjoying and disliked* indicates that it must be possible to combine the categories $(\mathsf{s}\backslash\mathsf{np})/\mathsf{np}$ and $((\mathsf{s}\backslash\mathsf{np})\backslash(\mathsf{s}\backslash\mathsf{np}))/\mathsf{np}$ to form a transitive category. The following rule provides exactly this functionality:

(57) *Backward crossed substitution* ($<\mathbf{S}_\times$)

$$\mathsf{Y/Z} \quad \mathsf{(X\backslash Y)/Z} \;\Rightarrow_{\mathbf{S}}\; \mathsf{X/Z}$$

With this rule available in the system, the derivation of (55b) then proceeds as follows:

(58)

$$
\begin{array}{llccccl}
\textit{game} & \textit{that} & \textit{John} & \textit{watched} & \textit{without enjoying} & \\
\cline{2-5}
 & \mathsf{(n\backslash n)/(s/np)} & \mathsf{np} & \mathsf{(s\backslash np)/np} & \mathsf{((s\backslash np)\backslash(s\backslash np))/np} & \\
\cline{3-3}\cline{4-5}
 & & \mathsf{s/(s\backslash np)} & \multicolumn{2}{c}{\mathsf{(s\backslash np)/np}} & {>}\mathbf{T},\ {<}\mathbf{S}_\times \\
\cline{3-5}
 & & \multicolumn{3}{c}{\mathsf{s/np}} & {>}\mathbf{B} \\
\cline{2-4}
 & \multicolumn{3}{c}{\mathsf{n\backslash n}} & & {>}
\end{array}
$$

There are of course other directional rule instantiations of the substitution combinator. For example, the forward harmonic substitution rule appears as follows:

(59) *Forward substitution* ($>\mathbf{S}$)

$$\mathsf{(X/Y)/Z} \quad \mathsf{Y/Z} \;\Rightarrow_{\mathbf{S}}\; \mathsf{X/Z}$$

This rule is implicated in the following sorts of parasitic gaps in English (Steedman, 1996):

(60) team that I persuaded every detractor of to support

Leaving aside how we form the category $\mathsf{np/np}$ for the sequence *every detractor of* until later (see §6.3), the derivation using forward substitution works as follows:

(61)

$$
\begin{array}{lcccccl}
\textit{team} & \textit{that} & \textit{I} & \textit{persuaded} & \textit{every detractor of} & \textit{to support} & \\
\cline{2-6}
 & \mathsf{(n\backslash n)/(s/np)} & \mathsf{np} & \mathsf{((s\backslash np)/(s\backslash np))/np} & \mathsf{np/np} & \mathsf{(s\backslash np)/np} & \\
\cline{3-3}\cline{4-5}
 & & \mathsf{s/(s\backslash np)} & \multicolumn{2}{c}{\mathsf{((s\backslash np)/(s\backslash np))/np}} & & {>}\mathbf{T},\ {>}\mathbf{B} \\
\cline{4-6}
 & & & \multicolumn{3}{c}{\mathsf{(s\backslash np)/np}} & {>}\mathbf{S} \\
\cline{3-5}
 & & \multicolumn{3}{c}{\mathsf{s/np}} & & {>}\mathbf{B} \\
\cline{2-4}
 & \multicolumn{3}{c}{\mathsf{n\backslash n}} & & & {>}
\end{array}
$$

There are two more substitution rules, which are the directional opposites of the two already discussed:

(62) *Backward substitution* ($<\mathbf{S}$)

$\mathsf{Y\backslash Z} \quad \mathsf{(X\backslash Y)\backslash Z} \;\Rightarrow_{\mathbf{S}}\; \mathsf{X\backslash Z}$

(63) *Forward crossed substitution* ($>\mathbf{S}_{\times}$)

$\mathsf{(X/Y)\backslash Z} \quad \mathsf{Y\backslash Z} \;\Rightarrow_{\mathbf{S}}\; \mathsf{X\backslash Z}$

These rules are never utilized by the grammar of English since it lacks lexical categories that can serve as input to them, nor can it form such categories through the other combinatory rules. However, they are implicated in parasitic gap constructions in other languages, as discussed in §5.4. For more discussion on the substitution rules and a more detailed account of parasitic gaps in English, see Steedman (1996).

2.2.5 Limits on the Space of Possible Rules

As we have seen, CCG defines a number of syntactic rules that are directly related to the combinators $\mathbf{B}$, $\mathbf{T}$, and $\mathbf{S}$. The Principle of Combinatory Type Transparency(31) ensures that a wide range of hypothetically possible rules are excluded since they do not respect the workings of any of the combinators. For example, the following rule is excluded:

(64) $\mathsf{X\backslash Y} \quad \mathsf{Y/Z} \;\Rightarrow\; \mathsf{Y/X}$

However, the rules discussed thus far are not the *only* rules that could be created which faithfully obey the semantics of the combinators. For example, consider the following hypothetical instantiation of $\mathbf{B}$:

(65) $\mathsf{X\backslash Y} \quad \mathsf{Y/Z} \;\Rightarrow\; \mathsf{X\backslash Z}$

To exclude such rules, Steedman (2000b) employs the following principles:

(66) *The Principle of Consistency*

All syntactic combinatory rules must be consistent with the directionality of the principal function.

(67) *The Principle of Inheritance*

If the category that results from the application of a combinatory rule is a function category, then the slash defining directionality for a given argument in that category will be the same as the one(s) defining directionality for the corresponding argument(s) in the input function(s).

The Principle of Consistency excludes (65) because the principal function $\mathsf{X}\backslash\mathsf{Y}$ seeks its argument to the left whilst the rule states that its argument is located to its right. The Principle of Inheritance also rules out (65) because the leftward slash on the result category $\mathsf{X}\backslash\mathsf{Z}$ does not match that of the input category Y/Z.

2.2.6 An Aside on Generalized Composition

When generalizing the harmonic composition rules, the directionality of the arguments which are included in the schematization is important for ensuring that the generalized rules do not induce permutativity. For example, if the generalized rule for backward crossed composition is given as (68) instead of as (41), it can have an effect similar to a crossed composition rule, as shown in (69).

(68) $(\mathsf{Y}\backslash\mathsf{Z})\$_1 \quad \mathsf{X}\backslash\mathsf{Y} \quad \Rightarrow_{\mathbf{B}^n} \quad (\mathsf{X}\backslash\mathsf{Z})\$_1$ $\qquad (<\mathbf{B}^n)$

(69)
$$\begin{array}{cccc}
\textit{Brazil} & \textit{defeated} & \textit{yesterday} & \textit{the team that it beat previously}\\
\hline
\mathsf{np} & (\mathsf{s}\backslash\mathsf{np})/\mathsf{np} & \mathsf{s}\backslash\mathsf{s} & \mathsf{np}\\
\cline{2-3}
 & \multicolumn{2}{c}{(\mathsf{s}\backslash\mathsf{np})/\mathsf{np}\ {}^{<\mathbf{B}}} &
\end{array}$$

Though the sentence itself is grammatical, the derivation should be made possible by the permutative backward crossed composition rule and the adverb category $(\mathsf{s}\backslash\mathsf{np})\backslash(\mathsf{s}\backslash\mathsf{np})$ for *yesterday*, similar to (52). A harmonic composition rule generalized in the manner of (68) ceases to be order-preserving, leading to a loss of control over permutativity in the grammar.[2]

On the other hand, it seems that the crossed rules can be declared without specifying the directionality of the schematized arguments since they are already non-order preserving. For example, we can generalize forward crossed composition (54) as follows:

(70) $\mathsf{X}/\mathsf{Y} \quad (\mathsf{Y}\backslash\mathsf{Z})\$_1 \quad \Rightarrow_{\mathbf{B}^n} \quad \mathsf{X}\backslash\mathsf{Z}\$_1$ $\qquad (>\mathbf{B}^n)$

It should also be noted that although the form of the generalized rules (40) and (70) indicates that composition is unbounded, it should be viewed as a schema over several bounded instances of the rule (Steedman, 2000b). This is important since allowing unbounded composition increases the generative power of CCG beyond mild context-sensitivity (Weir, 1988).

[2] On pages 143, 169, and 193 of Steedman (2000b), Steedman gives the generalized backward harmonic composition as it is shown in (68), which is in error.

2.3 Multiset-CCG

CCG categories specify a rigid order of combination that limits the amount of variability in word order that a given CCG lexicon can express without using a great deal of categorial ambiguity. Motivated by the freedom in word order exhibited by Turkish, Hoffman (1995) proposes Multiset Combinatory Categorial Grammar (MULTISET-CCG), an extension of CCG which allows categories to express greater liberty in the word orders they will accept. MULTISET-CCG provides a very important backdrop for much of the work in this dissertation, and this section provides a very brief explanation of the MULTISET-CCG system and shows how its constructs can be utilized even in English.

Consider the two sentences of (71), in which the position of the particle *up* can change with respect to the object of the verb.

(71) a. Marcos picked up the ball.

b. Marcos picked the ball up.

To handle this variation in standard CCG, we could assume that the order in (71a) is basic and capture it with the category $((\mathsf{s}\backslash\mathsf{np})/\mathsf{np})/\mathsf{prt}$. The alternative order (71b) would then arise by type-raising the object *the ball* and then using backward crossed composition to combine it with the verb before the verb consumes the particle.

We could however consider an alternative in which the strict order specified by the verbal category is relaxed, permitting the two elements to permute more straightforwardly. This is the strategy adopted by Hoffman in creating MULTISET-CCG. To do so, she first redefines the manner in which categories are constructed so that arguments are placed into multisets which interact with the rules of MULTISET-CCG to allow permutation.

Definition 2 (Multiset-CCG categories). *Given $\mathcal{A}$, a finite set of atomic categories, the set of categories $\mathcal{C}$ is the smallest set such that:*

- $\mathcal{A} \subseteq \mathcal{C}$
- $\mathsf{X}_0\{|^1\mathsf{X}_1, \ldots, |^n\mathsf{X}_n\}, \in\ \mathcal{C}$, *if* $\mathsf{X}_0, \mathsf{X}_1, \ldots, \mathsf{X}_n \in \mathcal{C}$, *for* $n \geq 1$, $|^i \in \{\backslash, /, |\}$

The non-directional slash $|$ can unify with both $\backslash$ and $/$. Note that the notation I employ for multiset categories is slightly different from that which Hoffman uses so that it remains consistent with my own use of multisets in categories in later chapters.

For example, a category which will allow the particle and object to permute in (71) is the following:

(72) *picked* $\vdash$ $(\mathsf{s}\{\backslash\mathsf{np}\})\{/\mathsf{np}, /\mathsf{prt}\}$

In order to use such categories we need to redefine the rules of the system to operate on them. Hoffman gives the following rules for functional application:

(73) MULTISET-CCG *functional application*

a. $\mathsf{X}(\alpha \uplus \{/\mathsf{Y}\}) \quad \mathsf{Y} \quad \Rightarrow \quad \mathsf{X}\alpha$ (>)

b. $\mathsf{Y} \quad \mathsf{X}(\alpha \uplus \{\backslash\mathsf{Y}\}) \quad \Rightarrow \quad \mathsf{X}\alpha$ (<)

Greek letters such as α indicate multisets of arguments (which can be empty, in which case results such as $\mathsf{X}\alpha$ are trivially assumed to reduce to X). A difference between (73) and the way Hoffman defines the rule is that (73) uses *multiset* union $\uplus$ rather than normal set union $\cup$, which blocks the set α from containing multiple instances of Y. This is just a minor point which represents more accurately the intent of Hoffman's formulation of the rule.

The use of set-based categories means that the simple λ-calculus can no longer be used for representing meaning. Hoffman thus binds semantic arguments via standard unification, similar to Zeevat *et al.* (1987) and Zeevat (1988). This strategy works well with semantic representation frameworks such as Hybrid Logic Dependency Semantics (Kruijff, 2001) or Minimal Recursion Semantics (Copestake *et al.*, 1999, 2001), though it should also in principle be possible to modify the λ-calculus so that it can handle λ-abstractions over multi-sets.

The derivations for the alternations in (71) then proceed as follows:

(74)
$$\begin{array}{lllll}
\textit{Marcos} & \textit{picked} & \textit{up} & \textit{the} & \textit{ball} \\
\mathsf{np} & (\mathsf{s}\{\backslash\mathsf{np}\})\{/\mathsf{np}, /\mathsf{prt}\} & \mathsf{prt} & \mathsf{np}\{/\mathsf{n}\} & \mathsf{n} \\
 & \multicolumn{2}{c}{(\mathsf{s}\{\backslash\mathsf{np}\})\{/\mathsf{np}\} \;>} & \multicolumn{2}{c}{\mathsf{np} \;>} \\
 & \multicolumn{4}{c}{\mathsf{s}\{\backslash\mathsf{np}\} \;>} \\
\multicolumn{5}{c}{\mathsf{s} \;<}
\end{array}$$

(75)
$$\begin{array}{lllll}
\textit{Marcos} & \textit{picked} & \textit{the} & \textit{ball} & \textit{up} \\
\mathsf{np} & (\mathsf{s}\{\backslash\mathsf{np}\})\{/\mathsf{np}, /\mathsf{prt}\} & \mathsf{np}\{/\mathsf{n}\} & \mathsf{n} & \mathsf{prt} \\
 & & \multicolumn{2}{c}{\mathsf{np} \;>} & \\
 & \multicolumn{3}{c}{(\mathsf{s}\{\backslash\mathsf{np}\})\{/\mathsf{prt}\} \;>} & \\
 & \multicolumn{4}{c}{\mathsf{s}\{\backslash\mathsf{np}\} \;>} \\
\multicolumn{5}{c}{\mathsf{s} \;<}
\end{array}$$

Hoffman uses this strategy to provide an analysis of local scrambling in Turkish, as discussed in §4.2.1. To cover long distance scrambling and other constructions, she defines versions of the type-raising and composition rules that operate on multiset categories, and with these rules the generative power of MULTISET-CCG surpasses CCG's mildly context-sensitive capacity. However, I will wait to introduce these rules until the discussion on Turkish in Chapter 4, and turn now to the logical tradition of categorial grammar.

2.4 Categorial Type Logic

Of the categorial extensions of the AB calculus other than CCG, Categorial Type Logic (CTL) has received the most development and interest into the present day. Categorial Type Logic is a general term used by Moortgat (1997) to refer to the entire family of type logical proposals which have built on the logical approach to categorial grammar initiated by Lambek (1958). Some of the specific proposals include hybrid logical grammars (Hepple, 1994), type-logical grammar (Morrill, 1994), term-labeled categorial type systems (Oehrle, 1994), and many-dimensional or multi-modal categorial grammars (Oehrle, 1995; Moortgat, 1997). The development of CTL started with Van Benthem (1988) and Moortgat (1988), and, independently, work by Oehrle (1988) and Oehrle and Zhang (1989). Hepple's 1990 dissertation is another important formative work for CTL.

I provide here a brief introduction to CTL by presenting a product-free version using the natural deduction notation. For more thorough introductions, see Hepple (1994), Morrill (1994), Moortgat (1997), and Oehrle (to appear). Also, Bernardi (2002) provides a clear, concise, and up-to-date introduction. It should be noted that in this brief introduction I do not provide extensive motivation for the CTL perspective. Instead, I simply define the CTL apparatus and demonstrate how it works on a few examples. The presentation is intended to act as either (i) a brief reminder of how CTL works using CCG-friendly notation for those who are familiar but not intimate with CTL, or (ii) a supplement to some of the above introductions that hopefully clarifies some potential points of confusion for readers who are entirely new to CTL.

CTL extends categorial grammar with a resource-sensitive perspective based on Linear Logic (Girard, 1987) in order to introduce controlled use of operations like associativity and permutativity. The main idea behind CTL is to distinguish various modes of grammatical composition which can exhibit quite different logical properties. For AB, CCG, and MULTISET-CCG, there is only one binary mode with its forward and backward varieties (i.e. \ and /). In CTL, we can distinguish any number of different kinds of unary and binary modes of grammatical composition by using modalities, each of which may have its own unique behavior. Our first step, then, is to provide a definition of categories in CTL, in contrast to that given for AB and CCG in Definition 1, that adds logical modalities to the definition of categories.

Definition 3 (CTL Categories). *Given $\mathcal{A}$, a finite set of atomic categories and $\mathcal{M}$, a finite set of modalities, the set of categories $\mathcal{C}$ is the smallest set such that:*

- $\mathcal{A} \subseteq \mathcal{C}$
- $(\mathsf{X}\backslash_i \mathsf{Y}), (\mathsf{X}/_i \mathsf{Y}) \in \mathcal{C}$ *if* $\mathsf{X}, \mathsf{Y} \in \mathcal{C}$ *and* $i \in \mathcal{M}$
- $\Diamond_i \mathsf{X}, \Box_i^{\downarrow} \mathsf{X} \in \mathcal{C}$ *if* $\mathsf{X} \in \mathcal{C}$ *and* $i \in \mathcal{M}$

Note that I use Steedman's result-first notation for categories and that I am using the product-free version of CTL.

For example, if we let the set of modalities be $\{0,1\}$, we can create the following sorts of lexical entries:

(76) a. *defeated* $\vdash (\mathsf{s}\backslash_1 \mathsf{np})/_1 \mathsf{np}$

b. *today* $\vdash (\mathsf{s}\backslash_1 \mathsf{np})\backslash_0 (\mathsf{s}\backslash_1 \mathsf{np})$

c. *that* $\vdash (\mathsf{n}\backslash_1 \mathsf{n})/_1 (\mathsf{s}/_1 \mathsf{np})$

d. *that* $\vdash (\mathsf{n}\backslash_1 \mathsf{n})/_1 (\mathsf{s}/_1 \Diamond_0 \Box_0^{\downarrow} \mathsf{np})$

Actually, these categories are better characterized as logical formulas, as is perhaps more evident in other category notation styles such as that of Hepple (1997), where, for example, a transitive category like (76a) would be represented as $(\mathsf{np} \xrightarrow{1} \mathsf{s}) \xleftarrow{1} \mathsf{np}$

In rule-based categorial systems like CCG, parsing involves a process of derivation, whereas in CTL parsing takes the form of a deductive process of type inference over lexical assignments (Moortgat, 1999). To this end, CTL employs a resource-sensitive style of inference based on Linear Logic (Girard, 1987). The proof terms that the system operates on are *sequents* that pair a structure $\mathcal{S}$ with a category formula $\mathcal{F}$ as follows:

(77) $\mathcal{S} \vdash \mathcal{F}$

$\mathcal{S}$ is a *structured antecedent* which contains the premises which were used to prove the formula $\mathcal{F}$. Lexical entries such as those in (76) exemplify the most basic kind of sequent.[3] However, we can build complex configurations by using the rules of inference available in CTL, to which we now turn.

All CTL systems share a universal component, referred to as the *base logic*, which defines the basic behavior common to all of the connectives $\backslash_i$, $/_i$, $\Diamond_i$, and $\Box_i^{\downarrow}$. The most recognizable part of the base logic for those who are familiar only with AB or CCG are the slash elimination rules, which correspond to the application rules of AB. However, there are important differences. Most importantly, slash elimination is not necessarily tied to string adjacency as it is in functional application. Instead, these rules include a structure building component that organizes the antecedents of the premises into a new structured antecedent. As we will see later, these two substructures can potentially move away from each other due to the actions of restructuring rules.

(78) *Slash elimination schemas*:

a. $$\frac{\Gamma \vdash \mathsf{X}/_i\mathsf{Y} \qquad \Delta \vdash \mathsf{Y}}{(\Gamma \ \circ_i \ \Delta) \vdash \mathsf{X}}\ [/_iE]$$

b. $$\frac{\Delta \vdash \mathsf{Y} \qquad \Gamma \vdash \mathsf{X}\backslash_i\mathsf{Y}}{(\Delta \ \circ_i \ \Gamma) \vdash \mathsf{X}}\ [\backslash_iE]$$

where $i \in \mathcal{M}$

These rules range over all modalities i in the set of modalities. Though I have omitted semantics from these rules, CTL of course builds semantic terms in parallel with syntax in an entirely compositional manner.

With these rules, we can provide the following proof that the sentence *Brazil defeated Belgium today* is of type s:

(79) $$\dfrac{\mathit{Brazil} \vdash \mathsf{np} \qquad \dfrac{\dfrac{\mathit{defeated} \vdash (\mathsf{s}\backslash_1\mathsf{np})/_1\mathsf{np} \quad \mathit{Belgium} \vdash \mathsf{np}}{(\mathit{defeated}\ \circ_1\ \mathit{Belgium}) \vdash \mathsf{s}\backslash_1\mathsf{np}}\,[/_1E] \qquad \mathit{today} \vdash (\mathsf{s}\backslash_1\mathsf{np})\backslash_0(\mathsf{s}\backslash_1\mathsf{np})}{((\mathit{defeated}\ \circ_1\ \mathit{Belgium})\ \circ_0\ \mathit{today}) \vdash \mathsf{s}\backslash_1\mathsf{np}}\,[\backslash_0E]}{(\mathit{Brazil}\ \circ_1\ ((\mathit{defeated}\ \circ_1\ \mathit{Belgium})\ \circ_0\ \mathit{today})) \vdash \mathsf{s}}\,[\backslash_1E]$$

[3]Actually, such entries are abbreviations that obviate the need to use the identity axiom when providing proofs. That is, we are viewing only part of the multi-dimensional signs on either side of $\vdash$ in a sequent — the orthography in the antecedent and the syntactic category in the consequent. See Oehrle (to appear) for further explication of different notational conventions for displaying proofs.

It is important to note that the order of the elements which have taken part in the inference is reflected in the structured antecedents of each term and not in the proof itself. Thus, the following is a possible application of the leftward elimination rule:

(80) $$\frac{\textit{won} \vdash \mathsf{s}\backslash_1\mathsf{np} \quad \textit{Brazil} \vdash \mathsf{np}}{(\textit{Brazil} \circ_1 \textit{won}) \vdash \mathsf{s}}[\backslash_1 E]$$

Wherever possible, the order of the proof terms is of course conventionally lined up with the order induced in the structured antecedent, as I have done in (79); however, there are some structural operations which induce permutativity and a person unfamiliar with CTL should always keep in mind that it is not the proof itself, but instead the structured antecedent which reveals the word orders induced by CTL grammars.

In (79), we have only built structure, using the rules of inference available in the base logic. If that were all the system permitted, it would only be a partial system of implicational reasoning like AB and CCG. The CTL base logic thus also permits hypothetical reasoning. That is, we have not only lexical entries available as assumptions from the lexicon but also hypothesized elements. These elements can be consumed during the course of a proof and thus become part of the structured antecedent built during the process. Of course, in order for the proof to succeed, these hypotheses must eventually be discharged, and for this to happen, they must be located on the periphery of the structured antecedent in order for the slash introduction rules (81) to apply. A hypothesized element on the right periphery may be discharged with a rightward slash (81a), and one on the left periphery may be discharged with a leftward slash (81b).

(81) *Slash introduction schemas*:

a. $$\frac{\begin{array}{r}\cdots[x \vdash \mathsf{Y}]\\ \vdots\\ (\Gamma \circ_i x) \vdash \mathsf{X}\end{array}}{\Gamma \vdash \mathsf{X}/_i\mathsf{Y}}[/_i I]$$

b. $$\frac{\begin{array}{l}[x \vdash \mathsf{Y}]\cdots\\ \vdots\\ (x \circ_i \Gamma) \vdash \mathsf{X}\end{array}}{\Gamma \vdash \mathsf{X}\backslash_i\mathsf{Y}}[\backslash_i I]$$

Note that just as eliminating a slash with a modality i builds structure by connecting two antecedents with the mode $\circ_i$, an introduced slash inherits its modality from the structure which produced it.

Hypothetical reasoning combined with slash introduction already provides some interesting consequences. For example, the CCG rule of type-raising is a theorem of the system as it stands, as demonstrated by the following proof, which involves hypothesizing a function which consumes the argument and is then subsequently discharged:

(82)
$$\dfrac{\dfrac{\mathit{Brazil} \vdash \mathsf{np} \qquad [x_1 \vdash \mathsf{s}\backslash_1\mathsf{np}]^{\dagger}}{(\mathit{Brazil} \circ_1 x_1) \vdash \mathsf{s}}[\backslash_1 E]}{\mathit{Brazil} \vdash \mathsf{s}/_1(\mathsf{s}\backslash_1\mathsf{np})}[/_1 I]^{\dagger}$$

Note that the hypotheses are marked by identifying symbols such as † and ‡ when they are introduced and discharged to improve readability of proofs which have multiple hypothesized elements.

Despite these interesting qualities, the base logic still has a fairly hands-off approach to the structured antecedents of proof terms and as such it is not more flexible than AB. However, it is possible to augment the base logic by defining *structural rules* that reconfigure the antecedent set of premises and thus create systems with varying levels of flexibility. We have already seen how CCG is a highly associative system thanks to the interaction of the rules of composition and type-raising, and it is possible to achieve a similar effect in CTL with the use of structural rules of associativity. The following rules will permit structures built by the mode $\circ_1$ to be associatively restructured:

(83) *Right Association*:
$$\dfrac{(\Delta_a \circ_1 (\Delta_b \circ_1 \Delta_c)) \vdash \mathsf{X}}{((\Delta_a \circ_1 \Delta_b) \circ_1 \Delta_c) \vdash \mathsf{X}}[RA]$$

(84) *Left Association*:
$$\dfrac{((\Delta_a \circ_1 \Delta_b) \circ_1 \Delta_c) \vdash \mathsf{X}}{(\Delta_a \circ_1 (\Delta_b \circ_1 \Delta_c)) \vdash \mathsf{X}}[LA]$$

For readers unfamiliar with CTL, it may help to recall the operators of basic arithmetic in order to see what is going on with CTL's structural rules. Operators such as $+, -, \times$, and $\div$ each have their own interpretive properties as functions over their arguments, and they interact with the arithmetic rules to allow formulas to be manipulated into other equivalent formulas. At the most basic level, we have the following sorts of axioms:

(85) a. $(x + (y + z)) = ((x + y) + z)$ (addition is associative)

b. $(x + y) = (y + x)$ (addition is commutative)

c. $(x \times (y + z)) = ((x \times y) + (x \times z))$ (multiplication is distributive)

Equalities such as those in (85) allow complex expressions to be reformulated so that they can be simplified and interpreted. They are reflections of the fundamental structural and functional properties of the various operators. Syntactically, arithmetic formulas can be viewed as tree structures and the equalities in (85) can thus be viewed as tree rewriting operations. Consider the associativity of addition from this perspective:

(86)
$$\begin{array}{ccccc} & + & & & \\ / & & \backslash & & \\ x & & + & & \\ & / & & \backslash & \\ & y & & z & \end{array} \quad \Longleftrightarrow \quad \begin{array}{ccccc} & & & + & \\ & & / & & \backslash \\ & & + & & z \\ & / & & \backslash & \\ & x & & y & \end{array}$$

The use of reconfiguring arithmetic expressions in this manner is that it permits some reductions to be performed where they otherwise would not be possible. For example, to simplify the expression $(1 + (\sqrt{3} + 4))$ we can use the commutativity and associativity of addition to bring the 1 and the 4 together:

(87) $$(1 + (\sqrt{3} + 4)) = (1 + (4 + \sqrt{3})) = ((1 + 4) + \sqrt{3})$$

The resulting expression can then be reduced to $(5+\sqrt{3})$ by applying the $+$ operator to its two integer arguments. In a similar manner, CTL's structural rules allow the input resources to be reconfigured so that proofs may proceed where they otherwise would halt.

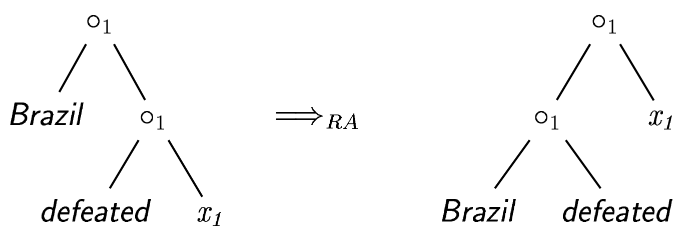
Just as CCG needs only the application rules to handle English object extraction, CTL can do the same with only the base logic. However, also as with CCG, something more is needed for a CTL-based analysis of object extraction, and structural rules provide the extra flexibility that is required. With CCG, we saw that the object extraction in (43) occurred by a process of type-raising the subject and composing it with the verb to produce the category s/np. In the example CTL system defined here, we obtain the same effect by use of the base logic in combination with the rule of Right Association (83).

As an example, consider the relativization *team that Brazil defeated*. We begin by introducing entries from the lexicon and hypothesizing the missing object of *defeated*. We then combine the premises using the slash elimination rules of the base logic, restructure the binary tree built up during the proof using the structural rule of Right Association, and then release the assumption using the rightward slash introduction rule.

(88)
$$\dfrac{\mathit{team} \vdash \mathsf{n} \qquad \dfrac{\mathit{that} \vdash (\mathsf{n}\backslash_1\mathsf{n})/_1(\mathsf{s}/_1\mathsf{np}) \qquad \dfrac{\dfrac{\dfrac{\mathit{Brazil} \vdash \mathsf{np} \qquad \dfrac{\mathit{defeated} \vdash (\mathsf{s}\backslash_1\mathsf{np})/_1\mathsf{np} \qquad [x_1 \vdash \mathsf{np}]^\dagger}{(\mathit{defeated} \circ_1 x_1) \vdash \mathsf{s}\backslash_1\mathsf{np}}[/_1E]}{(\mathit{Brazil} \circ_1 (\mathit{defeated} \circ_1 x_1)) \vdash \mathsf{s}}[\backslash_1E]}{((\mathit{Brazil} \circ_1 \mathit{defeated}) \circ_1 x_1) \vdash \mathsf{s}}[RA]}{(\mathit{Brazil} \circ_1 \mathit{defeated}) \vdash \mathsf{s}/_1\mathsf{np}}[/_1I]^\dagger}{(\mathit{that} \circ_1 (\mathit{Brazil} \circ_1 \mathit{defeated})) \vdash \mathsf{n}\backslash_1\mathsf{n}}[/_1E]}{(\mathit{team} \circ_1 (\mathit{that} \circ_1 (\mathit{Brazil} \circ_1 \mathit{defeated}))) \vdash \mathsf{n}}[\backslash_1E]$$

The crucial step for the extraction is where the structural rule RA applies and puts the assumption on the periphery so that it may be released by slash introduction — before that it is buried in an inaccessible position. This can be seen more clearly if we look at the structured antecedents as binary trees:

(89)



Here we see that the hypothesized element has moved to the periphery and is thus available for the rightward slash introduction rule. The parallel with traces and movement operations in GB/Minimalismshould be clear from this example, and in fact much recent research has explored the connections between Minimalist operations and the use of modalities and structural rules in CTL. See Cornell (1997), Vermaat (1999) and Lecomte (2001) and the collection of papers in Retoré and Stabler (to appear) devoted to the connections between categorial grammars and Minimalism.

In addition to the binary connectives with their residuals, CTL also has unary connectives that are typically used for features and fine-grained structural control. Their behavior is governed by the following rules:

(90) a. $$\dfrac{\Gamma \vdash \Box_i^{\downarrow}\mathsf{X}}{\langle\Gamma\rangle^i \vdash \mathsf{X}}[\Box_i^{\downarrow}E]$$

b. $$\dfrac{\langle\Gamma\rangle^i \vdash \mathsf{X}}{\Gamma \vdash \Box_i^{\downarrow}\mathsf{X}}[\Box_i^{\downarrow}I]$$

c. $$\frac{\Gamma \vdash \mathsf{X}}{\langle\Gamma\rangle^i \vdash \Diamond_i \mathsf{X}}\,[\Diamond_i I]$$

d. $$\frac{\Delta \vdash \Diamond_i \mathsf{X} \qquad \Gamma[\langle\mathsf{X}\rangle^i] \vdash \mathsf{Y}}{\Gamma[\Delta] \vdash \mathsf{Y}}\,[\Diamond_i E]$$

There are a few things to mention regarding notation in these rules. First, angled brackets in the structured antecedent, such as $\langle\Gamma\rangle^i$, is how a unary mode of composition is indicated in the structured antecedent, just as an indexed circle such as $(\Gamma \circ_i \Delta)$ indicates a binary mode of composition. Next, an antecedent formula such as $\Gamma[\Delta]$ indicates that there is some substructure Δ *within* the larger structure Γ. Third, the use of a category within an antecedent in (90d) may seem unexpected to someone who has seen only the proofs I have provided in this section, in which the antecedent appears to consist of lexical material whilst the consequent contains categories. So, it is worth mentioning again that we actually have full signs on each side of sequents, but in order to make proofs more readable, orthography is generally used to stand in for signs in the antecedent and categories to represent the sign of the consequent. However, for the definition of the $\Diamond$ elimination rule (90d), the category must be referenced in the antecedent.

Unary modalities can be used to implement fine-grained structural management regimes that introduce powerful operations such as permutativity in a controlled fashion (Moortgat, 1999). For example, consider the the following relative clause, in which the extraction site is not on the periphery:

(91) team$_i$ that Brazil defeated t$_i$ today

In addition to requiring associative restructuring such as that employed in (88), a proof for this extraction will require permutation of the hypothesized element with the adverb so that it can reach the periphery. To do this, we define the following structural rule of permutation keyed to the unary mode $\langle\cdot\rangle^0$:

(92) *Right Permutation*: $$\frac{((\Delta_a \circ_1 \langle\Delta_b\rangle^0) \circ_0 \Delta_c) \vdash \mathsf{X}}{((\Delta_a \circ_0 \Delta_c) \circ_1 \langle\Delta_b\rangle^0) \vdash \mathsf{X}}\,[\Diamond_0 RPerm]$$

With this rule, only resources which have undergone composition under the unary mode $\langle\cdot\rangle^0$ and lie in the correct configuration with respect to the binary modes $\circ_0$ and $\circ_1$ will be able to permute. As the proof in (93) and (94) shows, the form of this rule leads to the string *Brazil defeated today* having the category $\mathsf{s}/_1\Diamond_0\Box_0^{\downarrow}\mathsf{np}$, which can then be consumed by the category for the relativizer given in (76d). The proof is split into two parts due to space considerations.

$$(93)\quad \dfrac{[x_2 \vdash \Diamond_0\Box^{\downarrow}_0 np]^{\ddagger} \qquad \dfrac{\dfrac{\dfrac{\mathit{Brazil} \vdash np \qquad \dfrac{\dfrac{\mathit{defeated} \vdash (s\backslash_1 np)/_1 np \qquad \dfrac{[x_1 \vdash \Box^{\downarrow}_0 np]^{\dagger}}{\langle x_1\rangle^0 \vdash np}[\Box^{\downarrow}_0 E]}{(\mathit{defeated} \circ_1 \langle x_1\rangle^0) \vdash s\backslash_1 np}[/_1 E] \qquad \mathit{today} \vdash (s\backslash_1 np)\backslash_0(s\backslash_1 np)}{((\mathit{defeated} \circ_1 \langle x_1\rangle^0) \circ_0 \mathit{today}) \vdash s\backslash_1 np}[\backslash_0 E]}{(\mathit{Brazil} \circ_1 ((\mathit{defeated} \circ_1 \langle x_1\rangle^0) \circ_0 \mathit{today})) \vdash s}[\backslash_1 E]}{(\mathit{Brazil} \circ_1 ((\mathit{defeated} \circ_0 \mathit{today}) \circ_1 \langle x_1\rangle^0)) \vdash s}[\Diamond_0 RPerm]}{\dfrac{\dfrac{(\mathit{Brazil} \circ_1 ((\mathit{defeated} \circ_0 \mathit{today}) \circ_1 x_2)) \vdash s}{(((\mathit{Brazil} \circ_1 \mathit{defeated}) \circ_0 \mathit{today}) \circ_1 x_2) \vdash s}[RA]}{((\mathit{Brazil} \circ_1 \mathit{defeated}) \circ_0 \mathit{today}) \vdash s/_1\Diamond_0\Box^{\downarrow}_0 np}[/_1 I]^{\ddagger}}[\Diamond_0 E]^{\dagger}$$

$$(94)\quad \dfrac{\mathit{team} \vdash n \qquad \dfrac{\mathit{that} \vdash (n\backslash_1 n)/_1(s/_1\Diamond_0\Box^{\downarrow}_0 np) \qquad \begin{array}{c}(93)\\ \vdots \\ ((\mathit{Brazil} \circ_1 \mathit{defeated}) \circ_0 \mathit{today}) \vdash s/_1\Diamond_0\Box^{\downarrow}_0 np\end{array}}{(\mathit{that} \circ_1 ((\mathit{Brazil} \circ_1 \mathit{defeated}) \circ_0 \mathit{today})) \vdash n\backslash_1 n}[/_1 E]}{(\mathit{team} \circ_1 (\mathit{that} \circ_1 ((\mathit{Brazil} \circ_1 \mathit{defeated}) \circ_0 \mathit{today}))) \vdash n}[\backslash_1 E]$$

If no entries in the lexicon have the form $\Diamond_0\Box^{\downarrow}_0 np$ then the only way for a proof to carry on past the application of the permutation rule is for an extracting category such as that of the relativizer to consume the category eventually built as a result of the rule.

These themes of resource-sensitivity and fine-grained control will play a major role in the modifications to CCG that I propose in this dissertation.

2.5 Generative Power

One of the most salient differences between the various frameworks presented in this chapter is their respective generative strengths. This section briefly addresses the ideological conflict that has arisen between these frameworks on this issue.

The perceived position of natural language on the Chomsky hierarchy (for those who believe it has such a position) has fluctuated during the past five decades. During the the 1960's and 1970's, it was generally accepted by linguists that grammars for natural languages required greater than context-free power. Due to this perception, AB categorial grammar was somewhat sidelined after it was prove to be context-free by Bar-Hillel, Gaifman, and Shamir in 1964 – even Bar-Hillel himself gave up on categorial grammar because of this result. Bar-Hillel should not have despaired so easily – the purported proofs that natural language was greater than context-free were shown later to be flawed by Pullum and Gazdar (1982). Ultimately, the belief that context-free power is insufficient for natural language syntax was upheld by the proofs of Huybregts (1984) and Shieber (1985) that the crossing dependencies of Swiss German and Dutch are not context-free. Rambow (1994) and Hoffman (1995) argue that even greater power is needed to cover long distance scrambling in German and Turkish, respectively. However, the apparent supposition that the categorial approach could not somehow be generalized to create systems of greater power was itself incorrect, as evidenced by the flowering of extended categorial grammar formalisms in the 1980's. See Wood (1993) for elaboration on the various extensions.

The combinatory rules employed by CCG increase the context-free power of AB by a small but linguistically significant amount. With this extra power, CCG falls into the class of formalisms which have mildly context-sensitive generative power, along with Tree-Adjoining Grammar (Joshi, 1988), Linear Indexed Grammar (Gazdar, 1988), and Head Grammar (Pollard, 1984), which were all shown to be weakly equivalent by Vijay-Shanker and Weir (1994). Mild context-sensitivity is the weakest generative strength that is still powerful enough to handle the crossing dependencies of Dutch and Swiss German.

The categorial climb up the Chomsky hierarchy does not end with CCG. Motivated by scrambling phenomena, MULTISET-CCG modifies the manner in which categories are constructed and defines rules of combination for them which allow greater flexibility, resulting in a formalism with more than mildly context-sensitive power and less than fully context-sensitive power (Hoffman, 1995). CTL permits even more powerful operations, allowing it to attain Turing-complete power in its most general form (Carpenter, 1995). Moot (2002) provides a much finer characterization of the different generative strengths produced by CTL systems which conform to certain constraints, including the result that systems which use non-expanding structural rules are only context-sensitive. This result is particularly interesting since it appears that nearly all linguistic applications of CTL have obeyed this constraint (Moot, 2002).

Figure 2.1 presents a pictorial summary of where each of the formalisms discussed here sits on the Chomsky hierarchy.

The issue of generative capacity has proven itself to be a sticking point between the CCG and CTL traditions, which otherwise share many of the same intuitions and analyses. CTL's Turing-complete expressive power puts it at odds with CCG's

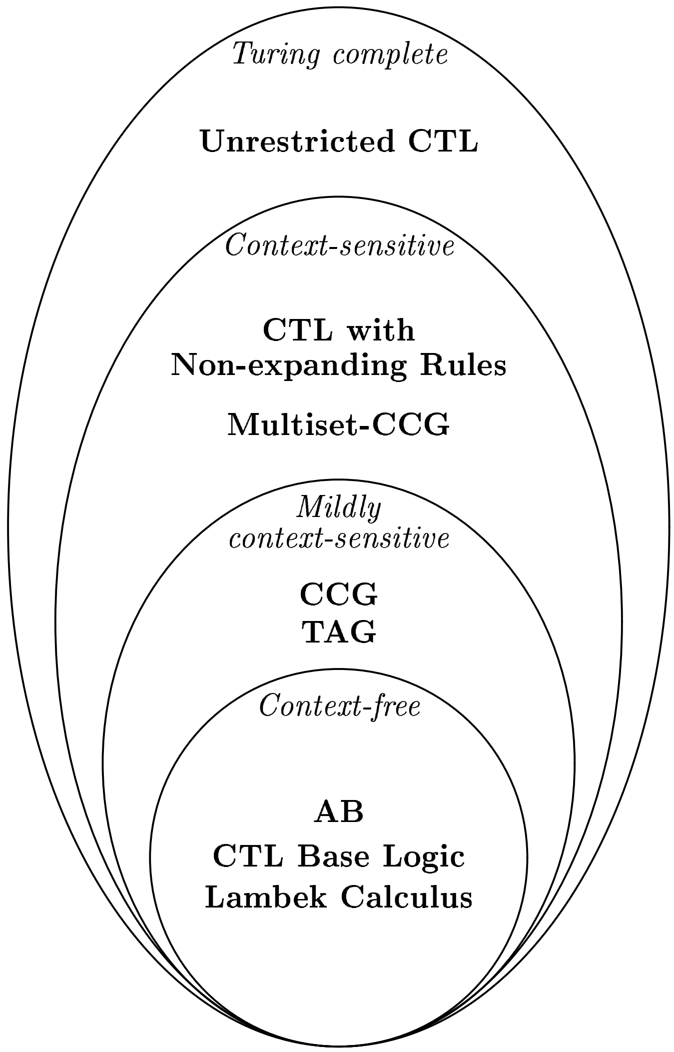


Figure 2.1: The positions of several formalisms on the Chomsky hierarchy.

conservatism, as is evident in the following comment by Steedman:

> Although the (weakly context-free) Lambek calculus provides an interesting starting point...., the addition of any of the non-order-preserving rules to its axioms immediately makes it collapse into permutation closure, as Moortgat (1988) has shown. (Steedman, 1996, p. 44)

Actually, the modern descendants of the Lambek calculus do not fall prey to such a collapse because permutation operations can be introduced in a controlled fashion and specified to induce permutation in only very particular configurations. Nonetheless, the fact that CTL is general enough to define rules which induce permutation closure is still problematic for Steedman, who argues that taking a commited stance on restricted generative capacity limits the range of analyses which may be put forth according to the theory and therefore boosts their predictive power:

> We should also try to minimize power in the modules of the theory, consistent with the primary goal of building interpretations that capture predicate-argument relations correctly. If we can do so with grammars that are provably low on the Chomsky hierarchy of formal languages and automata-theoretic power, then we are on stronger theoretical ground than if we adopted theories that achieve the same coverage at the expense of greater power, because greater automata-theoretic power increases the variety of alternative constructions and phenomena that we *could* capture beyond the point where there seems to be any motivation from empirical or imaginable semantic dependencies. (Steedman, 2000b, p. 23)

Thus, the basic linguistic claim for mildly context-sensitive formalisms in general is that their restricted formal power provides inherent limitations on theories which are built up around them and thereby enforce a wide-range of formal universals which do not even need to be explicitly stated. Such formalisms sit on the lower bound of natural language complexity without venturing any further – they are expressive enough, but cannot do everything. Theories that employ a low power formalism have fewer degrees of freedom than those which assume higher powered mechanisms as their basis, and they have thus far proven suitable for extensive coverage of empirical linguistic data.

For CTL researchers, issues of generative capacity are generally not considered to be of prime theoretical importance. For example, Morrill states that

> there is no absolute measure as to what constitutes a restricted class of languages, so that while observations of generative capacity are perhaps interesting, there seems no sense in which they are important. (Morrill, 1994, p. 257)

Morrill further notes that we might consider generative scales other than the Chomsky hierarchy, such as a categorial hierarchy stratified by structural rules. Though there certainly have been investigations into the abstract complexity of particular CTL systems, such as the important proof in Pentus (1997) of the context-freeness of the Lambek calculus (Lambek, 1958), as well as Carpenter (1995) and Moot (2002), the CTL tradition does not generally place much *linguistic* significance on generative capacity.

One of the contributions of this dissertation is cross-linguistic support for the position that restricted generative power *is* linguistically relevant, and furthermore that mild context-sensitivity is sufficient. One side of the argument involves demonstrating that mildly context-sensitive power is sufficient for handling a variety of constructions from a number of languages for which more powerful mechanisms have been proposed. The other side is that the analyses we propose have inherent limitations that lead to interesting linguistic predictions.

It should be noted that generative capacity does not provide an argument for the primacy of CCG *per se*. For example, a CTL grammar which uses only particular kinds of structural rules is every bit as capable of producing such linguistic predictions. In fact, this is part of the true elegance of the type-logical approach – it is able to mix substructural logics in a resource-sensitive manner that permits the creation of systems with power ranging from its context-free base logic all the way to Turing-completeness. Thus, we are able to assume a conservative base for a CTL grammar and introduce more powerful operations of grammatical composition in a *controlled* manner. Indeed, another main contribution of this dissertation is to show that even a formalism with limited power like CCG must incorporate fine-grained resource-sensitivity to gain similar control over its finite range of operators.

2.6 Dependency Relations

It is generally recognized that making distinctions between heads and dependendents is important not only for linguistic description, but also as an important functional component of grammatical systems. Formalisms in the Dependency Grammar family work with heads and dependents as core ingredients of syntax, and categorial grammar formalisms are well-suited to encoded dependency-based approaches (Wood, 1993; Moortgat and Morrill, 1991; Kruijff, 2001). Though I will not use heads and dependents explicitly in syntax, I will nonetheless make use of dependency relations of Functional Generative Description (FGD) (Sgall *et al.*, 1986) throughout this dissertation in order to distinguish different dependents for descriptive purposes.

FGD distinguishes nearly 40 different dependency relations which are partitioned into two classes: *inner participants* and *free modifiers*. The former are the kinds of dependents which are obligatorily selected by a head and which must be unique within the head's domain. Examples of relations denoting inner participants are **Actor**, **Patient**, and **Addressee**. An **Actor** is a participant which is doing or directly causing something, a **Patient** corresponds to the dependent that is affected by the action denoted by its verbal head, and the **Addressee** expresses the recipient of the action. For example, an English ditransitive sentence expresses all of these roles:

(95) [Roberto Carlos]$^{\mathbf{Actor}}$ passed [Ronaldo]$^{\mathbf{Addressee}}$ [the ball.]$^{\mathbf{Patient}}$

Other inner participant roles include **Effect**, **Origin**, **Partitive**, and **Identity**. The term **Addressee** may appear inappropriate to readers familiar with other theories of roles since it could be interpreted as only applying to a dependent that is, for example, being spoken to, rather than a recipient in general. The use of the term is, however, used in the more general sense in FGD, and I adhere to this usage.

Because I will be focusing primarily on syntactic concerns, I will not be displaying logical forms throughout most of this dissertation. However, since CCG assumes that every syntactic step involves a matching semantic effect, the syntactic analyses must of course be shown to capture the right dependencies. To reflect this without making logical forms explicit, I will at times use the convention of annotating syntactic categories with inner participant dependency roles. For example, the category for a verb such as *passed* appears as follows:

(96) $((\mathsf{s}\backslash\mathsf{np}^{\mathbf{Act}})/\mathsf{np}^{\mathbf{Pat}})/\mathsf{np}^{\mathbf{Addr}}$

Thus, the outermost argument of the category will act as the **Addressee** dependent of the verbal head, the next as the **Patient**, and the innermost argument as the **Actor**. It should be stressed that such annotations have *no syntactic effect whatsoever* in the analyses I propose – they are there purely for the reader to understand the categories and their combinations better. This is particularly useful in a language such as Dutch in which all arguments are found in the same direction and one can easily get lost in the categories if such relations are not made explicit.

The class of free modifiers includes roles such as **Purpose**, **Cause**, **Time**, and **Direction** (Sgall *et al.*, 1986). These primarily modify verbal heads and are optionally specified in determining the meaning of the head. For example, the **Time:For How Long** dependent[4] in (97a) can be omitted (97b), but the **Actor** cannot be (97c).

(97) a. [Rivaldo]$^{\textbf{Actor}}$ played [for 73 minutes.]$^{\textbf{Time:For How Long}}$

b. [Rivaldo]$^{\textbf{Actor}}$ played.

c. *Played [for 73 minutes.]$^{\textbf{Time:For How Long}}$

Kruijff (2001) models dependency relations in logical forms as terms of hybrid modal logic (Blackburn, 2000), and furthermore he shows how these relations can project entailments about temporal relations and aspectual classifications. Kruijff builds these logical forms by using a resource-sensitive categorial proof theory (i.e. a CTL based system), and Baldridge and Kruijff (2002) shows how the dependency based perspective on semantic interpretation can be coupled with CCG to produce logical forms with greater efficiency.

Using dependency roles for descriptive purposes is particular useful for languages such as Tagalog and Toba Batak, for which it has proven extremely difficult to determine which arguments correspond to notions such as subject (Schachter, 1976; Hoekstra, 1986; Guilfoyle *et al.*, 1992; Kroeger, 1993). This is apparent in recent papers such as Maclachlan and Nakamura (1997), in which theta roles such as *Agent* and *Theme* are used instead of subject and object. These correspond to what I will call the **Actor** and **Patient** dependents, respectively, in Tagalog sentences, though it should be noted that theta roles often do not correlate well with dependency relations.

2.7 Summary

This chapter has provided an introduction to categorial grammar and several of its specific traditions. As is clear from the discussion, different categorial perspectives can take on quite different characters, motivated as they are by different phenomena and interests. Such diversity is certainly healthy for the field, and a major part of the contribution of this dissertation is to take advantage of developments from different

[4] **Time:For How Long** is a subtype of **Time**.

perspectives to create a formulation of categorial grammar that enjoys attractive aspects from all of these systems.

Chapter 3

Substantive Universals

The explanatory adequacy of any theory of grammar will ultimately rest on its ability to provide linguistic universals that hold across the grammars of all languages and hence make strong predictions about the space of possible human languages. Chomsky (1965) classifies universals as being of two types: *formal* universals that arise from the abstract properties and conditions of a grammatical system and *substantive* universals that state that objects of a particular kind must be drawn from a fixed class of items. Chomsky gives the proposal that transformational rules must be included in the syntactic component as an example of a formal universal, and the proposal that certain fixed syntactic categories such as noun and verb are utilized in the grammar of any language as an example of a substantive universal.

It is not always possible to definitively declare a given universal as formal or substantive; in fact, Chomsky's dual classification may be better characterized as a continuum from more to less abstract universals. The more abstract formal universals are generally stronger in their predictive force. They are more likely to dictate the nature of entire classes whereas more substantive universals typically only carve out a select group of objects from a potentially infinite set in a class. That is to say that formal universals set the stage upon which substantive ones are formulated. Nonetheless, determining substantive universals is also extremely important for restricting an abstract formal system to account for only natural language grammars — without a theory of substantive constraints, we could in principle write grammars for programming languages such as C++ and Java using CCG lexicons.

In this chapter, I discuss a number of proposals for substantive universals in CCG that play an important role in the linguistic analyses provided in this dissertation. CCG provides an abstract formalization of categories and rules, and it is conceivable to utilize the CCG formalization in conjunction with different substantive theories. However, since a CCG-based grammar for any given language is almost entirely determined by its lexicon, any effective proposal must limit the unbounded range of possible categories to a finite set of linguistically relevant categories. It is also crucial to structure the lexicon in a manner that reduces the redundancy between different entries that share information. Proposals for both of these areas are provided in this chapter, based largely on work in other frameworks. Finally, the combinatory rules themselves are also potential targets for substantive constraints. In §3.3, I place such a constraint on type-raising and discuss the rule restrictions permitted by CCG in the context of substantive universals.

3.1 Substantive Lexical Categories

Most work in categorial grammar has focused on formal aspects and has sought to define the form which categories and rules may take and demonstrate the linguistic significance of these choices. For example, Steedman (2000b) shows that taking the combinatory rules of CCG as formal universals leads to a strong prediction regarding the connection between the residues of extraction, coordination, parentheticals, and intonational constituents — namely that they are all mediated by the ability of the system to form the same incomplete constituents in their respective contexts. While this line of inquiry has of course required a robust notion of the content of lexicons for different languages, there has yet to be a truly detailed cross-linguistic study into the categorial lexicon. The GB framework, on the other hand, has benefitted from extensive cross-linguistic research, so it may thus be advantageous to consider utilizing some of that work to help provide a theory of the lexicon. Steedman points to such a possibility:

> In categorial terms, such theories can be seen as predominantly concerned with predicate-argument structure and hence with elements of semantic interpretation or Logical Form, rather than syntax proper. To the extent that such theories provide a systematic account of the relation between interpretations in this sense and syntactic categories, they provide what amounts to a theory of the categorial lexicon — a component of the present theory that continues to be lacking in this and preceding discussions of CCG. (Steedman, 2000b, p. 257)

Properly speaking, the above statement pertains to a theory of lexical categories, not to a full theory of the lexicon. This section provides some concrete suggestions for how a theory such as GB could be used to create lexical categories, as well as how an inheritance-based approach using typed feature structures can provide a rich and flexible definition of atomic categories.

Steedman (2000b) proposes several other substantive constraints that limit the kinds of lexicons that are permitted according to the CCG theory. One of the most basic assumptions about the possible categories that can be used in natural language grammars is that they are paired with underlying semantic types and that the mapping between syntactic and semantic arguments is tight. This is encapsulated in the following principle:

(98) *The Principle of Categorial Type Transparency*
For a given language, the semantic type of the interpretation with a number of language-specific directional parameter settings uniquely determines the syntactic category of a category.

An immediate implication of this assumption is that the lexical function categories of natural language have finite arity — for example, the maximum number of arguments for English verbs appears to be four, from verbs like *bet*, as in *John bet Bill five dollars that Brazil would defeat China* (Steedman, 2000b). Thus, even though the definition of categories permits them to have unbounded arity, they are substantively limited by the finite nature of the compositional semantic interpretations that they construct.

Steedman (2000b) also assumes the Principle of Lexical Head Government, which is related to the Projection Principle of GB and the Condition on Elementary Tree Minimality of TAG (Frank, 1992, 2002), and amounts to a statement that CCG is lexicalized.

(99) *The Principle of Lexical Head Government*
Both bounded and unbounded syntactic dependencies are specified by the lexical syntactic type of their head.

The term *head* in this principle refers to the words which correspond to functors in logical form, such as verbs.

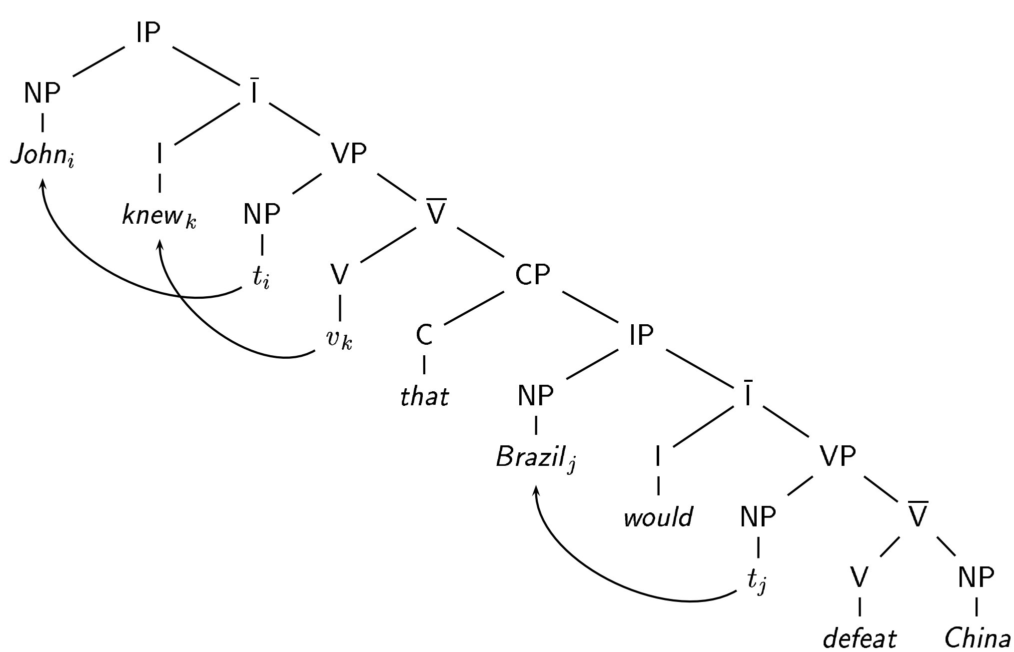
Even with these principles, the number of possible categories is vast, and we require further substantive universals to constrain them to just those needed for natural language. Frank (1992, 2002) provides such a theory for TAG by incorporating GB-theoretic notions and analyses with the TAG machinery. Even though CCG's categories may not at first glance appear to be amenable to using GB-based notions, it is actually fairly straightforward to capture a number of distinctions. Though I will not attempt a proper treatment of GB within CCG, even very basic usage of certain core ideas in GB do prove useful in providing more constrained analyses of cross-linguistic data. It also has the advantage of hopefully making categorial analyses more accessible to readers with a GB background. In particular, I will utilize the standard GB clausal spine to encode clausal levels within CCG categories.

I will not provide any sort of introduction to GB here. My utilization of ideas from GB in this dissertation can be understood without familiarity with the specifics of GB, and the reader is directed to (Chomsky, 1981), Chomsky and Lasnik (1993) and Chomsky (1995) for introductions to the framework. I will adopt two essential notions from GB:

1. the distinction between heads whose specifiers are empty ($\overline{\mathsf{X}}$) or filled (XP)
2. the distinction between complementizer phrases (CP), inflectional phrases (IP), and non-inflected verb phrases (VP).

For example, consider the following GB-theoretic structure for the sentence *John knew that Brazil would defeat China*:

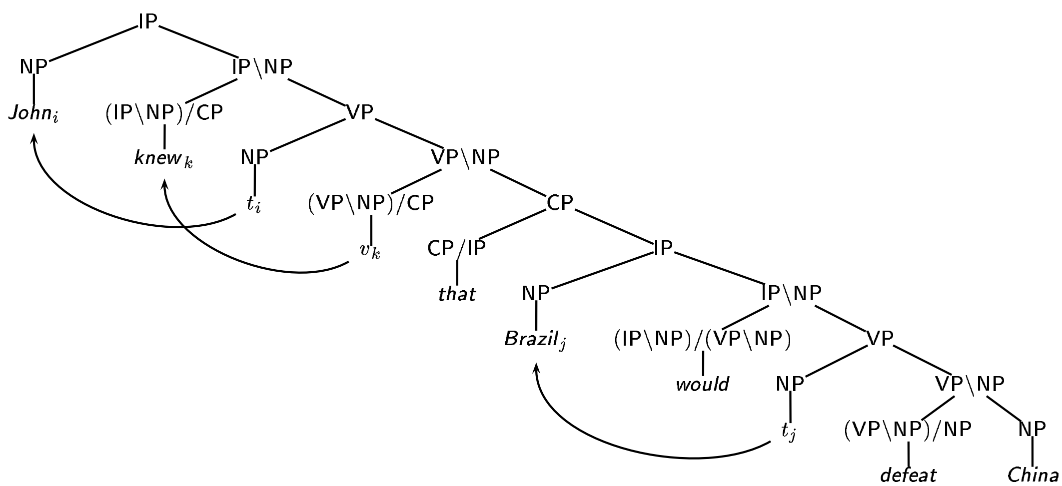
(100)



Under standard assumptions, *knew* would actually remain in V and the inflection would lower to attach to the verb (Pollock, 1989). However, my (quite limited) use of the GB clausal spine will assume that inflected elements do indeed "move" to I in the categorial encoding of the spine, thus reflecting the analysis given in (100). In doing so, I ignore the standard GB analysis regarding the distribution of adverbs in French and English which assumes that adverbs always attach to VP. While it would certainly be possible to remain more faithful, I assume that adverbs can potentially attach to any verbally-headed phrase.

Such an analysis can actually be used to generate a set of lexical items and CCG categories by adapting the algorithm defined by Hockenmaier *et al.* (2001) and Hockenmaier and Steedman (2002a) for creating categories from tree structures. The algorithm proceeds top-down and uses information about headedness to replace atomic phrase structure labels with categories that encode the syntactic arguments of the nodes in the tree. It is sensitive to noun phrase traces, but does not incorporate any notion of verb movement. However, it should be possible to modify the algorithm to deal with verb movement and relabel the nodes in (100) with categories as follows:

(101)



We can then read the lexical items off this new structure, thereby producing a CCG lexicon that encodes the GB analysis.

(102) a. *John* ⊢ NP

b. *knew* ⊢ (IP\NP)/CP

c. *that* ⊢ CP/IP

d. *Brazil* ⊢ NP

e. *would* ⊢ (IP\NP)/(VP\NP)

f. *defeat* $\vdash$ (VP\NP)/NP

g. *China* $\vdash$ NP

With this lexicon, we can clearly provide a CCG derivation using forward and backward application that is essentially (100) turned upside-down. The advantage for CCG is that it paves the way for a methodology for creating CCG categories in a cross-linguistic manner. The advantage for GB is that comparative language studies for bounded phenomena can be carried out in a familiar methodology that maintains the same sort of vocabulary (movement, traces, specifiers, etc.) whilst gaining CCG's theory of unbounded movement and flexible constituency (as well as compositional semantics and computational efficiency). For example, complex right-node raising examples come for free given the miniature lexicon (102) and the rules of CCG. The following derivation shows how the system provides not only a derivation for *John knew that Brazil would defeat* required as input for the coordination seen in (103), but also that it can do so completely incrementally:

(103) (*John knew that Brazil would defeat*), and (*Bill predicted that Turkey would tie with*), China.

(104)
```
 John        knew        that     Brazil       would            defeat      Turkey
------    -----------    -----   --------  ---------------   ------------   ------
  NP       (IP\NP)/CP    CP/IP      NP     (IP\NP)/(VP\NP)    (VP\NP)/NP      NP
------->T                        ------->T
IP/(IP\NP)                       IP/(IP\NP)
------------------------>B
        IP/CP
---------------------------------->B
            IP/IP
           -------------------------------->B
                      IP/(IP\NP)
                    ------------------------------------------>B
                                  IP/(VP\NP)
                                 ------------------------------------------------->B
                                                 IP/NP
                                                ------------------------------------->
                                                                  IP
```

We thus see the categorial calculus driving a GB clausal analysis and providing the grammatical compositions that are not as straightforwardly obtained with the rigid assumptions about phrase structure and composition structure correspondences inherent in the standard GB approach.

This fairly direct encoding of GB node labels into distinct atomic categories introduces some inconveniences for the working categorial grammarian. For one thing, adverbs would need to have several categories in order to reflect when they attach to CP, IP, and VP. Another problem is that we need the further category (IP\NP)/IP for sentences in which *knew* takes a clause that is not introduced by

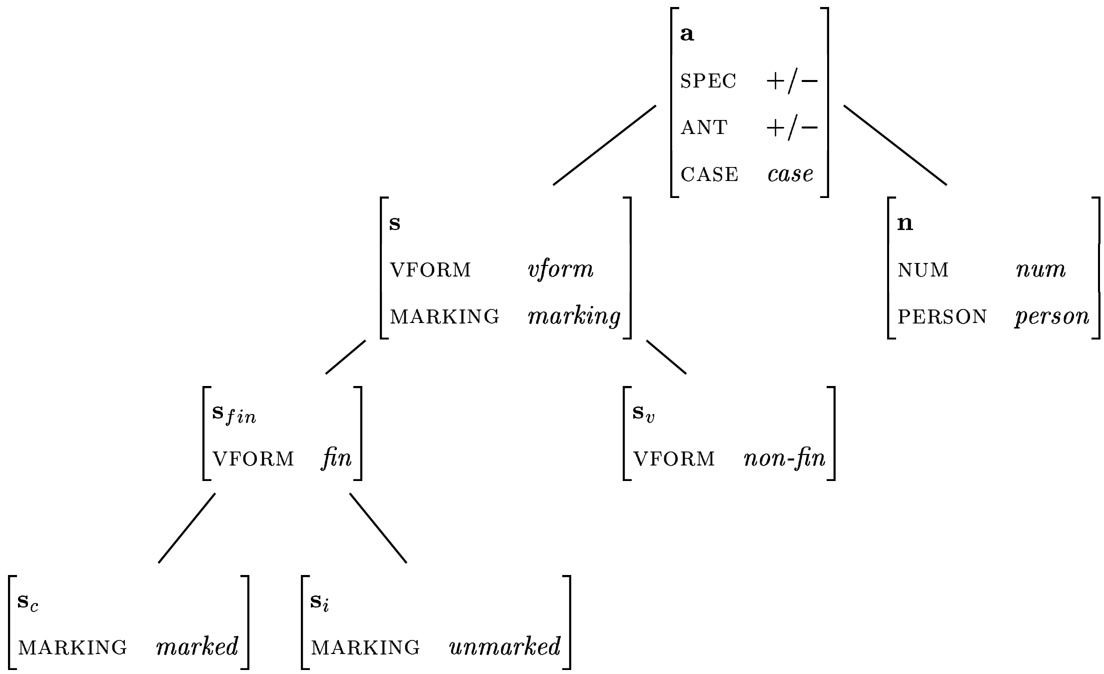


Figure 3.1: Inheritance hierarchy for atomic categories.

that. It thus appears that greater flexibility with respect to atomic category types is desirable.

Steedman (2000b) utilizes the category s with the features $\pm$CP and $\pm$IP to encode some of the intuitions familiar from GB, but he does not use these features to much effect and we can instead bring the flavor of GB analyses into CCG categories in a more perspicous way which bestows further advantages on our conception of the lexicon: a hierarchy of typed feature structures. Typed feature structures encoded in inheritance hierarchies are the foundation upon which the data structures of Head-driven Phrase Structure Grammar (HPSG) (Pollard and Sag, 1987, 1994) are built, and Villavicencio (2002) has already made extensive use of them in the context of Unification-Based Generalized Categorial Grammar. I will not provide an introduction to these concepts here, and instead point the reader to Shieber (1986), Pollard and Sag (1994), Ginzburg and Sag (2000) and Copestake (2002).

To bring these threads of GB, HPSG, and CCG together, the strategy is to use non-recursive typed feature structures familiar from HPSG as the data structures for atomic categories. The types of these feature structures are named using the GB labels corresponding to the featural configurations found within, and the CCG categories will be written using these type names. The inheritance hierarchy shown in Figure 3.1 outlines an initial, though incomplete, suggestion for atomic categories.

Throughout the remainder of this dissertation, I will use the type names directly in categories to indicate the use of the feature structure associated with them, including the use of a to refer to any atomic category. The attribute SPEC indicates whether the category has a filled specifier, i.e. bar level. A noun category with a filled specifier corresponds to the traditional category np. ANT signifies *antecedent-government*, which marks a category as being lexical or not, as discussed in §5.2.6. CASE appears on the type a rather than just n to accomodate case-marked sentences in languages like Turkish and Tagalog.

The sub-hierarchy rooted in the type s declares features related to verbal categories. *vform* is the traditional HPSG attribute for *verb form* and takes values such as *fin(ite)*, *non-fin(ite)*. Though not shown in this hierarchy, subtypes of the *non-fin* value include past-participles, infinitives, and gerunds, and I will make use of atomic categories such as s_{ppt}, s_{inf}, and s_{ger} to represent such categories. MARKING encodes the distinction between different clause types, such as complementizer phrases (*marked*) and and inflectional phrases (*unmarked*).

As examples, the fully-specified feature structure corresponding to an accusatively-marked, inflected sentence, or IP, is shown in (105a), and that for the English pronoun *I* is given in (105b).

(105) a. $\begin{bmatrix} \mathbf{s}_i & \\ \text{SPEC} & + \\ \text{ANT} & - \\ \text{CASE} & \textit{acc} \\ \text{VFORM} & \textit{fin} \\ \text{MARKING} & \textit{unmarked} \end{bmatrix}$ b. $\begin{bmatrix} \mathbf{n} & \\ \text{SPEC} & + \\ \text{ANT} & - \\ \text{CASE} & \textit{nom} \\ \text{NUM} & \textit{sing} \\ \text{PERSON} & \textit{first} \end{bmatrix}$

Work in CCG has always assumed that categories are bundles of features of this sort, but this has generally not been made explicit. To achieve the level of granularity that is characteristic of the detail found in most work in HPSG, CCG grammars can thus import both formal and linguistic developments for encoding features and regularities in the lexicon. The Unification Categorial Grammar framework of Zeevat *et al.* (1987); Zeevat (1988); Moens *et al.* (1989) started this, and CCG can clearly benefit from making this aspect of categories more explicit.

With this basic hierarchy, I am actually *not* proposing to faithfully encode GB analyses with CCG categories — instead I am capturing some of the overall flavor of the GB approach, but with much greater flexibility and generality. In effect, I am using the GB notions as labels for fairly standard HPSG data structures. Nonetheless, if one did wish to be more faithful to GB, the lexicon of (102) demonstrates how this could be accomplished.

For the remainder of this dissertation, I assume several conventions for writing categories to reduce clutter in derivations.

Convention 5 (Filled and unfilled). *The following abbreviations are used to indicate the value of the* SPEC *feature:*

- $\mathsf{s}_{spec=+}$ *is written as* $\ddot{\mathsf{s}}$.
- $\mathsf{s}_{spec=-}$ *is written as* $\dot{\mathsf{s}}$.
- $\mathsf{n}_{spec=+}$ *is written as* np.
- $\mathsf{n}_{spec=-}$ *is written as* n.

The single and double dots above the s *categories are intended to be reminiscent of single and double bar notation from* $\overline{X}$*-theory. I do not make use of underspecified* n *categories in this dissertation, and thus use the standard categories to represent the* SPEC *distinction.*

Convention 6 (Obvious attributes). *When a particular feature needs to be highlighted, it is written as a subscript on the category, such as* $\mathsf{s}_{VFORM=fin}$. *When the relevant attribute is clear from context, it is omitted as in* s_{fin} *for* $\mathsf{s}_{VFORM=fin}$ *and* n_{nom} *for* $\mathsf{n}_{CASE=nom}$.

Convention 7 (Abbreviated case values). *Case values are often abbreviated to just the first letter of the case name, e.g.* n_n *for* n_{nom} *and* n_a *for* n_{acc}.

Convention 8 (Positive-negative values).

- *Features which take on the values + and − are normally written with the value in front of the attribute, e.g.* n_{+ANT} *for* $\mathsf{n}_{ANT=+}$.
- *The underspecified value for + and − valued attributes is represented as* $\pm$.

With the atomic category hierarchy and the above conventions, the lexicon given in (102) then appears as in (106).

(106) a. *John* $\vdash$ np

b. *knew* $\vdash (s_i\backslash np)/s_{fin}$

c. *that* $\vdash \ddot{s}_c/s_i$

d. *Brazil* $\vdash$ np

e. *would* $\vdash (s_i\backslash np)/(s_v\backslash np)$

f. *defeat* $\vdash (s_v\backslash np)/np$

g. *China* $\vdash$ np

Note that these categories are not a significant departure from those generally assumed in CCG. They are essentially shorthand for more traditional-looking categories that would require more space to write out. For example, the category for *would* might instead appear as $(s_{VFORM=fin}\backslash np)/(s_{VFORM=non-fin}\backslash np)$.

Because of the properties of unification, a given type with a value specified, such as $\ddot{s}_i$ ($\equiv s_{i+SPEC}$) can unify with s_i ($\equiv s_{i\pm SPEC}$), the same type with the value underspecified, to produce a result with the value specified. Furthermore, a more general type such as s_{fin} can unify with both of its subtypes, s_c and s_i. This allows the complementizer to be optional for the complements of verbs such as *knew*:

(107)
$$
\begin{array}{cccc}
\textit{John} & \textit{knew} & \textit{that} & \textit{Brazil would defeat China} \\
\hline
np & (s_i\backslash np)/s_{fin} & \ddot{s}_c/s_i & s_i \\
 & & \multicolumn{2}{c}{\underline{\qquad\qquad\qquad\qquad}>} \\
 & & \multicolumn{2}{c}{\ddot{s}_c} \\
 & \multicolumn{3}{c}{\underline{\qquad\qquad\qquad\qquad\qquad\qquad}>} \\
 & \multicolumn{3}{c}{s_i\backslash np} \\
\multicolumn{4}{c}{\underline{\qquad\qquad\qquad\qquad\qquad\qquad\qquad}<} \\
\multicolumn{4}{c}{s_i}
\end{array}
$$

(108)
$$
\begin{array}{ccc}
\textit{John} & \textit{knew} & \textit{Brazil would defeat China} \\
\hline
np & (s_i\backslash np)/s_{fin} & s_i \\
 & \multicolumn{2}{c}{\underline{\qquad\qquad\qquad\qquad\qquad}>} \\
 & \multicolumn{2}{c}{s_i\backslash np} \\
\multicolumn{3}{c}{\underline{\qquad\qquad\qquad\qquad\qquad\qquad}<} \\
\multicolumn{3}{c}{s_i}
\end{array}
$$

In addition to many other things, the distinction between finite and non-finite sentential categories is important for providing even a basic account of *do*-support in English:

(109) *China not won.

(110) China did not win.

If we assume that *not* has the category $(s_v\backslash np)/(s_v\backslash np)$, then it cannot take *won* as its argument since *won* is a tensed verb and therefore its value for VFORM is *fin*, as shown in (111). Thus, negation applies to the non-finite base form of the verb, and *did* provides the inflection, shown in (112).

(111) *China $[\text{not}]_{(\mathsf{s}_v\backslash\mathsf{np})/(\mathsf{s}_v\backslash\mathsf{np})}$ $[\text{won}]_{\mathsf{s}_i\backslash\mathsf{np}}$.

(112) China $\text{did}_{(\mathsf{s}_i\backslash\mathsf{np})/(\mathsf{s}_v\backslash\mathsf{np})}$ $[\text{not}]_{(\mathsf{s}_v\backslash\mathsf{np})/(\mathsf{s}_v\backslash\mathsf{np})}$ $[\text{win}]_{\mathsf{s}_v\backslash\mathsf{np}}$.

This is in fact a fairly standard unification-based analysis, and indeed it should be quite feasible to translate the fine-grained analyses of verbal types available in the HPSG literature into categories that employ similar features. However, the rough distinctions I have made here are sufficient for the purposes of this dissertation, and I leave the wider task of providing a more detailed account for future work.

To provide an analysis of English topicalization, we can utilize the distinction between s_c and s_i to create a category for the fronted argument that appears very much like a type-raised noun phrase: $\ddot{\mathsf{s}}_c/(\ddot{\mathsf{s}}_i/\mathsf{np})$. This category permits the following derivation for *China, Brazil defeated*:

(113)
$$
\begin{array}{ccc}
\textit{China} & \textit{Brazil} & \textit{defeated} \\
\hline
\ddot{\mathsf{s}}_c/(\ddot{\mathsf{s}}_i/\mathsf{np}) & \mathsf{np} & (\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np} \\
 & \overline{\mathsf{s}_i/(\mathsf{s}_i\backslash\mathsf{np})}^{\,>\mathbf{T}} & \\
 & \multicolumn{2}{c}{\overline{\mathsf{s}_i/\mathsf{np}}^{\,>\mathbf{B}}} \\
\multicolumn{2}{c}{\overline{\ddot{\mathsf{s}}_c}^{\,>}} &
\end{array}
$$

Fronting categories play a major role in the analysis of Dutch given in Chapter 5, based on the analysis of Steedman (2000b).

The categories given in (21) for relativization are refined to apply only to finite, non-topicalized phrases which are missing an argument:

(114) a. $(\mathsf{n}\backslash\mathsf{n})/(\mathsf{s}_i\backslash\mathsf{np})$

b. $(\mathsf{n}\backslash\mathsf{n})/(\mathsf{s}_i/\mathsf{np})$

With these categories, both of the relativized clause types exemplified in (20) can be derived while blocking the following ungrammatical relativizations due to the fact that the category s_i cannot unify with s_{inf}:

(115) *team $[\text{that}]_{(\mathsf{n}\backslash\mathsf{n})/(\mathsf{s}_i/\mathsf{np})}$ $[\text{Brazil to defeat}]_{\mathsf{s}_{inf}/\mathsf{np}}$

(116) *team $[\text{that}]_{(\mathsf{n}\backslash\mathsf{n})/(\mathsf{s}_i\backslash\mathsf{np})}$ $[\text{to defeat}]_{\mathsf{s}_{inf}\backslash\mathsf{np}}$ China

I do not wish to argue that the use of GB-theoretic notions is the only or even the best way to proceed with creating a theory of the lexicon. This should be seen as a categorial adoption of useful distinctions made available from a framework which has benefitted from extensive cross-linguistic work, as well as an attempt to bridge the dialogue gap between the approaches. In actuality, I am simply using labels familiar from GB to name types which have already been used in HPSG, and that should hopefully improve the compatibility of HPSG and CCG analyses.

This substantive refinement of atomic categories is particularly important for the analysis of Dutch provided in Chapter 5. It should be noted that the feature structures of atomic categories must not be unbounded — otherwise there is potential to increase the generative power of the system. Complex categories are built as in Definition 1 on page 15 using the more articulated notion of atomic categories provided in Figure 3.1. For the remainder of this dissertation, all of the categories discussed are defined with respect to the hierarchy given in Figure 3.1.

3.2 Structuring the Lexicon

Even with a theory of lexical categories, CCG still lacks an explicit theory of the lexicon which declares how the lexicon is structured so that systematic relationships are encoded with minimal redundancy. The need for such a theory is not disputed, but it has been thus far neglected in the CCG tradition. Fortunately, work in other categorial frameworks has addressed this issue, and this work can be readily incorporated into a CCG-based approach.

Steedman (2000b) provides one principle that guides the form which the lexicon may take:

(117) *The Principle of Head Categorial Uniqueness*
A single nondisjunctive lexical category for the head 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.

This principle forces a strong minimality condition on the size of the lexicon. Recalling the discussion on the limitations of the AB calculus in §2.1.4, it entirely rules out the possibility noted there of using categorial ambiguity to overcome limitations of the grammatical machinery. Even with the extra power granted by the combinatory rules of CCG, this principle serves as a guiding force to ensure that the lexicon contains, to the maximum extent possible, only one category per head. This rules out, for example, the use of an extra category for topicalization such as $(\mathsf{s}_c\backslash\mathsf{np}^{\mathbf{Pat}})\backslash\mathsf{np}^{\mathbf{Act}}$ in addition to the usually assumed $(\mathsf{s}_i\backslash\mathsf{np}^{\mathbf{Act}})/\mathsf{np}^{\mathbf{Pat}}$.[1] This contrasts with the TAG approach of providing numerous elementary trees to cover the use of a given head in different contexts.

The Principle of Head Categorial Uniqueness is not inviolable, and at times it is necessary to provide extra categories in order to capture the data. However, it nonetheless penalizes any analysis which does so, and we are to generally disprefer analyses which use many categories to those which accomplish the same task with fewer.

Though (117) cuts down the number of categories we can place in the lexicon, it says nothing about their organization and how information is shared or distributed throughout it. An idea for how one aspect of the lexicon can be structured comes from Foster (1990). He defines a generalized categorial grammar which allows immediate dominance and linear precedence statements (Gazdar *et al.*, 1985) to be declared over partially ordered category definitions. Foster's lexicon thus projects underspecified categories into sets of ordered categories which can be used by the grammatical machinery. Each category comes with constraints that define how these ordered categories take form and the same constraints will be operative for different categories. It thus is crucial for Foster to exploit redundancy so that the same constraints do not need to be restated for each category. To overcome this, Foster suggests that the lexicon is structured into a lattice of categories so that information is shared between them. For example, the category for intransitive verbs contains pointers to the atomic categories s and np, and the category for transitive verbs contains a pointer to the category for intransitive verbs.

Though Foster does not explicate how these pointers would be implemented in a computational lexicon, the idea of course finds a very natural expression in the inheritance hierarchies of typed feature structures used in HPSG (Pollard and Sag, 1987). As Villavicencio (2002) shows, categories defined as typed default feature structures (TDFSs) (Lascarides *et al.*, 1996) can be organized in an inheritance hierarchy in which categories inherit the bulk of their definition from categories of lesser arity. She proposes the hierarchy in Figure 3.2, based on that of Pollard and Sag (1987), for structuring the categories of the English lexicon. Each subtype in

[1] The use of dependency relations to differentiate syntactic arguments is discussed in §2.6.

this hierarchy extends the TDFS of its parent by adding a new argument to the subcategorization list.

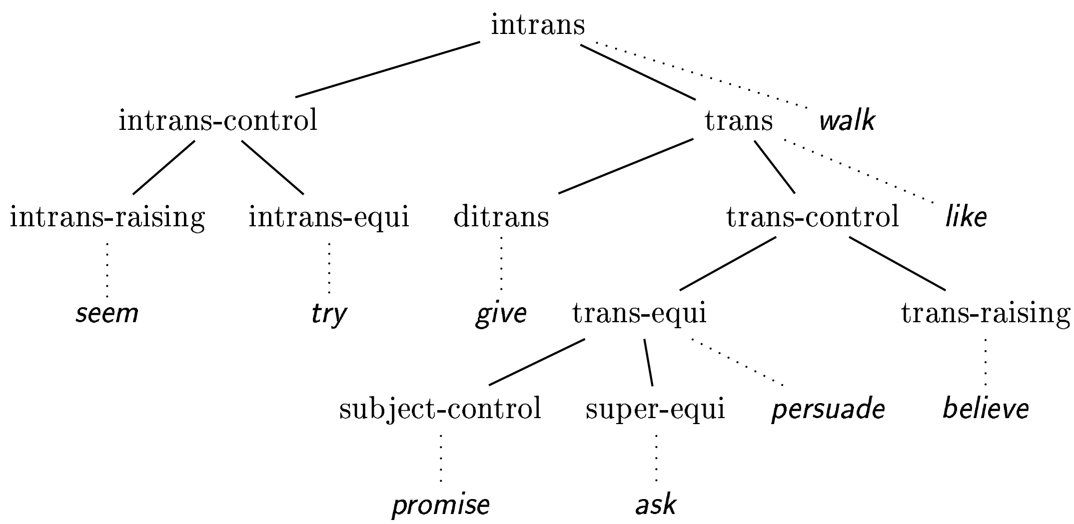


Figure 3.2: Villavicencio's verbal hierarchy for the English lexicon

Villavicencio (2002) shows that the use of such a hierarchy has important implications for language acquisition. Once the information defining the intransitive category has been set, it is propagated to all subtypes automatically. Villavicencio exploits the hierarchical structure of her lexicon to implement a Principles and Parameters-based learning theory in which the acquisition of the lexicon amounts to setting parameters which are left underspecified in the feature structures that make up the hierarchy. For example, one parameter which must be set is the general directionality of the arguments in the lexicon — once set, the value is propagated to all categories. However, because Villavicencio uses *default* inheritance, the general directionality parameter can be overridden by, for example, the directionality of the subject in the intransitive category.

Defining categories via an inheritance hierarchy furthermore simplifies the formulation of a linking theory. As syntactic arguments are incrementally added to each subtype, their semantic import is also specified. This ensures that linking generalizations are encoded higher in the hierarchy and are subsequently propagated to subtypes. Following Davis (1996), Villavicencio implements this with a separate hierarchy for linking types. It is also possible to build the linking theory directly into the hierarchy of lexical categories. Taking advantage of this aspect of the hierarchy requires a flexible representation for logical forms that permits new aspects of meaning to be added incrementally. For example, a transitive verb typically adds a **Patient** argument to the syntax, the semantics of which must be incorporated with the semantic component of the intransitive verb type which the transitive verb type inherits from. Villavicencio employs Minimal Recursion Semantics (Copestake *et al.*, 1999, 2001; Copestake, 2002) for these purposes. I utilize Hybrid Logic Dependency Semantics (Kruijff, 2001; Baldridge and Kruijff, 2002), as discussed in Chapter 8.

From the perspective of writing large grammars for natural languages, using an inheritance hierarchy of this nature translates into tremendous benefits for grammar creation and maintenence. If the analysis of particular phenomena changes (as is often the case in the course of wide-coverage grammars), it can be modified in just a few places since the hierarchy encodes the shared structure of the lexical categories. The potential to introduce errors is also reduced since type-checking keeps many inconsistencies from being introduced as changes are made. The actual form which any given category takes as a result of the information encoded in the hierarchy can then be checked with a grammar development environment such as the LKB (Copestake, 2002), which shows not only the hierarchy, but also allows the resulting categories in their expanded form to be inspected.

The evolution of the XTAG system (XTAG-group, 1999; Doran *et al.*, 2000), which processes TAG grammars, provides an excellent example of the need to develop robust mechanisms for avoiding redundancy in the declaration of lexicons. A great deal of information is represented redundantly in XTAG, which frequently led to the laborious task of changing the same basic specification in hundreds of places for the wide-coverage English grammar. While this could sometimes be ameliorated by writing simple scripts to match and replace certain patterns, the upkeep of unstructured lexicons is certainly far more onerous than for those which encode information only once and propagate it appropriately. For these reasons, recent developments on the XTAG system have focused on improving this situation by defining grammars in terms of *blocks* which are used to generate elementary trees (Doran *et al.*, 2000) and using metarules to generate all members of a tree family from a small set of base trees (Becker, 2000). Also, Wintner and Sarkar (2002) created a *post hoc* type-signature for the untyped feature structures in the XTAG English grammar and used it to check the correctness of the features declared in the grammar. This exercise revealed many errors in the grammar and thus provides a practical example of the value of using a strict typing regime on feature structures.

Another important component in a theory of the lexicon is a system of lexical rules, which Villavicencio (2002) exploits to handle certain morphological processes and also to encode inversion, imperative and dative alternations. Thus, the regular-ity in which verbs of all arities drop their subject argument in their imperative form is expressed just once in the lexical rule for forming imperative categories.

There are thus several important ways in which regularities can be captured in the lexicon, which is absolutely essential for an extremely lexicalist perspective such as that assumed in categorial grammars. I will leave aside discussion on the structure of the lexicon for most of this dissertation since the discussion will primarily focus on what categories can do once we have them in hand and are ready to use them. However, it should be kept in mind that this in no way implies that a theory of the lexicon is not viewed as an important component in formulating a linguistic theory, and indeed I assume that something like Villavicencio's structured lexicon underlies the categories which I discuss. I return to this issue in Chapter 8, discussing some of the limitations of Villavicencio's approach with respect to cross-linguistic concerns and showing how the developments of the dissertation relate to providing a structured lexicon.

3.3 Substantive Constraints on Combinatory Rules

The type-raising rules have a peculiar tendency of striking fear into the hearts of some and loathing in the minds of others. Indeed, if type-raising is completely general, it can lead to problems such as two categories being type-raised over each other *ad infinitum*. The good news, however, is that in practice we do not actually need to do anything of the sort. Steedman (2000b) suggests that the possible set of categories that can be created by type-raising rules is substantively restricted to functions over categories which are parametrically licensed for the given grammar. Thus, type-raising in English cannot create the category $(\mathsf{s}\backslash\mathsf{np})/((\mathsf{s}\backslash\mathsf{np})\backslash\mathsf{np})$ since no category exists or can be created which seeks two noun-phrase arguments to the left.

For this dissertation, I will actually assume an even more limited form of type-raising that restricts the result category to be one which is rooted in a sentential category with a filled specifier. The type-raising rules are thus given in their more specific form as follows:

(118) *Forward type-raising* ($>\mathbf{T}$)

$\mathsf{X} \;\Rightarrow_{\mathbf{T}}\; \ddot{\mathsf{s}}\$_1/(\ddot{\mathsf{s}}\$_1\backslash\mathsf{X})$

where $\ddot{\mathsf{s}}\$\backslash\mathsf{X}$ is a parametrically licensed category for the language

(119) *Backward type-raising* ($>\mathbf{T}$)

$$\mathsf{X} \;\Rightarrow_{\mathbf{T}}\; \ddot{\mathsf{s}}\$_1\backslash(\ddot{\mathsf{s}}\$_1/\mathsf{X})$$

where $\ddot{\mathsf{s}}\$/\mathsf{X}$ is a parametrically licensed category for the language

This extra, substantive constraint on type-raising blocks the creation of type-raised categories which seek nominally rooted functions, such as $\mathsf{n}/(\mathsf{n}\backslash\mathsf{n})$. While such categories have been invoked in past work, the local associativity introduced by categories with multisets takes away the need for them, as discussed in §6.3.

Using such a specific instantiation of the type-raising rule might appear to some to be an unfair shackling of the system's general potential. However, it constitutes a substantive universal and thus is intended to hold for the grammars of all natural languages. While not as satisfying as a formal universal that arises directly from the system, it nonetheless constitutes a step toward narrowing the system to all and only natural language grammars.

Standing in contrast to a universal constraint such as this, the theory of CCG assumed by Steedman (2000b) also permits restrictions to be placed on the applicability of any of the combinatory rules in any grammar. For example, Steedman limits permutativity in the grammar of English by defining backward crossed composition as in (120) so that it cannot combine nominally rooted categories.

(120) *Backward crossed composition* ($<\mathbf{B}_\times$)

$$\mathsf{Y}/\mathsf{Z} \;\; \mathsf{X}\backslash\mathsf{Y} \;\Rightarrow_{\mathbf{B}}\; \mathsf{X}/\mathsf{Z}$$

where $\mathsf{X} = \mathsf{Y} = \mathsf{s}\backslash\$$

This blocks the grammar of English from providing a derivation for a string like **a powerful by Rivaldo shot* that produces the same interpretation as *a powerful shot by Rivaldo*. See §5.2.4 for more details.

The ability to place restrictions on rules in this manner significantly increases the number of ways in which grammars for different languages can differ parametrically according to the theory. Nonetheless, it is often the case that the same restrictions must be stated on multiple rules for a given grammar, and many restrictions are the same across different grammars. This situation thus appears to beg for substantive universals to limit the space of possible restrictions and predict the co-occurence of certain rules in individual grammars and across grammars. However, rather than taking this route, I instead propose that the control achieved by rule restrictions should be projected from the lexicon by incorporating the multi-modal perspective on grammatical composition provided by CTL. Under this formulation, the rules can only apply to the resources that are appropriate for them, rather than being mediated by a substantive theory of restrictions and rule interactions. A major consequence of this move is that we have the *option* to adopt an invariant rule component and thereby push *all* variation into the lexicon. Though this heightens the need for a structured lexicon, it casts the CCG rules as formal universals rather than parametric options and increases the predictive power of the theory as a formal device, as demonstrated in Chapter 5 (particularly in §5.4) and the linguistic analyses of Turkish, Tagalog, and Toba Batak in later chapters.

3.4 Summary

This chapter has explored part of the space of possible substantive universals available to the CCG theory. I first looked at constraints on the class of lexical categories in order to refine it to those needed for natural language. After discussing general principles defined by Steedman (2000b), I outlined a basic methodology for translating GB-theoretic structures into categorial lexicons. Rather than use the results of this sort of translation directly, however, I recast them in terms of distinctions imported from work in HPSG and defined a hierarchy for atomic categories that allows for significant underspecification when using these types. Next, I discussed previous work on structuring the lexicon, which further reduces the scope of cross-linguistic variation and guides the process of constructing lexicons for language learning. Finally, I discussed constraints placed on the CCG rule-base, suggesting on the one hand that the rules for type-raising should be *universally* restricted in their application, but arguing on the other that the previous practice in CCG of restricting rules for individual grammars should give way to a universally applicable set of rules.

In the next chapter, I discuss two major linguistic phenomena – scrambling and syntactic asymmetries – in English, Tagalog, Toba Batak, and Turkish. While CCG offers promising approaches to these phenomena, they nonetheless point to certain inadequacies of the formalism and thereby provide the impetus for modifications to CCG which relax the strict order dictated by categories constructed according to Definition 1 while permitting fine-grained lexical control over the applicability of the CCG rules.

Chapter 4

Linguistic Motivation

This chapter outlines the linguistic phenomena which motivate this dissertation and discusses some previous approaches to them. In particular, the phenomena considered here are the syntactic extraction asymmetries exhibited by English, Tagalog and Toba Batak and the scrambling behavior observed in Tagalog and Turkish. These phenomena highlight the advantages and disadvantages of different categorial approaches, and therefore act as the principal motivators for the steps taken in this dissertation to define MULTI-MODAL CCG, a categorial framework which inherits the advantages of the various perspectives. MULTI-MODAL CCG in turn permits a unified explanation of the appearance of syntactic asymmetries in parametrically diverse languages and provides an intuitive analysis of local scrambling which generalizes to limited long distance scrambling.

I first discuss argument extraction asymmetries, in which semantically coherent logical forms cannot be realized linguistically because of restrictions on syntactic form. Then, I outline scrambling, a phenomenon which permits the same propositional content to be packaged in different ways.

4.1 Extraction Asymmetries

§4.1.1 introduces the well known subject/object asymmetry in extraction from English relative clauses and illustrates the CCG explanation of the asymmetry. §4.1.2 then discusses the striking asymmetries which appear in the Austronesian languages Tagalog and Toba Batak.

4.1.1 An Asymmetry in English Relative Clauses

The syntactic types of English verbs treat subjects differently from all other arguments in categorial grammars – subjects are sought to the left, while the others are sought to the right, as is evident in the following standardly assumed categories:[1]

(121) a. *defeated* $\vdash (\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np}$

b. *gave* $\vdash ((\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np})/\mathsf{np}$

c. *knew* $\vdash (\mathsf{s}_i\backslash\mathsf{np})/\mathsf{s}_{fin}$

d. *told* $\vdash ((\mathsf{s}_i\backslash\mathsf{np})/\mathsf{s}_{fin})/\mathsf{np}$

This fact in combination with the directional sensitivity of the combinatory rules of CCG provides a compelling explanation of the asymmetry between the subject and the object with respect to relativization from embedded clauses, also known as the Fixed Subject Constraint (Bresnan, 1977). Under the analysis of Steedman (1996) for relative clauses in English, this asymmetry falls out as a prediction. The relativizer *that* has two categories in English, given in (114) and repeated here:

(122) a. *that* $\vdash (\mathsf{n}\backslash\mathsf{n})/(\mathsf{s}_i\backslash\mathsf{np})$

b. *that* $\vdash (\mathsf{n}\backslash\mathsf{n})/(\mathsf{s}_i/\mathsf{np})$

With these categories and the rules of CCG, we have the derivations in (123) and (124) of a subject relative and object relative, respectively.

(123)
$$\begin{array}{llll}
\textit{team} & \textit{that} & \textit{defeated} & \textit{China} \\
\hline
 & (\mathsf{n}\backslash\mathsf{n})/(\mathsf{s}_i\backslash\mathsf{np}) & (\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np} & \mathsf{np} \\
\hline
 & & \multicolumn{2}{c}{\mathsf{s}_i\backslash\mathsf{np}} \quad > \\
\hline
 & \multicolumn{3}{c}{\mathsf{n}\backslash\mathsf{n}} \quad >
\end{array}$$

(124)
$$\begin{array}{llll}
\textit{team} & \textit{that} & \textit{Brazil} & \textit{defeated} \\
\hline
 & (\mathsf{n}\backslash\mathsf{n})/(\mathsf{s}_i/\mathsf{np}) & \mathsf{np} & (\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np} \\
\hline
 & & \ddot{\mathsf{s}}_i/(\ddot{\mathsf{s}}_i\backslash\mathsf{np}) \quad >\mathbf{T} & \\
\hline
 & & \multicolumn{2}{c}{\ddot{\mathsf{s}}_i/\mathsf{np}} \quad >\mathbf{B} \\
\hline
 & \multicolumn{3}{c}{\mathsf{n}\backslash\mathsf{n}} \quad >
\end{array}$$

The subject/object asymmetry arises when we embed the extraction site, as in (125).

[1]Note that the atomic categories used here follow the hierarchy given in Figure 3.1 on page 59 and the conventions of §3.1.

(125) a. team$_i$ that John knew that Brazil would defeat t_i

b. *team$_i$ that John knew that t_i would defeat China

The object extraction in (125a) is fine, whereas the subject extraction in (125b) is not. This state of affairs is predicted for a configurational language with a lexicon that has categories which seek subjects to the left and all other complements to the right. To see this, first consider the derivation for (125a), which uses only the order preserving rules of CCG:

(126)
$$\begin{array}{ccccccc}
\textit{team} & \textit{that} & \textit{John} & \textit{knew} & \textit{that} & \textit{Brazil} & \textit{would defeat}\\
\cline{2-7}
 & (\mathsf{n}\backslash\mathsf{n})/(\mathsf{s}_i/\mathsf{np}) & \mathsf{np} & (\mathsf{s}_i\backslash\mathsf{np})/\mathsf{s}_{fin} & \ddot{\mathsf{s}}_c/\mathsf{s}_i & \mathsf{np} & (\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np}\\
\cline{3-3}\cline{6-6}
 & & {>}\mathbf{T} & & & {>}\mathbf{T} & \\
 & & \ddot{\mathsf{s}}_i/(\ddot{\mathsf{s}}_i\backslash\mathsf{np}) & & & \ddot{\mathsf{s}}_i/(\ddot{\mathsf{s}}_i\backslash\mathsf{np}) & \\
\cline{3-4}\cline{6-7}
 & & \multicolumn{2}{c}{{>}\mathbf{B}} & & \multicolumn{2}{c}{{>}\mathbf{B}}\\
 & & \multicolumn{2}{c}{\ddot{\mathsf{s}}_i/\mathsf{s}_{fin}} & & \multicolumn{2}{c}{\ddot{\mathsf{s}}_i/\mathsf{np}}\\
\cline{5-7}
 & & & & \multicolumn{3}{c}{{>}\mathbf{B}}\\
 & & & & \multicolumn{3}{c}{\ddot{\mathsf{s}}_c/\mathsf{np}}\\
\cline{3-7}
 & & \multicolumn{5}{c}{{>}\mathbf{B}}\\
 & & \multicolumn{5}{c}{\ddot{\mathsf{s}}_i/\mathsf{np}}\\
\cline{2-7}
 & \multicolumn{6}{c}{{>}}\\
 & \multicolumn{6}{c}{\mathsf{n}\backslash\mathsf{n}}
\end{array}$$

To extract an embedded subject, however, the forward crossed composition rule $>\mathbf{B}_\times$ (54), repeated here as (127), is required to compose the complementizer with the embedded sentence that is still missing its subject. Since CCG permits any of the combinatory rules to be entirely banned for any given grammar, we can assume that forward crossed composition is banned in English and therefore fail to derive (125a), as seen below:

(127) *Forward crossed composition* ($>\mathbf{B}_\times$)

$$\mathsf{X/Y} \quad \mathsf{Y\backslash Z} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X\backslash Z}$$

(128)
$$\begin{array}{ccccc}
\textit{*team} & \textit{that} & \textit{John knew} & \textit{that} & \textit{would defeat China}\\
\cline{2-5}
 & (\mathsf{n}\backslash\mathsf{n})/(\mathsf{s}_i\backslash\mathsf{np}) & \ddot{\mathsf{s}}_i/\mathsf{s}_{fin} & \ddot{\mathsf{s}}_c/\mathsf{s}_i & \mathsf{s}_i\backslash\mathsf{np}\\
\cline{3-4}
 & & \multicolumn{2}{c}{{>}\mathbf{B}} & \\
 & & \multicolumn{2}{c}{\ddot{\mathsf{s}}_i/\mathsf{s}_i} & \\
\cline{3-5}
 & & \multicolumn{3}{c}{*}
\end{array}$$

The extraction in (125b) is thus blocked, whereas it would have succeeded if the grammar of English did include forward crossed composition.

Steedman points out that while the rule of forward crossed composition is made available in CCG, we could not use it to specify a language which looks exactly like English except that it allows subject extraction. By including this rule, we would get not only embedded subject extraction where it was blocked before, but also a collapse of word order which would induce scrambled sentences such as (129), as shown in derivation (130). The noun phrases have been indexed only to show the dependencies and do not perform any syntactic work.

(129) John Brazil$_i$ knew that t_i would defeat China.

(130)

$$
\begin{array}{ccccc}
\textit{*John} & \textit{Brazil} & \textit{knew} & \textit{that} & \textit{would defeat China} \\
\hline
\mathsf{np}_1 & \mathsf{np}_2 & (\mathsf{s}_i\backslash\mathsf{np}_1)/\mathsf{s}_{fin} & \ddot{\mathsf{s}}_c/\mathsf{s}_i & \mathsf{s}_i\backslash\mathsf{np}_2 \\
 & & \multicolumn{2}{c}{\underline{\qquad\qquad\qquad}{>}\mathbf{B}} & \\
 & & \multicolumn{2}{c}{(\mathsf{s}_i\backslash\mathsf{np}_1)/\mathsf{s}_i} & \\
 & & \multicolumn{3}{c}{*{*}{*}\underline{\qquad\qquad\qquad\qquad\qquad}{>}\mathbf{B}_\times\ *{*}{*}} \\
 & & \multicolumn{3}{c}{(\mathsf{s}_i\backslash\mathsf{np}_1)\backslash\mathsf{np}_2} \\
 & \multicolumn{3}{c}{\underline{\qquad\qquad\qquad\qquad}{<}} & \\
 & \multicolumn{3}{c}{\mathsf{s}_i\backslash\mathsf{np}_1} & \\
\multicolumn{3}{c}{\underline{\qquad\qquad\qquad\qquad}{<}} & & \\
\multicolumn{3}{c}{\mathsf{s}_i} & &
\end{array}
$$

Thus, the ban on forward crossed composition is a forced move, and the subject/object asymmetry then arises as a side effect. Steedman claims that it is a necessary property of SVO languages which have verbal categories that seek subjects and objects in different directions. Though this is an attractive result, CCG actually has an escape hatch which could allow subject extraction for such languages – namely, the ability to use features that control the applicability of the combinatory rules. It is in fact this strategy that Steedman uses to allow subjects to be extracted when the complementizer is absent. Consider the following grammatical relativization, which differs from (125b) only in that the complementizer is missing:

(131) team$_i$ that John knew t_i would defeat China

Steedman (1996) accounts for such extractions by employing a feature $\pm$ANT and giving an extra category to *knew*. This category incorporates the extracted subject and consumes an verb phrase instead of a saturated sentential category:

(132) $\textit{knew} \vdash ((\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np}_{+ANT})/(\mathsf{s}_{fin}\backslash\mathsf{np})$

The feature +ANT marks an argument as being *antecedent governed*, which means that it cannot be applied to a lexical category due to the assumption that all lexical noun phrases have the value −ANT. However, extracting categories such as relativizers actually select for sentential categories which are missing +ANT noun phrases, thereby allowing the following derivation for (131):

(133)

team	that	John	knew	would defeat China
	$(\mathsf{n}\backslash\mathsf{n})/(\mathsf{s}_i/\mathsf{np}_{+ANT})$	np_{-ANT}	$((\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np}_{+ANT})/(\mathsf{s}_{fin}\backslash\mathsf{np})$	$\mathsf{s}_i\backslash\mathsf{np}$
		$>\mathbf{T}$: $\mathsf{np}^{\uparrow}_{-ANT}$	$>$: $(\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np}_{+ANT}$	
		$>\mathbf{B}$: $\ddot{\mathsf{s}}_i/\mathsf{np}_{+ANT}$		
	$>$: $\mathsf{n}\backslash\mathsf{n}$			

Note that this derivation uses the abbreviation of type-raised categories given in Convention 4 on page 28. In this particular case, $\mathsf{np}^{\uparrow}_{-ANT}$ abbreviates $\ddot{\mathsf{s}}_i/(\ddot{\mathsf{s}}_i\backslash\mathsf{np}_{-ANT})$.

Because of the use of $\pm$ANT, the extra category (132) for *knew* does not overgenerate the ungrammatical sentence (134) since *Brazil* in a lexical noun phrase and is thus specified as $-$ANT.

(134) *John [knew would defeat China]$_{(\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np}_{+ANT}}$ [Brazil]$_{\mathsf{np}_{-ANT}}$.

It would similarly be possible to provide an extra category for the complementizer that performs a very similar function (132) and thus permits a derivation for (125b). The important thing to note, however, is that such cases are treated as exceptions which require extra work out of the language learner, so it is not particularly surprising that the extra category would be invoked for *knew* but not for the complementizer *that*. In fact, there are some speakers for whom it appears that relativizations such as (125b) are not particularly bad.

Some SVO languages do allow embedded subject extractions parallel to (125b). However, a confounding factor for such languages is that they exhibit VOS order in addition to SVO (Rizzi, 1982; Steedman, 1996). This is the case in Italian, as the following data from Rizzi (1982) shows:

(135) Credo che verrà qualcuno.
(I) believe that will-come someone.
I believe that someone will come.

(136) Chi credi che verrà?
Who think (you) that will-come
Who do you think will come?

Sentence (135) shows that Italian subjects can be extraposed and that the verb *verrà* has the category s_i/np. Forward crossed composition is therefore not implicated in the derivation of (136) since *che* and *verrà* may combine through forward harmonic composition.

(137) Chi credi [che]$_{\ddot{\mathsf{s}}_c/\mathsf{s}_i}$ [verrà]$_{\mathsf{s}_i/\mathsf{np}}$.

Embedded subject extraction is thus possible without forward crossed composition and its associated scrambling effects in a language with a lexicon in which subjects are sought to the right. SVO word order can be obtained by assigning the subject a category that fronts the subject: $\ddot{\mathsf{s}}_i/(\mathsf{s}_i/\mathsf{np})$. This category reflects a GB analysis in which the subject has raised to a left-branching SPEC of IP.[2]

(138) $[\text{Qualcuno}]_{\ddot{\mathsf{s}}_i/(\mathsf{s}_i/\mathsf{np})}$ $[\text{verrà}]_{\mathsf{s}_i/\mathsf{np}}$.

4.1.2 Austronesian Asymmetries

As shown in the previous section, the subject/object asymmetry in English arises because of a limitation on forward crossed composition. This is unsurprising given the rigidity of English word order and the fact that forward crossed composition induces scrambled word orders. Given that restricting a permutation-inducing rule produces the English asymmetry, we next consider the Austronesian languages Tagalog and Toba Batak so that we can ultimately provide a cross-linguistic account of the manner in which asymmetries appear. The typological description of the asymmetries in these two languages is quite similar, but as we will see in Chapter 7, Tagalog's asymmetries arise because of limitations on forward harmonic composition, whereas Toba Batak's come about due to limitations of a crossed composition rule, like English.

The Austronesian family of languages includes a number of languages which exhibit strong extraction asymmetries. Tagalog and Toba Batak are particularly interesting to consider in relation to English because the asymmetry typically manifests itself with respect to certain dependency roles rather than to a specific syntactic role distinction like subject/object. Tagalog and Toba Batak verbs appear in a variety of voices, each of which marks a particular dependent as special either through position or case marking. For example, in the Active Voice form *bumili* of the Tagalog verb *bili* 'buy', the **Actor** argument recieves nominative case:

(139) Bumili ng libro ang titser.
AV-buy GEN book NOM teacher
The teacher bought a book.

Note that sentence (139) is also grammatical if the order of the arguments is switched: *bumili ang titser ng libro.*

---

[2]Alternatively, a standard SVO category such as that of English could be employed.

Interestingly, although it is possible to form a question regarding the **Actor** for a verb in Active Voice, as shown in (140), forming a question about **Patient** *ng libro* is ungrammatical (140b).

(140) a. Sino ang bumili ng libro?
who NOM AV-buy GEN book
Who bought a book?

b. *Ano ang bumili ang titser?
what NOM AV-buy NOM teacher
(for: *What did the teacher buy?*)

Toba Batak, a language spoken in northern Sumatra, exhibits very similar asymmetrical behavior, though its word order is more fixed than that of Tagalog.

(141) Mangida si John si Bill
AV-see PNM-John PNM-Bill
Bill saw John.

Like Tagalog, only the **Actor** argument may be questioned for a verb in Active Voice form.

(142) a. Ise mangida si John
who see PNM-John
Who saw John?

b. *Aha mangida si Bill
what see PNM-Bill
(for: *What did Bill see?*)

It would be rather limiting if it were not possible to ask about the **Patient** argument through some other means. To accomodate such queries, the extensive voice systems of Tagalog and Toba Batak permit a verb to single out any one of its dependents as special. For example, as an alternative to *bumili*, which is the Active Voice form of the root *bili* 'buy', Tagalog also has the Objective Voice form *binili*.

(143) Binili ng titser ang libro.
OV-buy GEN-teacher NOM-book
A teacher bought the book.

This process has been called topicalization and focussing in the literature on Tagalog, but I will instead use the term *distinguishing* to avoid the baggage that those terms come with, especially since it is not clear that the distinguished argument necessarily corresponds to any generally understood sense of *topic* or *focus*.

As can be seen, the marking on the two nominal arguments is the opposite of that in (139). Here, the **Patient** *libro* is preceded by the marker *ang* and the **Actor** *titser* by *ng*. What is particularly interesting at present about this voice alternation is that the extractability of the arguments is reversed. Contrast (140) with the following:

(144) a. Ano ang binili ng titser?
what NOM OV-buy GEN-teacher
What did a teacher buy?

b. *Sino ang binili ang libro?
who NOM OV-buy NOM-book
(for: *Who bought the book?*)

A very similar pattern arises in Toba Batak; however, instead of any change in the marking on the arguments, the distinguished argument is moved to the end of the sentence when the verb is in Objective Voice.

(145) Di-ida si Bill si John
OV-see PNM-Bill PNM-John
Bill saw John.

The order of the dependents thus changes, and now only the **Patient** argument can be extracted:

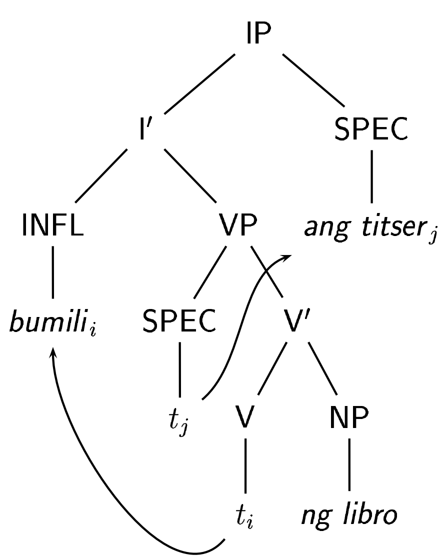
(146) a. *Ise diida si John
who OV-see PNM-John
(for: *Who saw John?*)

b. Aha diida si Bill
what OV-see PNM-Bill
What did Bill see?

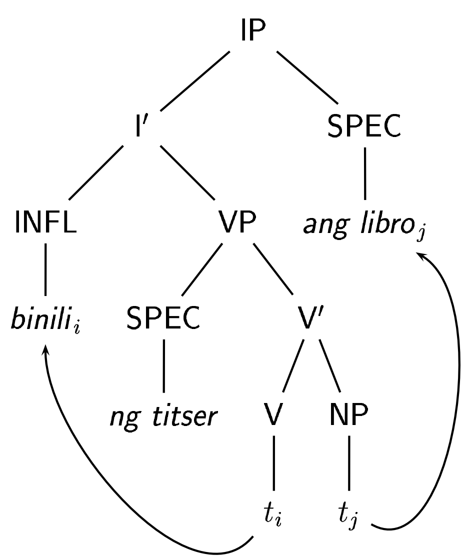
Asymmetries in Tagalog and Toba Batak are active in a much wider variety of constructions than just *wh*-extraction as shown above. In Tagalog, for example, we find asymmetrical behavior in relativization, floated quantifiers, and a fronting process called *ay*-inversion. These asymmetries have played a central role in the debate on whether or not Tagalog has an identifiable subject. According to the accessibility hierarchy of Keenan and Comrie (1977), the greater syntactic potential of nominative arguments implies that they are the subjects of Tagalog (Schachter 1976). However, as Schachter pointed out, the **Actor** argument appears to retain certain other properties normally associated with subjects even when it is not the nominative argument. The most important of these are the ability to bind reflexives and the fact that they are the most common targets of Equi-NP deletion. However, Kroeger (1993) convincingly argues that the case for the **Actor** argument as subject is greatly diminished in light of the fact that neither of the above properties is unique to **Actor** arguments. Kroeger amasses a great deal of other data which supports the view that the nominative argument is the subject. Having done this, Kroeger is able to give a concise analysis of Tagalog asymmetries in terms of grammatical relations with the Functional Uncertainty formalism of Kaplan and Zaenen (1989).

Guilfoyle *et al.* (1992) are concerned less with identifying the subject than with providing a configurational account of the apparent split in subject properties. They discuss four asymmetries—extraction, quantifier float, reflexivization, and control—and discuss how they might be captured by assuming that **Actor** arguments occupy SPEC of VP and that nominative noun phrases move to SPEC of IP. The tree configurations which Guilfoyle *et al.* assign to (139) and (143) are given in (147) and (148), respectively.

(147)



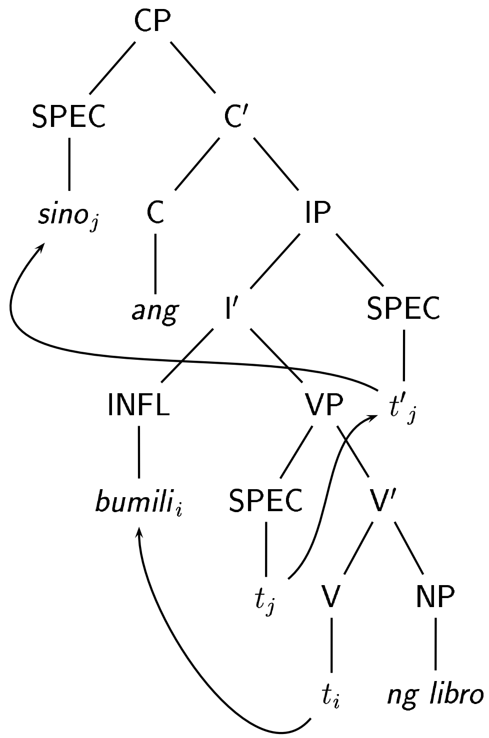
(148)



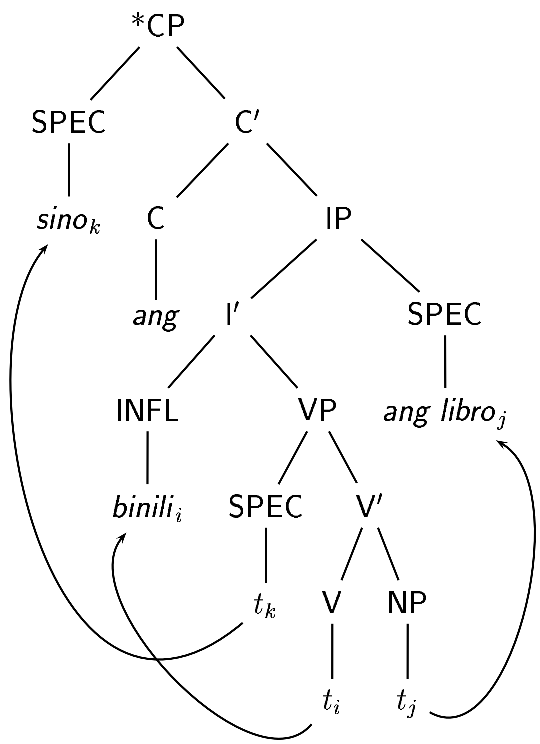
Regardless of verbal morphology, the **Actor** noun phrase is base generated in SPEC of VP and is thus always external to the V' projection. This fact in conjunction with standard binding theory assumptions explains why it is always a potential antecedent for reflexives and is always a potential target of Equi-NP deletion. On the other hand, the fact that the nominative argument may always raise to SPEC of IP is said to be responsible for its unique ability to be extracted and to launch floating quantifiers.

Guilfoyle *et al.* thus identify SPEC of IP as a pivotal configurational position for extractability, suggesting that "extraction obviously requires some type of locality to prevent movement out of the VP across a filled SPEC of IP" (p. 393). Nakamura (1994, 1998) adopts Guilfoyle *et al.*'s analysis of Tagalog phrase structure and provides a detailed syntactic account of Tagalog's asymmetries in terms of representational economy, arguing that the licit extraction (140a) competes with and subsequently blocks the illicit (144b). According to Nakamura, the fundamental reason for this is that the derivation of (140a), shown in (149), involves shorter links in the movement chain for the *wh*-word *sino* than the links in the derivation of (144b), shown in (150).

(149)



(150)



The length of a chain link is defined by Nakamura as the number of maximal projections that dominate the tail but not the head. We see in (149) that the *wh*-chain (t'_j, t_j) has length 1, whereas the comparable chain (sino$_k$, t_k) in (150) has length 2. For this reason, derivation (150) is blocked by (149) in obeyance with Nakamura's formulation of the *Minimal Link Condition* (Chomsky and Lasnik, 1993), which states that one derivation blocks another if they have comparable chain links and the link from the former is shorter than the latter.

Nakamura's 1998 analysis provides extensive coverage of Tagalog's extraction facts. Kroeger's 1993 analysis cannot handle some extractions out of certain voiceless constructions in Tagalog, whereas these come as a prediction of Nakamura's analysis. Nonetheless, Nakamura uses an interpretation of the Minimal Link Condition which amounts to a powerful transderivational constraint. He argues that his Tagalog analysis counters Chomsky's 1994 claim that the Minimal Link Condition should be applied strictly locally. In Chapter 7, I demonstrate that Tagalog's asymmetries can in fact be accounted for with local, very low power mechanisms, therefore undermining Nakamura's arguments that transderivational constraints are *necessary*.

4.2 Scrambling

Despite being extraordinarily restrictive with respect to extraction phenomena, Tagalog actually permits a great deal of freedom with respect to the order of arguments within clauses. For example, a ditransitive sentence such as (151) has another five orders in which the verb's arguments may appear, as shown in (152).

(151) Nagbigay ang lalaki ng libro sa babae.
gave-AV NOM-man GEN-book DAT-woman
'The man gave the book to the woman.'

(152)
a. Nagbigay ang lalaki sa babae ng libro.
b. Nagbigay ng libro ang lalaki sa babae.
c. Nagbigay sa babae ang lalaki ng libro.
d. Nagbigay ng libro sa babae ang lalaki.
e. Nagbigay sa babae ng libro ang lalaki.

According to (Schachter and Otanes, 1972), these six sentences "include exactly the same components, are equally grammatical, and are identical in meaning" (Schachter and Otanes, 1972, p. 83). Kroeger (1993) points out, however, that word order is not completely arbitrary and is influenced by a preference for **Actor** noun phrases to precede all others, nominative noun phrases to follow all others, and heavier noun phrases to precede lighter ones.

Given the limitations on extraction in Tagalog, it is thus interesting to consider how we might reconcile the its restrictiveness with such freedom in clausal word order. In fact, this is a substantial problem for an approach such as that of Guilfoyle *et al.* (1992) and Nakamura which rely on rigid tree configurations to ensure the extractability of certain arguments in Tagalog. In Chapter 7, I show that Tagalog's asymmetries can be handled without conflating basic word order and accessibility for extraction.

I will discuss scrambling in Turkish rather than Tagalog since Turkish exhibits not only local scrambling like that in (152), but also long distance scrambling in which arguments of a subordinate clause appear in the clausal domain of higher verbs. Hoffman (1995) provides an analysis of Turkish scrambling within the framework of Multiset-CCG, an extension of CCG. I argue, like Hoffman, that CCG is not optimally defined for handling scrambling-type languages, but unlike her, I argue that mild context-sensitive power actually is adequate to handle long distance scrambling, in line with Joshi *et al.* (2000) and Kulick (2000).

4.2.1 Local Scrambling

Like Tagalog, the arguments of a Turkish verb may be freely permuted within its clausal domain, as seen in the Turkish transitive sentence (153a) and its scrambled counterpart (153b), adapted from Hoffman (1995):

(153) a. Ayse kitabi okuyor
Ayse-Nom book-Acc read-Prog

b. Kitabi Ayse okuyor

Ayse reads the book.

The essence of local scrambling runs counter to the configurational assumptions of many grammatical formalisms, and numerous proposals have been made for dealing with the variability it introduces. In the transformational tradition, scrambling is seen either as a result of argument movement (see Laenzlinger (1998) for an overview) or as base-generation of the various word orders (Miyagawa, 1997). Uszkoreit (1987) gives a Generalized Phrase Structure Grammar account in which meta-rules convert an underspecified rule system into one which produces the scrambled orders through base generation. Foster (1990) uses a similar architecture within a unification-based categorial grammar to capture word order variation in Spanish. Reape (1994) rejects the conventional assumption that surface syntax determines word order and provides a flexible mapping from syntactic and logical structures to phonology. Carpenter (1999) takes Reape's notion of shuffling and recasts it in CTL to provide an analysis of German scrambling. Also following Reape, Bröker (1998) uses word order domains in Dependency Grammar to allow a separation of linear order and hierarchical relations. The (Unification) Categorial Grammar proposal of Karttunen (1989) transfers the work of syntactic subcategorization from the verbs to the noun phrases, a move related to type-raising which immediately yields local scrambling. Using a CTL-based system, Kruijff (2001) employs dependency-based modes of composition that interact with permutation-inducing structural rules that mark the informativity of scrambled dependents to produce articulated information structures.

In CCG, the most obvious way of handling local scrambling is base generation, which amounts to lexical ambiguity for verbs that allow scrambling. For example, if we assume the Turkish lexicon contains the two categories in (154) for *okuyor* 'read', we will capture both of the word orders in (153) as shown in derivations (157) and (158).

(154) a. *okuyor* $\vdash (\mathsf{s}_i\backslash\mathsf{np}_{nom})\backslash\mathsf{np}_{acc}$

b. *okuyor* $\vdash (\mathsf{s}_i\backslash\mathsf{np}_{acc})\backslash\mathsf{np}_{nom}$

(155) *Ayse* $\vdash \mathsf{np}_{nom}$

(156) *kitabi* $\vdash \mathsf{np}_{acc}$

(157)
$$\dfrac{\dfrac{\textit{Ayse}}{\mathsf{np}_{nom}} \quad \dfrac{\dfrac{\textit{kitabi}}{\mathsf{np}_{acc}} \quad \dfrac{\textit{okuyor}}{(\mathsf{s}_i\backslash\mathsf{np}_{nom})\backslash\mathsf{np}_{acc}}}{\mathsf{s}_i\backslash\mathsf{np}_{nom}}<}{\mathsf{s}_i}<$$

(158)
$$\dfrac{\dfrac{\textit{Kitabi}}{\mathsf{np}_{acc}} \quad \dfrac{\dfrac{\textit{Ayse}}{\mathsf{np}_{nom}} \quad \dfrac{\textit{okuyor}}{(\mathsf{s}_i\backslash\mathsf{np}_{acc})\backslash\mathsf{np}_{nom}}}{\mathsf{s}_i\backslash\mathsf{np}_{acc}}<}{\mathsf{s}_i}<$$

This solution might at first appear to violate the Principle of Head Categorial Uniqueness (117), but it actually does not since the principle only governs the relationship between a given category, its canonical bounded dependencies, and the unbounded dependencies that arise from its interaction with other categories. It is thus silent on the issue of using different categories to capture multiple bounded dependencies such as local scrambling.

Even so, this solution for local scrambling is unattractive since it essentially denies the relationship between the different orders and requires a justification of each category (for handling each word order) on its own merits if the approach is not to be *ad hoc*. Also, since CCG is strongly lexicalized, such ambiguity raises the complexity of a grammar which uses it and increases the burden on the language learner. Therefore, a solution which permits the specification of multiple local word orders with a single category is to be preferred to one which uses lexical ambiguity.

Specification via a single category can be achieved in a number of ways. One is to use lexical rules that use a single core category in the lexicon to generate multiple categories for actual use in derivations. A similar strategy is advocated by Foster (1990), who defines unordered categories in the lexicon which potentially project multiple ordered categories for use by the grammar. The difference between Foster's strategy and one which uses lexical rules is that he does not require any language specific rules in order to create ordered categories from an unordered core category. Nonetheless, both approaches retain the tight connection between the different orders in a principled manner, though they do not allow for strategies that take advantage of underspecified orders via processing mechanisms. I will thus pursue a strategy which allows a single category to project multiple orders directly from the lexicon, though it should be noted that I do not wish to suggest that lexical rules and other such mechanisms should not be used for other purposes.

An alternative to multiple categories is to use type-raising and crossed composition rules to allow order variation (Bozsahin, 1998). This permits both of the orders in (153) while using only the lexical entry (154a). In derivation (157), only the rule of functional application is used for the order in which the nominative argument precedes the accusative one. In the simple AB system, we would not be able to derive the order in which the arguments are reversed with only the category $(\mathsf{s}_i\backslash\mathsf{np}_{nom})\backslash\mathsf{np}_{acc}$. With CCG, however, the extra associativity and permutativity provided by the rules of forward type-raising and forward crossed composition enable a derivation of the order in which the accusative argument precedes the nominative one , as shown below.

(159)

$$
\begin{array}{ccc}
\textit{Kitabi} & \textit{Ayse} & \textit{okuyor} \\
\hline
\mathsf{np}_{acc} & \mathsf{np}_{nom} & (\mathsf{s}_i\backslash\mathsf{np}_{nom})\backslash\mathsf{np}_{acc} \\
 & \overline{\ddot{\mathsf{s}}_i/(\ddot{\mathsf{s}}_i\backslash\mathsf{np}_{nom})}^{\,>\mathbf{T}} & \\
 & \multicolumn{2}{c}{\overline{\ddot{\mathsf{s}}_i\backslash\mathsf{np}_{acc}}^{\,>\mathbf{B}_{\times}}} \\
\multicolumn{3}{c}{\overline{\ddot{\mathsf{s}}_i}^{\,<}}
\end{array}
$$

Crossed composition has been used by Steedman to account for phenomena such as heavy-NP shift and non-peripheral argument extractions (Steedman, 1987) and Dutch crossed dependencies (Steedman, 2000b). Derivation (159) demonstrates how the non-order-preserving effect of the crossed composition rules can lead to scrambled word orders. Unsurprisingly, the crossed composition rules must be used carefully in languages which have fairly rigid word order as the discussion on the subject/object asymmetry in §4.1.1 made clear. As discussed in Chapter 5, one of the goals of this dissertation is to show how we can license controlled use of rules like forward crossed composition in the English grammar by formulating the CCG system to have greater resource-sensitivity which allows fine-grained lexical control. Under this formulation, forward crossed composition is available to the English grammar without causing overgenerations such as (129).

With the scrambling freedom exhibited by Turkish, it would not be surprising to find a non-order preserving rule such as forward crossed composition working quite freely in its grammar. Nonetheless, the solution explicated above fails when we consider argument cluster coordination such as (48). Similar constructions exist in Tagalog and Turkish, and we can appeal to a similar analysis to handle them. However, we must ensure that our verbal categories have the appropriate type to serve as arguments for the argument cluster functions. Both of the sentences in (160) (from Hoffman (1995)) are grammatical in Turkish, but – as Hoffman points out – only (160a) can be derived if the Turkish lexicon contains only the category given for *okuyor* in (154a).

(160) a. Ayse kitabi, Fatma da gazeteyi okuyor
Ayse book-ACC, Fatma too newspaper-ACC read-PROG
Ayse is reading the book, and Fatma the newspaper.

b. Kitabi Ayse, gazeteyi de Fatma okuyor
book-ACC Ayse, newspaper-ACC too Fatma read-PROG
As for the book, Ayse is reading it, and the newspaper, Fatma.

Although type-raising and forward crossed composition do allow the order in which the accusative noun phrase precedes the nominative one to be derived with the category $(s_i\backslash np_{nom})\backslash np_{acc}$, they do not permit the related argument cluster coordination. The noun phrases must compose and then coordinate in order to have any chance of deriving the sentence (160b), but once that is done, forward crossed composition cannot be used as before since the complex argument cluster expects a verbal argument with a different argument ordering. The blocked derivation is given in (161). Note that the type-raised categories of the right conjunct are abbreviated, and the two application steps needed for the coordinator have been abbreviated by the single step marked with Φ.

(161)
$$
\begin{array}{cccccc}
\textit{Kitabi} & \textit{Ayse} & , & \textit{gazeteyi} & \textit{Fatma} & \textit{okuyor} \\
\hline
np_a & np_n & (X\backslash X)/X & np_a & np_n & (s_i\backslash np_n)\backslash np_a \\
\hline
\ddot{s}_i/(\ddot{s}_i\backslash np_a)\ {>}\mathbf{T} & (\ddot{s}_i\backslash np_a)/((\ddot{s}_i\backslash np_a)\backslash np_n)\ {>}\mathbf{T} & & np_a^{\uparrow}\ {>}\mathbf{T} & np_n^{\uparrow}\ {>}\mathbf{T} & \\
\hline
\multicolumn{2}{c}{\ddot{s}_i/((\ddot{s}_i\backslash np_a)\backslash np_n)\ {>}\mathbf{B}} & & \multicolumn{2}{c}{\ddot{s}_i/((\ddot{s}_i\backslash np_a)\backslash np_n)\ {>}\mathbf{B}} & \\
\hline
\multicolumn{5}{c}{\ddot{s}_i/((\ddot{s}_i\backslash np_a)\backslash np_n)\ {<}\Phi{>}} & \\
\hline
\multicolumn{6}{r}{*}
\end{array}
$$

The coordinated argument clusters require the function $(s_i\backslash np_a)\backslash np_n$, and the only one available is $(s_i\backslash np_n)\backslash np_a$. In order to handle local scrambling, we must therefore either accept multiple lexical entries for verbs whose arguments may scramble, or add new mechanisms to the grammatical machinery.

The MULTISET-CCG extension of CCG (Hoffman, 1995) is able to account for local scrambling with a single lexical item and can still handle both coordinations in (160). It does so by redefining the categories and rules of CCG to allow the use of multisets of arguments, as discussed in §2.3. MULTISET-CCG thus relaxes the strict ordering on a verb's arguments enforced by standard categories. For example, a transitive verb in Turkish such as *okuyor* would receive the category in (162). With the rule of backward function application given in (73b) and repeated here as (163), we can derive both of the orders in (153).

(162) $\textit{okuyor} \vdash s_i\{\backslash np_{nom}, \backslash np_{acc}\}$

(163) $Y \quad X(\alpha \uplus \{\backslash Y\}) \ \Rightarrow \ X\alpha$ (<)

The above entry and rule allow derivations of the sentences in (153) as follows:

(164) a.
$$
\begin{array}{ccc}
\textit{Ayse} & \textit{kitabi} & \textit{okuyor} \\
\hline
np_{nom} & np_{acc} & s_i\{\backslash np_{nom}, \backslash np_{acc}\} \\
\hline
 & \multicolumn{2}{c}{s_i\{\backslash np_{nom}\}\ <} \\
\hline
\multicolumn{3}{c}{s_i\ <}
\end{array}
$$

b. *Kitabi* *Ayse* *okuyor*

$$\dfrac{\dfrac{}{\mathsf{np}_{acc}} \quad \dfrac{\dfrac{}{\mathsf{np}_{nom}} \quad \dfrac{}{\mathsf{s}_i\{\backslash\mathsf{np}_{nom}, \backslash\mathsf{np}_{acc}\}}}{\mathsf{s}_i\{\backslash\mathsf{np}_{acc}\}}<}{\mathsf{s}_i}<$$

MULTISET-CCG thus makes no assumptions about canonical word order, and it captures local scrambling without lexical ambiguity. It also provides analyses of unbounded long distance scrambling, as discussed in the next section.

4.2.2 Long-distance Scrambling

The term long distance scrambling describes the appearance of an argument of a lower clause intermixed with the arguments of a higher clause. The use of long distance scrambling usually indicates a specific pragmatic function or status. Hoffman (1995) gives the following examples from Turkish, in which the argument *kitabi* 'book' of the lower verb *okudugunu* 'read' scrambles out of its base position in (165a) into the matrix clause on the left (165b) and the right (165c) (from Hoffman (1995)).

(165) a. Fatma [Esra'nın kitabi okudugunu] biliyor.
Fatma [Esra-GEN book-ACC read-GER-ACC] know-PROG
Fatma knows that Esra read the book.

b. *Kitabi*$_i$ Fatma [Esra'nın t$_i$ okudugunu] biliyor.
book-ACC$_i$ Fatma [Esra-GEN t$_i$ read-GER-ACC] know-PROG
As for the book, Fatma knows that Esra read it.

c. Fatma [Esra'nın t$_i$ okudugunu] biliyor *kitabi*$_i$.
Fatma [Esra-GEN t$_i$ read-GER-ACC] know-PROG *book*-ACC$_i$
Fatma knows that Esra read it, the book.

MULTISET-CCG is able to derive such sentences through the use of its composition rules, which join the elements of two multisets.

(166) MULTISET-CCG *composition rules*

a. $\mathsf{X}(\alpha \uplus \{/\mathsf{Y}\}) \quad \mathsf{Y}\beta \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X}(\alpha \uplus \beta)$ (>**B**)

b. $\mathsf{Y}\beta \quad \mathsf{X}(\alpha \uplus \{\backslash\mathsf{Y}\}) \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X}(\alpha \uplus \beta)$ (<**B**)

Consider MULTISET-CCG's backward composition rule (166b) and its use in deriving (165b), shown in derivation (167).

(167) *Kitabi* *Fatma* *Esra'nın* *okudugunu* *biliyor*

$$
\begin{array}{lllll}
\mathsf{np}_{acc} & \mathsf{np}_{nom} & \mathsf{np}_{gen} & \mathsf{s}_{v\,acc}\{\backslash\mathsf{np}_{gen}, \backslash\mathsf{np}_{acc}\} & \mathsf{s}_i\{\backslash\mathsf{np}_{nom}, \backslash\mathsf{s}_{v\,acc}\} \\
 & & \multicolumn{2}{c}{\mathsf{s}_{v\,acc}\{\backslash\mathsf{np}_{acc}\}\ <} & \\
 & & \multicolumn{3}{c}{\mathsf{s}_i\{\backslash\mathsf{np}_{nom}, \backslash\mathsf{np}_{acc}\}\ <\mathbf{B}} \\
 & \multicolumn{4}{c}{\mathsf{s}_i\{\backslash\mathsf{np}_{acc}\}\ <} \\
\multicolumn{5}{c}{\mathsf{s}_i\ <}
\end{array}
$$

The crucial step is when *biliyor* composes with the lower verb *okudugunu*, which is still missing its accusative argument. The composition rule (166b) places the nominative argument of *biliyor* in the same multiset with the accusative argument of *okudugunu*, thereafter rendering them indistinguishable with respect to the order in which they are sought. Effectively, this definition of composition performs a *domain union* operation (Reape, 1994) which merges the arguments of the daughter clause with those of the mother.

A further aspect of long distance scrambling is that more than one element of the same clause may undergo long distance scrambling. In the following sentence, *bu kitabi* has scrambled to the front and *benim* to the end.

(168) Bu kitabi$_j$ Fatma [t$_i$ [t$_j$ okumak] istedigimi] biliyor benim$_i$
This book-ACC$_j$ Fatma [t$_i$ [t$_j$ read-INF] want-GER-ACC know-PROG I-GEN$_i$
As for this book, Fatma knows that I want to read it.

Hoffman recognizes that sentences become increasingly difficult to process as arguments scramble further from their local positions. However, she claims, following Rambow (1994), that there is no absolute limit on the distance which an element may scramble nor on the number of elements which may scramble, and she therefore claims such difficulties should be viewed as a processing limitation rather than an inherent aspect of the competence grammar. Unbounded scrambling as such may be abstractly described by the following general schema, where each v_i subcategorizes for n_i (Joshi *et al.*, 2000).

(169) $\sigma(\mathsf{n}_1, \mathsf{n}_2, \ldots \mathsf{n}_n)\ \mathsf{v}_n, \mathsf{v}_{n-1}, \ldots \mathsf{v}_1$ (σ a permutation)

Rambow (1994) proves that such strings require greater than mildly context-sensitive power in order to assign appropriate structural descriptions to them. He defines Vector-TAG, an extension of TAG which meets this criteria by relaxing immediate dominance constraints in elementary trees. This work has been followed more recently by Joshi *et al.* (2000), who argue that the increase in power which Rambow proposed was not necessarily warranted. They present a variant of the TAG formalism, Tree-local Multi-component TAG, which not only handles local scrambling behavior, but also exhibits limitations on the depth of long distance scrambling as a property of the competence grammar.

MULTISET-CCG's rejection of the strict order engendered by standard CCG categories provides the inspiration for the steps taken in Chapter 6. However, since MULTISET-CCG is more powerful than CCG, I follow Joshi *et al.* (2000) in seeking a formulation of CCG which benefits from the attractive aspects of MULTISET-CCG while maintaining its mildly context-sensitive position on the Chomsky hierarchy. This adherence to restricted generative capacity excludes ungrammatical Turkish scrambling sentences without certain stipulations required by Hoffman. §6.2 shows that rejecting the extra power engendered by MULTISET-CCG's rules of composition and type-raising does not limit my formulation from handling the levels of scrambling for which native speaker judgements can be reliably obtained.

4.3 Limits on Permutativity

Though MULTISET-CCG retains versions of CCG's rules of type-raising and composition, it actually *loses* resource-sensitivity in comparison with CCG because it conflates the harmonic and crossed composition rules into just a forward and backward variety. Hoffman's own data shows that the grammar of Turkish does require sensitivity to this distinction — even though the arguments of verbs can permute quite freely, the arguments of adjectives such as *siyah* 'black' must appear to their immediate right (Hoffman's (27)):

(170) a. [Siyah kedi] geldi.
[black cat] come-PAST
The black cat came in.

b. *Kedi siyah geldi.
cat black come-PAST

c. *Siyah geldi kedi.
black come-PAST cat

Hoffman gives the adjective *siyah* 'black' the MULTISET-CCG category $\mathsf{np}_x\{/\mathsf{np}_x\}$, which excludes the order in (170b). However, the backward composition rule of MULTISET-CCG (166b) does allow a derivation of (170c), demonstrated below:

(171)

$$
\begin{array}{ccc}
\textsf{\textit{siyah}} & \textsf{\textit{geldi}} & \textsf{\textit{kedi}} \\
\hline
\mathsf{np}_x\{/\mathsf{np}_x\} & \mathsf{s}_i\{\backslash\mathsf{np}_{nom}\} & \mathsf{np}_{nom} \\
\hline
\multicolumn{2}{c}{\mathsf{s}_i\{/\mathsf{np}_{nom}\}} \; {<}\mathbf{B} & \\
\hline
\multicolumn{3}{c}{\mathsf{s}_i} \; {>}
\end{array}
$$

In order to block this derivation, Hoffman places a restriction on the backward composition rule which blocks this derivation:

(172) *Turkish Backward Composition* ($<\mathbf{B}$)

$$\mathsf{Y}\beta \quad \mathsf{X}(\alpha \uplus \{\backslash\mathsf{Y}\}) \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X}(\alpha \uplus \beta)$$

(except when $\mathsf{X} = \mathsf{s}$ and $\mathsf{Y}\beta = \mathsf{np}\{/\mathsf{np}\}$)

Instead of utilizing such a restriction, the extension of CCG developed in Chapter 5 and Chapter 6 blocks such derivations by the fact that the category of adjectives such as *siyah* 'black' does not permit other elements to come between an adjective and its argument. The specifications used to achieve this for Turkish are precisely the same ones which are employed in the English lexicon to allow adverbs to come between a verb and its direct object (173) while blocking a postnominal modifier from coming between an adjective and its nominal argument (174).

(173) a. Kahn blocked a powerful shot by Rivaldo skillfully.

b. Kahn blocked skillfully a powerful shot by Rivaldo.

(174) a. a powerful shot by Rivaldo

b. *a powerful by Rivaldo shot

We thus see that despite its liberalness with respect to argument order, Turkish does require control over other aspects of word order in a way that is similar to English.

4.4 Summary

In this chapter, I have outlined some of the high-level linguistic motivations for the formal developments pursued in this dissertation. The phenomena discussed here show how the grammars of natural languages are pushed to be quite restrictive in some ways and pulled to allow more freedom in others. The next chapter develops a resource-sensitive formulation of CCG which provides us with the means to handle asymmetries and limits on permutativity, and then Chapter 6 provides further modifications which allow an account of Turkish and Tagalog scrambling that fulfills the desiderata laid out in §4.2. These developments lead to a shift in perspective that pushes all cross-linguistic variation into the CCG lexicon by utilizing lexically internalized control over the combinatory rules of CCG.

Chapter 5

Modal Control in CCG

This chapter shows how incorporating formal devices from the Categorial Type Logic (CTL) tradition into CCG can significantly improve the linguistic suitability of the formalism. In much the same way that the multi-modal setting of CTL provides it with heightened lexical sensitivity to the structural rules defined for a given grammar, modalities can be incorporated into CCG to improve the resource-sensitivity of CCG's rules of combination and make them universally applicable. I will refer to the resulting variant of CCG as *Multi-Modal Combinatory Categorial Grammar.*

This chapter involves several CTL proofs, so readers unfamiliar with CTL should see §2.4 for a brief introduction and pointers to more extensive expositions regarding CTL. However, do note that the main thrust of the developments in this chapter can be appreciated without a full understanding of CTL, and readers who are not especially concerned with the CTL basis of the CCG rules I present should not agonize over the specifics of the proofs.

I begin with a brief discussion of how CCG can be simulated in CTL, and then turn this simulation on its head to bring modalities into CCG. The need to do so is demonstrated with respect to a number of constructions in English, and then I show that recasting Steedman's CCG analysis of Dutch in MULTI-MODAL CCG boosts the predictive power of the analysis. This is followed by a cross-linguistic illustration of the benefits of assuming a universal rule component that shows that the CCG composition and substitution rules are parametrically related. After a brief discussion on generative power and its relation to the use of modally refined rules. The chapter finishes with full listings of the rules of MULTI-MODAL CCG and the CTL rules which back them up.

5.1 Simulating CCG in CTL

This section is based on as yet unpublished work with Geert-Jan Kruijff, circulated as a draft entitled *Relating Categorial Type Logics and CCG through Simulation* (Kruijff and Baldridge, 2000). Here I provide an abbreviated explication of the basic ideas which guide the creation of a CTL system with modalities and structural rules that allow the behavior of CCG to be simulated.

In one derivational step, the composition rules of CCG perform what requires several proof steps in CTL. That is, to simulate a rule like forward harmonic composition in CTL requires the invocation of hypothetical reasoning, elimination, associativity, and introduction in CTL. (175) shows a CCG derivation for the composition of a modal verb with a transitive verb required for coordinations such as *Brazil will meet and should defeat China*, and (176) demonstrates the proof steps needed to achieve the same result in CTL, using the base logic defined in §2.4 and the structural rule of Right Association (83) given on page 41.

(175)
$$\dfrac{\dfrac{\textit{should}}{(\mathsf{s}_i\backslash\mathsf{np})/(\mathsf{s}_v\backslash\mathsf{np})} \quad \dfrac{\textit{defeat}}{(\mathsf{s}_v\backslash\mathsf{np})/\mathsf{np}}}{(\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np}}{>}\mathbf{B}$$

(176)
$$\dfrac{\dfrac{\dfrac{\textit{should} \vdash (\mathsf{s}_i\backslash_1\mathsf{np})/_1(\mathsf{s}_v\backslash_1\mathsf{np}) \qquad \dfrac{\textit{defeat} \vdash (\mathsf{s}_v\backslash_1\mathsf{np})/_1\mathsf{np} \qquad [x_1 \vdash \mathsf{np}]^{\dagger}}{(\textit{defeat} \circ_1 x_1) \vdash \mathsf{s}_v\backslash_1\mathsf{np}}[/_1E]}{(\textit{should} \circ_1 (\textit{defeat} \circ_1 x_1)) \vdash \mathsf{s}_i\backslash_1\mathsf{np}}[/_1E]}{((\textit{should} \circ_1 \textit{defeat}) \circ_1 x_1) \vdash \mathsf{s}_i\backslash_1\mathsf{np}}[RA]}{(\textit{should} \circ_1 \textit{defeat}) \vdash (\mathsf{s}_i\backslash_1\mathsf{np})/_1\mathsf{np}}[/_1I]^{\dagger}$$

CTL thus provides a more detailed view on the process of combining two types of the form $\mathsf{X/Y}$ and $\mathsf{Y/Z}$ into one such as $\mathsf{X/Z}$.

CTL rules rely on the logical behavior which different modes of grammatical composition induce on structural configurations in order to make use of operations of associativity and permutativity in grammars. These different modes give rise to multiple slash types, each of which has a leftward and rightward instantiation. CCG implicitly uses just one mode of grammatical composition that appears as one kind of slash with its leftward and rightward varieties. Even so, the fact that CCG rules can reference two or more slashes on their input side (e.g. the composition and substitution rules) can have similar effects to the use of different modalities on structural configurations involving two modalities.

Recall that the modes {0,1} utilized in the CTL introduction (§2.4) were used in the lexicon with both leftward and rightward slashes in categories like $(\mathsf{s}\backslash_1 \mathsf{np})/_1\mathsf{np}$. Eliminating these slashes with appropriate arguments produces structures like (177) which "forget" the directionality of the slashes that produced them.

(177) (*Brazil* $\circ_1$ (*defeated* $\circ_1$ *Belgium*))

Consider then a CTL grammar that uses the modalities $\{\triangleleft, \triangleright\}$ as the means of recording the direction of their associated lexical slashes.[1] Thus, a lexical category such as the English transitive would be $(\mathsf{s}\backslash_{\triangleleft}\mathsf{n})/_{\triangleright}\mathsf{n}$. Using purely the non-associative base logic, we would have the following proof:

(178)
$$\frac{\Gamma \vdash \mathsf{X}/_{\triangleright}\mathsf{Y} \qquad \dfrac{\Delta \vdash \mathsf{Y}/_{\triangleright}\mathsf{Z} \qquad [x_1 \vdash \mathsf{Z}]}{(\Delta \ \circ_{\triangleright} \ x_1) \vdash \mathsf{Y}}[/_{\triangleright}E]}{(\Gamma \ \circ_{\triangleright} \ (\Delta \ \circ_{\triangleright} \ x_1)) \vdash \mathsf{X}}[/_{\triangleright}E]$$

The structure thus retains the information about the directionality of the slashes which were eliminated to construct it. We are now in a position to create structural rules which reference the directionality of multiple categories to simulate the effects of CCG rules. For example, the following rule of Right Association allows this system with its two directional modalities to continue the above proof as a simulation of forward harmonic composition, similar to the proof in (176).

(179) *Right Association*
$$\frac{(\Delta_a \ \circ_{\triangleright} \ (\Delta_b \ \circ_{\triangleright} \ \Delta_c)) \vdash \mathsf{X}}{((\Delta_a \ \circ_{\triangleright} \ \Delta_b) \ \circ_{\triangleright} \ \Delta_c) \vdash \mathsf{X}}[RA]$$

Without this rule, the hypothesis would be trapped and we would be unable to introduce the slash back into the type of the category to produce the result X/Z.

The advantage of using the two modalities $\{\triangleleft, \triangleright\}$ instead of a single modality is that we can formulate more powerful structural rules without engendering permutation closure. Just as CCG has rules like forward crossed composition which are non-order preserving, we can use the following interaction postulate which permutes two nodes in the structure:

(180) *Mixed Left Permutation*
$$\frac{(\Delta_a \ \circ_{\triangleright} \ (\Delta_b \ \circ_{\triangleleft} \ \Delta_c)) \ \vdash \mathsf{X}}{(\Delta_b \ \circ_{\triangleleft} \ (\Delta_a \ \circ_{\triangleright} \ \Delta_c)) \ \vdash \mathsf{X}}[MLP]$$

[1] These modalities should in no way be interpreted as encoding a head/dependent distinction.

This rule is crucial for simulating forward crossed composition in the CTL setting, as shown below:

$$(181)\quad \dfrac{\dfrac{\dfrac{\Gamma \vdash \mathsf{X}/_{\triangleright}\mathsf{Y} \qquad \dfrac{[x_1 \vdash \mathsf{Z}]^{\dagger} \qquad \Delta \vdash \mathsf{Y}\backslash_{\triangleleft}\mathsf{Z}}{(x_1 \circ_{\triangleleft} \Delta) \vdash \mathsf{Y}}[\backslash_{\triangleright}E]}{(\Gamma \circ_{\triangleright} (x_1 \circ_{\triangleleft} \Delta)) \vdash \mathsf{X}}[/_{\triangleright}E]}{(x_1 \circ_{\triangleleft} (\Gamma \circ_{\triangleright} \Delta)) \vdash \mathsf{X}}[MLP]}{(\Gamma \circ_{\triangleright} \Delta) \vdash \mathsf{X}\backslash_{\triangleleft}\mathsf{Z}}[\backslash_{\triangleleft}I]^{\dagger}$$

Note that we could also redefine the CCG rule of forward crossed composition to produce the same structure, as shown in (182).

$$(182)\quad \Gamma \vdash \mathsf{X}/_{\triangleright}\mathsf{Y} \quad \Delta \vdash \mathsf{Y}\backslash_{\triangleleft}\mathsf{Z} \quad \Rightarrow_{\mathbf{B}} \quad (\Gamma \circ_{\triangleright} \Delta) \vdash \mathsf{X}\backslash_{\triangleleft}\mathsf{Z} \qquad (>\mathbf{B}_{\times})$$

Nonetheless, unlike CTL, CCG formulated in this way would still not be able to manipulate the structures it creates. Because it is a string calculus and not a logical system, CCG can only *build* such structures.

Since every mode $\circ_i$ has a left ($\backslash_i$) and right ($/_i$) slash associated with it (known as its *residuals*), it is possible to have the slashes $\backslash_{\triangleright}$ and $/_{\triangleleft}$ in addition to $\backslash_{\triangleleft}$ and $/_{\triangleright}$. The modalities on these slashes have opposite directionality to the slash itself. In our simulation of CCG, we assume that all slashes from the lexicon have matching directionality. However, the proof for type-raising over a function with the $\backslash_{\triangleleft}$ slash shows that the latter kind of slash nonetheless arise as a property of the system.

$$(183)\quad \dfrac{\dfrac{\mathit{Brazil} \vdash \mathsf{np} \qquad [x_1 \vdash \mathsf{s}\backslash_{\triangleleft}\mathsf{np}]^{\dagger}}{(\mathit{Brazil} \circ_{\triangleleft} x_1) \vdash \mathsf{s}}[\backslash_{\triangleleft}E]}{\mathit{Brazil} \vdash \mathsf{s}/_{\triangleleft}(\mathsf{s}\backslash_{\triangleleft}\mathsf{np})}[/_{\triangleleft}I]^{\dagger}$$

Steedman (2000b) provides a unification-based argument for why type-raising must be order-preserving, and extending his discussion to include modalities should yield the same behavior as the logical system with respect to which modality will decorate the slashes of the output category.

Slashes which have oppositely directed modalities provide an interesting and necessary degree of freedom when we turn this simulation around by bringing modalities into CCG.

5.2 Modalized CCG

The simulation of CCG in CTL points to the interesting possibility of bringing modalities into CCG in order to obtain the sort of fine-grained *lexical* control available in

CTL. It is quite common to find in CCG analyses that constraints need to be stipulated on the applicability of several of the combinatory rules for that fragment. By modifying the formal definition of CCG to utilize the modalities of CTL, we can do away with the use of rule restrictions to control the applicability of CCG's rules in favor of using categories that specify different modes of grammatical composition. I will show how this permits existing CCG analyses for English and Dutch to be improved. Ultimately, this will provide us with a principled and meaningful way of characterizing the manner in which syntactic extraction asymmetries arise in different languages while allowing controlled use of permutative operations.

Before presenting the modalized CCG system, I will motivate the multi-modal approach with a brief look at coordination, demonstrating that not all categories should have full access to the combinatory rules.

5.2.1 Restricting Associativity

CCG's composition and type-raising rules provide the kind of associativity which is needed to handle many constructions. However, while this associativity is crucial for forming the constituents which are to be coordinated, it causes problems when we consider the categories of coordinating words themselves. For example, we might consider giving *and* the following category:

(184) *and* $\vdash (\mathsf{s}_i\backslash\mathsf{s}_i)/\mathsf{s}_i$

Unfortunately, such a category can lead to overgeneration because it is able to consume its first argument and then compose, producing noun phrases such as **player that shoots and he misses.*

(185)
$$
\begin{array}{cccccc}
\textit{*player} & \textit{that} & \textit{shoots} & \textit{and} & \textit{he} & \textit{misses} \\
 & \overline{(\mathsf{n}\backslash\mathsf{n})/(\mathsf{s}_i\backslash\mathsf{np})} & \overline{\mathsf{s}_i\backslash\mathsf{np}} & \overline{(\mathsf{s}_i\backslash\mathsf{s}_i)/\mathsf{s}_i} & \overline{\mathsf{np}} & \overline{\mathsf{s}_i\backslash\mathsf{np}} \\
 & & & & \multicolumn{2}{c}{\overline{\qquad\qquad}{<}} \\
 & & & & \multicolumn{2}{c}{\mathsf{s}_i} \\
 & & & \multicolumn{3}{c}{\overline{\qquad\qquad\qquad\qquad}{>}} \\
 & & & \multicolumn{3}{c}{\mathsf{s}_i\backslash\mathsf{s}_i} \\
 & & \multicolumn{4}{c}{\overline{\qquad\qquad\qquad\qquad\qquad}{<}\mathbf{B}} \\
 & & \multicolumn{4}{c}{\mathsf{s}_i\backslash\mathsf{np}} \\
 & \multicolumn{5}{c}{\overline{\qquad\qquad\qquad\qquad\qquad\qquad}{>}} \\
 & \multicolumn{5}{c}{\mathsf{n}\backslash\mathsf{n}}
\end{array}
$$

This problem has led to the use of the ternary syncategorematic coordination rule (186), which forces the consumption of both arguments of the coordinator simulataneously.

(186) $\mathsf{X}\ \ \mathsf{CONJ}\ \ \mathsf{X} \Rightarrow_{\Phi}\ \mathsf{X}$ $\qquad (Coordination(\Phi))$

However, this rule is unsatisfactory for two main reasons. First, it does not correspond to one of the combinators **B**, **T**, and **S**, and is therefore a rule created outside the scope of the Principle of Combinatory Type Transparency (31). More importantly, it does not permit fine-grained control over the properties of the various coordinating words. For example, *and* and *but* have different semantic properties, not only between each other, but also between their different uses in sentential and nominal coordinations. Furthermore, lexical control is particularly important in languages which actually have different coordinators for sentential and phrasal coordination. An example of such a language is Malagasy, which uses *ary* to coordinate sentences (187) and *sy* to coordinate phrases (188) (from (Keenan, 1978, pp. 319-320)).

(187) Misotro taoka Rabe ary mihinam bary Rabe.
drink alcohol Rabe and eat rice Rabe
Rabe is drinking alcohol and Rabe is eating rice.

(188) Misotro taoka sy mihinam bary Rabe.
drink alcohol and eat rice Rabe
Rabe is drinking alcohol and eating rice.

It thus appears that coordination should be handled in the lexicon, and the multi-modal approach gives us the tools to support this strategy without leading to overgenerations such as (185). Assuming a non-associative mode of grammatical composition is quite standard in work in CTL — in fact, a CTL system which utilizes just one mode of composition that is non-associative is the well-known Non-associative Lambek calculus. In the present setting, I will implement this mode with the $\star$ modality, giving it access to only the forward and backward application rules of CCG, but crucially not to the composition rules. We can then use this mode to create refined categories for coordinating words that ensure that they cannot take part in composition steps. For example, the category for *and* can now be declared as (189) without the danger of overgenerations like (185).

(189) *and* $\vdash (\mathsf{s}_i \backslash_{\star} \mathsf{s}_i)/_{\star}\mathsf{s}_i$

In effect, the ternary coordination (186) amounts to covert use of $\backslash_{\star}$ and $/_{\star}$ confined to the meta-category CONJ. Note that for coordination, it would be impossible to use restrictions on the composition rules to block overgenerations like (185) since the restrictions we would need to state would block most of the other desired uses of the composition rules. Though this is not the only reason why the resource-sensitivity of CCG should be enhanced, it already provides a strong indication the need for it.

The next sections show how CCG can be reformulated to incorporate lexically specified derivational control via categories that license different modes of grammatical composition.

5.2.2 Modes and the Application Rules

The definition of CTL given in §2.4 declares a general form which can be used to create specific multi-modal systems that use a *particular* finite set of modalities and a *particular* finite set of structural rules that interact with those modalities. Because of this, CTL-based systems can vary greatly with respect to the power of the operations they use and the inventory of modalities which those operations have access to. In bringing a multi-modal perspective to CCG, it is thus necessary to choose the set of modalities and define the rules to respect the different modes of grammatical composition we wish to utilize. In this section, I propose one way of how this may be done.

In the preceding section, I showed that a non-associative mode is necessary for providing a lexical account of coordination. The modality that I used to enforce non-associativity is $\star$. This symbol was chosen simply because it has not, to my knowledge, been used for other purposes in the CTL literature. The full set of modalities which I will employ is the following:

(190) $\mathcal{M}_{MM-CCG} = \{\star, \diamond, \triangleleft\!\times, \times\!\triangleright, \triangleleft, \triangleright, \cdot\}$

The effect of each of these modalities will be explicated as I introduce each of the combinatory rules and define their interaction with the modalities. Some of the modes incorporate a directionality dimension, represented as $\triangleleft$ and $\triangleright$, as discussed in §5.1. Disregarding the directionality, the basic intent behind the modalities is as follows: the $\star$ modality already discussed is the least permissive and permits access only to the application rules; $\diamond$ allows access to the harmonic composition rules; $\times$ licenses the crossed composition rules; and finally, $\cdot$ licenses all of the rules.

To use modally refined slashes in CCG, the definition of categories is the same as that defined in Definition 3 on page 38, except that we assume that atomic categories are feature structures as shown in Figure 3.1 on page 59. I will begin by defining a

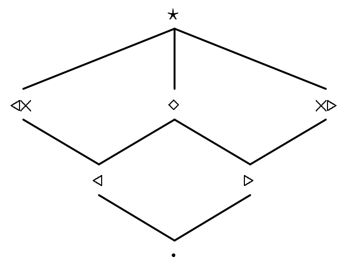


Figure 5.1: Hierarchy of modes.

small English lexicon to demonstrate the effects of each of the rules as we proceed through the modalization of the rules. The choice of the particular modalities on the slashes will become clear as more phenomena are discussed.

(191) a. *Rivaldo* $\vdash \mathsf{np}$

b. *kicked* $\vdash (\mathsf{s}_i\backslash_{\triangleleft}\mathsf{np})/_{\triangleright}\mathsf{np}$

c. *the* $\vdash \mathsf{np}/_{\diamond}\mathsf{n}$

d. *ball* $\vdash \mathsf{n}$

e. *skillfully* $\vdash (\mathsf{s}\backslash_{\triangleleft}\mathsf{np})/_{\triangleright}(\mathsf{s}\backslash_{\triangleleft}\mathsf{np})$

f. *that* $\vdash (\mathsf{n}\backslash_{\star}\mathsf{n})/_{\star}(\mathsf{s}_i/_{\cdot}\mathsf{np})$

The atomic categories and feature abbreviations used here are defined on page 59 in §3.1.

Instead of viewing the set of modalities as an unstructured collection of modalities, we can consider it to be an inheritance hierarchy, in which a modality has all the powers of its supertypes. I will assume that the hierarchy given in Figure 5.1 is operative for these purposes. The intuition behind a hierarchy such as this is that category with a particular modality may serve as input not only to all of the structural rules which reference the modality explicitly, but also rules which reference its supertypes.

We must next redefine the CCG rules to be sensitive to the modalities. It should be kept in mind that in every case the CCG deductive rules will be backed by CTL structural rules, and I will make some of these explicit as the discussion progresses. I start with the application rules, the simplest rule class. They are the equivalent of the slash elimination rules of CTL's base logic, and thus are defined to work with all modes of grammatical composition. Given the hierarchy in Figure 5.1, we can define the application rules so that they are keyed to the root modality $\star$:[2]

[2] Without the hierarchy, the definition would look more similar to that of the CTL rules:

(192) MULTI-MODAL CCG *application rules*

a. $\mathsf{X}/_{\star}\mathsf{Y} \quad \mathsf{Y} \;\Rightarrow\; \mathsf{X}$ $\qquad$ (>)

b. $\mathsf{Y} \quad \mathsf{X}\backslash_{\star}\mathsf{Y} \;\Rightarrow\; \mathsf{X}$ $\qquad$ (<)

Because the application rules are defined using the modality that is the root of the hierarchy, slashes which are decorated with any of the modalities can serve as input to them. Thus, we can now use the mini-lexicon given above to derive the sentence *Rivaldo skillfully kicked the ball.*

(193)

$$\begin{array}{ccccc}
\textit{Rivaldo} & \textit{skillfully} & \textit{kicked} & \textit{the} & \textit{ball} \\
\hline
\mathsf{np} & (\mathsf{s}\backslash_{\triangleleft}\mathsf{np})/_{\triangleright}(\mathsf{s}\backslash_{\triangleleft}\mathsf{np}) & (\mathsf{s}_i\backslash_{\triangleleft}\mathsf{np})/_{\triangleright}\mathsf{np} & \mathsf{np}/_{\diamond}\mathsf{n} & \mathsf{n}
\end{array}$$

$$\mathsf{np}/_{\diamond}\mathsf{n} \quad \mathsf{n} \;\Rightarrow_{>}\; \mathsf{np}$$

$$(\mathsf{s}_i\backslash_{\triangleleft}\mathsf{np})/_{\triangleright}\mathsf{np} \quad \mathsf{np} \;\Rightarrow_{>}\; \mathsf{s}_i\backslash_{\triangleleft}\mathsf{np}$$

$$(\mathsf{s}\backslash_{\triangleleft}\mathsf{np})/_{\triangleright}(\mathsf{s}\backslash_{\triangleleft}\mathsf{np}) \quad \mathsf{s}_i\backslash_{\triangleleft}\mathsf{np} \;\Rightarrow_{>}\; \mathsf{s}_i\backslash_{\triangleleft}\mathsf{np}$$

$$\mathsf{np} \quad \mathsf{s}_i\backslash_{\triangleleft}\mathsf{np} \;\Rightarrow_{<}\; \mathsf{s}_i$$

In order to cut down to some extent on modality clutter, I will assume the following convention for the remainder of this dissertation:

Convention 9 (Modality direction suppression). *The portion of a modality which specifies directionality will be suppressed on slashes with the same direction. Thus,* $\backslash_{\triangleleft}$, $/_{\triangleright}$, $\backslash_{\triangleleft\times}$, *and* $/_{\times\triangleright}$ *will be written as* $\backslash$, $/$, $\backslash_{\times}$, *and* $/_{\times}$, *respectively.*

The slashes $\backslash_{\triangleright}$, $/_{\triangleleft}$, $\backslash_{\times\triangleright}$, and $/_{\triangleleft\times}$ will continue to be written without abbreviation. Note that this is purely a matter of presentation — the slashes $\backslash_{\times}$ and $/_{\triangleleft\times}$ both carry the same modality $\triangleleft\times$, even though the $\triangleleft$ is suppressed on the former.

I now turn to how the inventory of modalities suggested in this section interact with the type-raising and composition rules.

5.2.3 Type-raising and Harmonic Composition

To derive a noun phrase such as *the ball that Henry kicked*, the rules of forward harmonic composition and forward type-raising are needed. The mode of grammatical composition for harmonic composition rules is associative, but not permutative,

---

a. $\mathsf{X}/_{i}\mathsf{Y} \quad \mathsf{Y} \;\Rightarrow\; \mathsf{X}$ $\qquad$ (>)

b. $\mathsf{Y} \quad \mathsf{X}\backslash_{i}\mathsf{Y} \;\Rightarrow\; \mathsf{X}$ $\qquad$ (<)

where $i \in \mathcal{M}_{MM-CCG}$

and I use the modality $\diamond$ to designate this. Thus, the modalized version of the composition rules are the following:

(194) MULTI-MODAL CCG *harmonic composition rules*

a. $\mathsf{X}/_{\diamond}\mathsf{Y} \quad \mathsf{Y}/_{\diamond}\mathsf{Z} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X}/_{\diamond}\mathsf{Z}$ (>**B**)

b. $\mathsf{Y}\backslash_{\diamond}\mathsf{Z} \quad \mathsf{X}\backslash_{\diamond}\mathsf{Y} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X}\backslash_{\diamond}\mathsf{Z}$ (<**B**)

Taking our cue from the CTL simulation of CCG discussed in §5.1, these formulations of composition stem from structural rules of associativity. For example, rule (194a) is essentially the reflection of the following rule:

(195) *Right Association*

$$\frac{(\Delta_a \ \circ_{\diamond} (\Delta_b \ \circ_{\diamond} \Delta_c)) \vdash \mathsf{X}}{((\Delta_a \ \circ_{\diamond} \Delta_b) \ \circ_{\diamond} \Delta_c) \vdash \mathsf{X}} [RA]$$

Clearly, this rule would allow a proof isomorphic to that of (176) if all instances of the modality 1 in that proof are replaced with $\diamond$.

Type-raising is a theorem of the base logic, so it does not require any further mention in the CTL system, but we need to declare it explicitly for CCG.

(196) MULTI-MODAL CCG *type-raising*

a. $\mathsf{X} \quad \Rightarrow_{\mathbf{T}} \quad \mathsf{Y}/_{i}(\mathsf{Y}\backslash_{i}\mathsf{X})$ (>**T**)

b. $\mathsf{X} \quad \Rightarrow_{\mathbf{T}} \quad \mathsf{Y}\backslash_{i}(\mathsf{Y}/_{i}\mathsf{X})$ (<**T**)

I use a modality variable to reflect the result of the CTL proof of type-raising as shown in (183), in which we see that the modality on the introduced slash is the same as that on the eliminated slash. However, when using type-raising in a derivation, I will use the contextually appropriate modality for clarity of exposition. Also, when type-raising is not the focus of a derivation, I will continue to employ the abbreviation of type-raised categories given in Convention 4 on page 28.

When the categories license harmonic composition, the system may exhibit the resulting associativity. For example, we can now give a derivation for object extractions such as *ball that Rivaldo kicked*.

(197)

$$
\begin{array}{cccccl}
\textit{the} & \textit{ball} & \textit{that} & \textit{Rivaldo} & \textit{kicked} & \\
\cline{1-1}\cline{2-2}\cline{3-3}\cline{4-4}\cline{5-5}
\mathsf{np}/_{\diamond}\mathsf{n} & \mathsf{n} & (\mathsf{n}\backslash_{\star}\mathsf{n})/_{\star}(\mathsf{s}_i/_{\cdot}\mathsf{np}) & \mathsf{np} & (\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np} & \\
\cline{4-4}
 & & & \mathsf{s}_i/_{\triangleleft}(\mathsf{s}_i\backslash\mathsf{np}) & & >\mathbf{T} \\
\cline{4-5}
 & & & \multicolumn{2}{c}{\mathsf{s}_i/\mathsf{np}} & >\mathbf{B} \\
\cline{3-5}
 & & \multicolumn{3}{c}{\mathsf{n}\backslash_{\star}\mathsf{n}} & > \\
\cline{2-5}
 & \multicolumn{4}{c}{\mathsf{n}} & < \\
\cline{1-5}
\multicolumn{5}{c}{\mathsf{np}} & >
\end{array}
$$

With this derivation, the way in which the hierarchy of modalities works can be seen more clearly. For example, since $\triangleleft$ and $\triangleright$ are supertypes of $\diamond$, the application of the forward harmonic composition rule to the category inputs $\mathsf{s}_i/_{\triangleleft}(\mathsf{s}_i\backslash_{\triangleleft}\mathsf{np})$ and $(\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np}$ succeeds. Had either of the slashes carried the $\star$ modality, $\triangleleft\!\times$ modality, or $\times\!\triangleright$ modality, the rule could not have applied. This is exactly why the use of the category $(\mathsf{s}_i\backslash_{\star}\mathsf{s}_i)/_{\star}\mathsf{s}_i$ for sentential coordination stops overgenerations such as (185), as discussed in §5.2.1.

Enhanced resource-sensitivity does not cause a loss in the *desired* derivational potential of the original CCG system — the categories of (191) and the rules defined so far permit the same mostly incremental derivation of *Rivaldo skillfully kicked the ball* as standard CCG.

5.2.4 Control Over Permutativity

The multi-modal setting brings further benefits. With standard CCG, we may choose to include or omit any of the combinatory rules for a given grammar. For example, Steedman (1996, 2000b) bans the forward crossed composition rule ($>\mathbf{B}_{\times}$) from the grammar of English, for reasons which are explained in §4.1.1. However, one of the goals of the present enterprise is to eliminate the use of rule restrictions. The desired effect of such restrictions must instead arise due to the modalities found on the categories of the lexicon.

To carry this proposal forward, we need to provide modalized versions of the crossed composition rules. Although it has been argued that the grammar for English cannot include forward crossed composition, its twin backward crossed composition is needed for deriving constructions such as heavy shift, as discussed in §2.2.3 and shown in derivation (52). However, backward crossed composition cannot apply universally in English since it would cause the grammar to accept strings such as **a powerful by Rivaldo shot* with the meaning *a powerful shot by Rivaldo*.

(198) *a [powerful]$_{\mathsf{n/n}}$ [by Rivaldo]$_{\mathsf{n\backslash n}}$ [shot]$_{\mathsf{n}}$

Steedman (1996) blocks such sequences by permitting backward crossed composition to apply only when both the primary and secondary categories are rooted in s, as shown in (199).

(199) *Backward crossed composition* ($<\mathbf{B}_{\times}$)

$$\mathsf{Y/Z} \quad \mathsf{X\backslash Y} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X/Z}$$

where $\mathsf{X} = \mathsf{Y} = \mathsf{s}\backslash\$$

See Convention 2 on page 26 for the definition of the \$ used in the restriction of the rule.

Rather than allowing certain CCG rules to be active or inactive in a given grammar, we can pass the control of such reflexes into the lexicon, just as we did for application-only constructions such as coordination. It is here that we employ the modalities $\triangleleft\!\times$ and $\times\!\triangleright$ in defining the crossed composition rules.

(200) MULTI-MODAL CCG *crossed composition rules*

a. $\mathsf{X}/_{\times}\mathsf{Y} \quad \mathsf{Y}\backslash_{\times}\mathsf{Z} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X}\backslash_{\times}\mathsf{Z}$ ($>\mathbf{B}_{\times}$)

b. $\mathsf{Y}/_{\times}\mathsf{Z} \quad \mathsf{X}\backslash_{\times}\mathsf{Y} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X}/_{\times}\mathsf{Z}$ ($<\mathbf{B}_{\times}$)

An immediate payoff is that we can now use rules like backward crossed composition in a grammar like English without having to place restrictions to limit its applicability. Instead, categories are declared in the lexicon to have slash types that are either compatible or incompatible with crossed composition. This provides a clean solution to the problem noted in §4.3, in which post-verbal adverbs can come between a verb and its direct object (173), but a post-nominal modifier cannot appear between an adjective and its nominal argument (174). For example, the category of *skillfully* is that in (201), while that of *by* is the one in (202).

(201) *skillfully* $\vdash (\mathsf{s\backslash np})\backslash(\mathsf{s\backslash np})$

(202) *by* $\vdash (\mathsf{n}\backslash_{\star}\mathsf{n})/_{\triangleleft}\mathsf{np}$

The category defined for English transitive verbs, $(\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np}$, can serve as input to rule (200b), thus allowing a derivation equivalent to (52) in MULTI-MODAL CCG, as shown in (203). However, (204) shows that it is impossible to derive (198) because of the slash type on the category given for *by*.

(203) Kahn [blocked]$_{(\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np}}$ [skillfully]$_{(\mathsf{s\backslash np})\backslash(\mathsf{s\backslash np})}$ a powerful shot by Rivaldo.

(204) *a [powerful]$_{\mathsf{n}/_{\diamond}\mathsf{n}}$ [by Rivaldo]$_{\mathsf{n}\backslash_{\star}\mathsf{n}}$ [shot]$_{\mathsf{n}}$

Of course, the harmonic slash specified for the adjective *powerful* in (204) also blocks crossed rules from applying, resulting in a double block on the use of backward crossed composition (and forward crossed composition as well, actually).

Note that the analysis of heavy shift suggested here will permit adverbs to come between the verb and non-heavy direct objects, which is often claimed to be ungrammatical for English. Not only do I find it hard to label them as ungrammatical, it is actually not uncommon to hear such utterances from native speakers of American English. Here is one real-life example:

(205) We've laid out very carefully the targets.
– *Former United States Defense Secretary William Cohen, quoted in the New York Times, 17th of December 1998*

Nonetheless, if we did desire an analysis which only permitted shifting when the direct object is heavy, the mechanisms developed in the next section provide the means to enforce such behavior.

Now that the resource-sensitive system has been introduced for the primary rule classes, I return to more evidence to that given in §5.2.1 that we need control not just over permutativity, but over associativity as well.

5.2.5 More Non-associative Contexts

The application-only modality $\star$ provides a clean solution to a problem noted by Milward (1994) for lexicons which do not rigidly distinguish between bare nouns and noun phrases. There is considerable convenience for CCG lexicons made for computational tasks in allowing underspecification of this distinction (Bierner, 2001), but it can give rise to some undesirable word order consequences. Under this non-standard assumption, relativizers are functions over noun phrases (instead of bare nouns), and a grammar which handles (206) will also produce the disallowed order (207) with the same reading.

(206) The players that came from Spain angrily left.

(207) *The players angrily that came from Spain left.

This is derivable because the verb can combine via backward composition with the relative clause before the latter combines with the subject:

(208)
$$\begin{array}{cccc}
\textit{*The players} & \textit{angrily} & \textit{that came from Spain} & \textit{left}\\
\mathsf{np} & (\mathsf{s}\backslash\mathsf{np})/(\mathsf{s}\backslash\mathsf{np}) & \mathsf{np}\backslash\mathsf{np} & \mathsf{s}_i\backslash\mathsf{np}\\
 & & \multicolumn{2}{c}{\mathsf{s}_i\backslash\mathsf{np}\quad {<}\mathbf{B}}\\
 & \multicolumn{2}{c}{\mathsf{s}_i\backslash\mathsf{np}\quad {>}} & \\
\multicolumn{3}{c}{\mathsf{s}_i\quad {<}} &
\end{array}$$

Given that *that* has the category $(\mathsf{np}\backslash_{\star}\mathsf{np})/_{\star}(\mathsf{s}_i\backslash\mathsf{np})$, this derivation is blocked in MULTI-MODAL CCG because the category $\mathsf{np}\backslash_{\star}\mathsf{np}$ of the constituent *that came from Spain* cannot serve as input to the rule of harmonic backward composition (194b).

(209) *The players angrily [that came from Spain]$_{\mathsf{np}\backslash_{\star}\mathsf{np}}$ [left]$_{\mathsf{s}_i\backslash\mathsf{np}}$

Of course, the derivation for (206) is still permitted since the relative clause can combine with *The players* via backward application.

It should be noted that Milward's example is not actually a problem for non-modalized CCG if relativizers only take bare nouns as arguments instead of full noun phrases. This strategy would of course require lexical ambiguity or a unary rule that turned the category of bare nouns into noun phrases for combination with the verb. Nonetheless, MULTI-MODAL CCG would support an analysis such as that given in Bierner (2001) without the above-mentioned overgeneration.

The category $(\mathsf{np}\backslash_{\star}\mathsf{np})/_{\star}(\mathsf{s}\backslash\mathsf{np})$ given above for relativizers also blocks illicit double relativizations such as the following:

(210) *goals$_j$ that I saw the players$_i$ that t$_i$ scored t$_j$

(211) *player$_i$ that I saw the card$_j$ that the referee gave t$_i$ t$_j$

Such relative clauses could otherwise be derived if nouns can type-raise over nominally rooted functions such as relativizers:

(212)
$$\begin{array}{cccccc}
\textit{*goals} & \textit{that} & \textit{I saw} & \textit{the players} & \textit{that} & \textit{scored}\\
 & (\mathsf{np}\backslash\mathsf{np})/(\mathsf{s}/\mathsf{np}) & \mathsf{s}_i/\mathsf{np} & \mathsf{np} & (\mathsf{np}\backslash\mathsf{np})/(\mathsf{s}\backslash\mathsf{np}) & (\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np}\\
 & & & \mathsf{np}/(\mathsf{np}\backslash\mathsf{np})\quad {>}\mathbf{T} & \multicolumn{2}{c}{(\mathsf{np}\backslash\mathsf{np})/\mathsf{np}\quad {>}\mathbf{B}}\\
 & & & \multicolumn{2}{c}{\mathsf{np}/\mathsf{np}\quad {>}\mathbf{B}} & \\
 & & \multicolumn{2}{c}{\mathsf{s}_i/\mathsf{np}\quad {>}\mathbf{B}} & & \\
 & \multicolumn{2}{c}{\mathsf{np}\backslash\mathsf{np}\quad {>}} & & &
\end{array}$$

With modalized rules, the slashes decorated by the $\star$ modality on the relativizer ensure that not even the first step of the above derivation can go through:

(213) $^*[\text{that}]_{(\mathsf{np}\backslash_\star\mathsf{np})/_\star(\mathsf{s}_i\backslash\mathsf{np})}$ $[\text{scored}]_{(\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np}}$

Note that the categories of relativizers are quite similar to those of coordinating words, which also utilize the non-associative mode.

By making similar assumptions about the modes specified by the categories of prepositional phrases, the following sort of overgeneration is also blocked because the verbs cannot backward compose into the unsaturated prepositional phrases:

(214) *The fan [in the field]$_{\mathsf{np}\backslash_\star\mathsf{np}}$ [left]$_{\mathsf{s}_i\backslash\mathsf{np}}$ and [in the stadium]$_{\mathsf{np}\backslash_\star\mathsf{np}}$ [stayed]$_{\mathsf{s}_i\backslash\mathsf{np}}$.

The non-associative slash is of course used in CTL approaches as well to restrict certain categories to the base logic. Also, the use of the non-associative slash is very much like the null adjunction label used in Tree-Adjoining Grammar (TAG) (Joshi, 1988; Abeillé and Rambow, 2000) to lexically block certain adjunctions and permit the rule of adjunction (the TAG correlate of composition) to apply universally.

5.2.6 Extraction of Direct and Indirect Objects from English Ditransitive Verbs

The proposal sketched above shows that we can get rid of restrictions on CCG rules like backward crossed composition and instead specify such behavior in the lexicon. However, in addition to being used for heavy-shift, backward crossed composition is also crucial for extraction of indirect objects of English ditransitive verbs.

(215) player$_i$ that the referee gave t_i a yellow card

Under the standard assumption that the category of the English ditransitive is $((\mathsf{s}_i\backslash\mathsf{np}^{\mathbf{Act}})/\mathsf{np}^{\mathbf{Pat}})/\mathsf{np}^{\mathbf{Addr}}$, the derivation proceeds as follows:[3]

(216)

$$
\begin{array}{llll}
\textit{(player that)} & \textit{the referee} & \textit{gave} & \textit{a yellow card} \\
\hline
 & \mathsf{np}^{\mathbf{Act}} & ((\mathsf{s}_i\backslash\mathsf{np}^{\mathbf{Act}})/\mathsf{np}^{\mathbf{Pat}})/\mathsf{np}^{\mathbf{Addr}} & \mathsf{np}^{\mathbf{Pat}} \\
 & \overline{\ddot{\mathsf{s}}_i/(\ddot{\mathsf{s}}_i\backslash\mathsf{np}^{\mathbf{Act}})}\ {>}\mathbf{T} & & \overline{\ddot{\mathsf{s}}_i\backslash(\ddot{\mathsf{s}}_i/\mathsf{np}^{\mathbf{Pat}})}\ {<}\mathbf{T} \\
 & \multicolumn{2}{c}{\overline{(\ddot{\mathsf{s}}_i/\mathsf{np}^{\mathbf{Pat}})/\mathsf{np}^{\mathbf{Addr}}}\ {>}\mathbf{B}^2} & \\
 & \multicolumn{3}{c}{\overline{\ddot{\mathsf{s}}_i/\mathsf{np}^{\mathbf{Addr}}}\ {<}\mathbf{B}_\times}
\end{array}
$$

[3]The superscripts **Act**, **Pat**, and **Addr** (for the dependency roles **Actor**, **Patient**, and **Addressee**, respectively) are included only to help distinguish the arguments for the reader and do not perform any grammatical function. See §2.6, page 49 for more details.

However, as it stands, this analysis also predicts that a sentence such as (217) is grammatical under the reading that the player receives the card and not vice versa. Even when the **Addressee** argument is heavy, such a sentence is clearly unacceptable.

(217) *[The referee gave a yellow card]$_{\mathsf{s}_i/\mathsf{np}^{\mathbf{Addr}}}$ [the aggressive player.]$_{\mathsf{np}^{\mathbf{Addr}}}$

To block such cases, Steedman (2000b) replaces the backward crossed composition rule (199) with the two featurally distinct instantiations (218) and (219).

(218) *Backward crossed composition I* ($<\mathbf{B}_\times$)
$\mathsf{Y/Z}_{+SHIFT} \quad \mathsf{X\backslash Y} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X/Z}_{+SHIFT}$
where $\mathsf{X} = \mathsf{Y} = \mathsf{s}\backslash\$$

(219) *Backward crossed composition II* ($<\mathbf{B}_\times$)
$\mathsf{Y/Z}_{-SHIFT,+ANT} \quad \mathsf{X\backslash Y} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X/Z}_{-SHIFT,+ANT}$
where $\mathsf{X} = \mathsf{Y} = \mathsf{s}\backslash\$$

Notice that the restrictions on X and Y in the input categories are the same in both rules.

The rule (218) permits shifting of an argument over other elements such as adverbs, thus allowing derivations such as (52). The rule (219) allows the creation of constituents which are marked for *antecedent government* via the feature $\pm$ANT, which was discussed in §4.1.1, page 74, with respect to exception subject extractions. The assumption is that all lexical noun phrases have the value $-$ANT for this feature and that the targets of extracting categories bear the unspecified value $\pm$ANT. By marking an argument of a category as $-$SHIFT, only the version of backward crossed composition given in (219) will be able to apply, with the result that the argument must thenceforth be antecedent governed. Assuming the categories in (220), Steedman (2000b) captures that fact that the indirect object can be leftward-extracted, as in (221), but not rightward-extracted, as in (222).

(220) a. *that* $\vdash (\mathsf{n\backslash n})/(\mathsf{s/np}_{\pm ANT})$

b. *gave* $\vdash ((\mathsf{s}_i\backslash\mathsf{np}^{\mathbf{Act}})/\mathsf{np}^{\mathbf{Pat}})/\mathsf{np}^{\mathbf{Addr}}_{-SHIFT}$

(221)
$$
\begin{array}{llll}
\textit{player} & \textit{that} & \textit{the referee gave} & \textit{a yellow card} \\
\cline{2-4}
 & (\mathsf{n}\backslash\mathsf{n})/(\mathsf{s}/\mathsf{np}_{\pm ANT}) & (\ddot{\mathsf{s}}_i/\mathsf{np}^{\mathbf{Pat}})/\mathsf{np}^{\mathbf{Addr}}_{-SHIFT} & \mathsf{np}^{\mathbf{Pat}} \\
\cline{4-4}
 & & & \ddot{\mathsf{s}}_i\backslash(\ddot{\mathsf{s}}_i/\mathsf{np}^{\mathbf{Pat}}) \quad {<}\mathbf{T} \\
\cline{3-4}
 & & \multicolumn{2}{c}{\ddot{\mathsf{s}}_i/\mathsf{np}^{\mathbf{Addr}}_{-SHIFT,+ANT} \quad {<}\mathbf{B}_{\times}} \\
\cline{2-4}
 & \multicolumn{3}{c}{\mathsf{n}\backslash\mathsf{n} \quad {>}}
\end{array}
$$

(222) *[The referee gave a yellow card]$_{\ddot{\mathsf{s}}_i/\mathsf{np}^{\mathbf{Addr}}_{-SHIFT,+ANT}}$ [a player.]$_{\mathsf{np}^{\mathbf{Addr}}_{-ANT}}$

To cover this data in the multi-modal setting also requires the use of multiple instantiations for backward crossed composition. One option would be to adopt essentially the same analysis using the features $\pm$SHIFT and $\pm$ANT. The improvement over Steedman's analysis would then be that these multiple instantiations would not declare (the same) restrictions on the rules' input categories. However, there is a degree of freedom in the modalities which I have defined which has not been thus far exploited and which can in fact cover the extraction facts. Recall that I have assumed that slashes from the lexicon maintain the same directionality on both the slash and the modality, e.g. $\backslash_{\triangleleft}$ and $/_{\triangleright}$ (which have been represented as just $\backslash$ and $/$ due to Convention 9). The appearance of slashes such as $\backslash_{\triangleright}$ and $/_{\triangleleft}$ has only arisen through type-raising. I will now explore the consequences of assuming that the latter types of slashes can be used directly in the lexicon.

Given that the modalities $\triangleleft\!\times$ and $\times\!\triangleright$ can appear on both leftward and rightward slashes, there are actually four possible instantiations of a rule like backward crossed composition. Let us assume that all rules in which the directionality of one of the slashes is opposite that of its modality induce a result which must be antecedent governed. The spectrum of backward crossed composition rules then looks as follows:

(223)
a. $\mathsf{Y}/_{\times}\mathsf{Z} \quad \mathsf{X}\backslash_{\times}\mathsf{Y} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X}/_{\times}\mathsf{Z}$ $\quad ({<}\mathbf{B}_{\times})$

b. $\mathsf{Y}/_{\times}\mathsf{Z}_{+ANT} \quad \mathsf{X}\backslash_{\times\triangleright}\mathsf{Y} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X}/_{\times}\mathsf{Z}_{+ANT}$ $\quad ({<}\mathbf{B}^{ant}_{\times})$

c. $\mathsf{Y}/_{\triangleleft\times}\mathsf{Z}_{+ANT} \quad \mathsf{X}\backslash_{\times}\mathsf{Y} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X}/_{\triangleleft\times}\mathsf{Z}_{+ANT}$ $\quad ({<}\mathbf{B}^{ant}_{\times})$

d. $\mathsf{Y}/_{\triangleleft\times}\mathsf{Z}_{+ANT} \quad \mathsf{X}\backslash_{\times\triangleright}\mathsf{Y} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X}/_{\triangleleft\times}\mathsf{Z}_{+ANT}$ $\quad ({<}\mathbf{B}^{ant}_{\times})$

To simulate the last three rules in CTL, it is necessary to bring unary modalities into the simulation. They play an important part in CTL's handling of not only particular aspects of structural control, but also feature percolation. Unary modes are often used to implement a "lock and key" strategy which ensures that a structure conforms to a given criteria — see Oehrle (to appear) for an example which uses this technique to capture Dutch crossed dependencies. I suggest here that we can use a similar strategy that reveals how Steedman's +ANT feature is very much like a unary lock on an extracted argument.

Structural rules in CTL can involve interaction of unary and binary modes, as discussed in §2.4. We can thus provide the following rule, Antecedent-Governed Right Permutation ($AGRP{-}I$), as the CTL basis for (223b).

(224) $$\frac{((\Delta_a \circ_{\times\triangleright} \langle\Delta_b\rangle^{ant}) \circ_{\times\triangleright} \Delta_c) \vdash \mathsf{X}}{((\Delta_a \circ_{\times\triangleright} \Delta_c) \circ_{\times\triangleright} \langle\Delta_b\rangle^{ant}) \vdash \mathsf{X}}\,[AGRP{-}I]$$

We also need two additional similar structural rules to back up the rules (223c,d). The effect of (224) is demonstrated below.

(225) $$\dfrac{[x_2 \vdash \Diamond_{ant}\Box^{\downarrow}_{ant}\mathsf{Z}]^{\ddagger} \qquad \dfrac{\dfrac{\dfrac{\Delta \vdash \mathsf{Y}/_{\times}\mathsf{Z} \qquad \dfrac{[x_1 \vdash \Box^{\downarrow}_{ant}\mathsf{Z}]^{\dagger}}{\langle x_1\rangle^{ant} \vdash \mathsf{Z}}\,[\Box^{\downarrow}_{ant}E]}{(\Delta \circ_{\times\triangleright} \langle x_1\rangle^{ant}) \vdash \mathsf{Y}}\,[/_{\times}E] \qquad \Gamma \vdash \mathsf{X}\backslash_{\times\triangleright}\mathsf{Y}}{((\Delta \circ_{\times\triangleright} \langle x_1\rangle^{ant}) \circ_{\times\triangleright} \Gamma) \vdash \mathsf{X}}\,[\backslash_{\times\triangleright}E]}{((\Delta \circ_{\times\triangleright} \Gamma) \circ_{\times\triangleright} \langle x_1\rangle^{ant}) \vdash \mathsf{X}}\,[AGRP{-}I]}{\dfrac{((\Delta \circ_{\times\triangleright} \Gamma) \circ_{\times\triangleright} x_2) \vdash \mathsf{X}}{(\Delta \circ_{\times\triangleright} \Gamma) \vdash \mathsf{X}/_{\times}\Diamond_{ant}\Box^{\downarrow}_{ant}\mathsf{Z}}\,[/_{\times}I]^{\ddagger}}\,[\Diamond_{ant}E]^{\dagger}$$

Thus, given these modalities, the rule can only ever apply to create a constituent which is marked for antecedent government. The use of the +ANT feature on Z in the rules (223b-d) parallels the structural effect of the unary mode $\langle\cdot\rangle^{ant}$.

Given the rules in (223), we are now in a position to provide an alternative analysis of the asymmetry discussed previously. We translate the relativizer category (220a) into the multi-modal setting (226a), and assign *gave* the category (226b).

(226) a. *that* $\vdash (\mathsf{n}\backslash_{\star}\mathsf{n})/_{\star}(\mathsf{s}/_{.}\mathsf{np}_{\pm ANT})$

b. *gave* $\vdash ((\mathsf{s}_i\backslash\mathsf{np}^{\mathbf{Act}})/\mathsf{np}^{\mathbf{Pat}})/_{\triangleleft}\mathsf{np}^{\mathbf{Addr}}$

(227) $$\dfrac{\dfrac{\textit{that}}{(\mathsf{n}\backslash_{\star}\mathsf{n})/_{\star}(\mathsf{s}/_{.}\mathsf{np}_{\pm ANT})} \qquad \dfrac{\dfrac{\textit{the referee gave}}{(\ddot{\mathsf{s}}_i/\mathsf{np}^{\mathbf{Pat}})/_{\triangleleft}\mathsf{np}^{\mathbf{Addr}}} \qquad \dfrac{\dfrac{\textit{a yellow card}}{\mathsf{np}^{\mathbf{Pat}}}}{\ddot{\mathsf{s}}_i/_{\triangleright}(\ddot{\mathsf{s}}_i\backslash_{\triangleright}\mathsf{np}^{\mathbf{Pat}})}{<}\mathbf{T}}{\ddot{\mathsf{s}}_i/_{\triangleleft}\mathsf{np}^{\mathbf{Addr}}_{+ANT}}{<}\mathbf{B}^{ant}_{\times}}{\mathsf{n}\backslash_{\star}\mathsf{n}}{>}$$

player

The category furthermore handles the difference in adverb placement evident in the following two sentences:

(228) a. The referee [gave the player]$_{(\ddot{s}_i\backslash np)/np}$ [today]$_{(s\backslash np)\backslash(s\backslash np)}$ a well-deserved red card.

b. *The referee [gave]$_{((s_i\backslash np)/np)/_{\triangleleft}np}$ [today]$_{(s\backslash np)\backslash(s\backslash np)}$ the aggressive player a well-deserved red card.

The use of the slash $/_{\triangleleft}$ on the external argument of the ditransitive category means that the category cannot be cross composed into until that argument is consumed, producing an effect quite similar to that achieved by Steedman's SHIFT feature.[4]

We have thus used a degree of freedom already available in the modal setting and dispensed with the ±SHIFT feature. This approach is applicable to another idiosyncractic extraction context.

(229) Which country did João travel to skeptically and return from joyfully?

(230) *João traveled to skeptically the country that hosted the cup final.

(231) João traveled to, and Maria returned from, the country that hosted the cup final.

Steedman (2000b) assigns the following category to *to*:

(232) $\textit{to} \vdash pp/np_{-SHIFT}$

(232) blocks the illicit sentence (230) via the same mechanism that blocks (222). On the other hand, (231) is permited by category (232) because backward crossed composition is not required and therefore does not induce a positive value for ANT. The category for a constituent like *traveled to* is $(s_i\backslash np)/np_{-SHIFT,\pm ANT}$, whereas the category for *travel to skeptically* is $(s_i\backslash np)/np_{-SHIFT,+ANT}$ because of the need to use the backward crossed composition rule (219). Nonetheless, this category is exactly what is needed to derive (229) since the extracting category of *which* is the following:

(233) $\textit{which} \vdash (\ddot{s}_c/(s_i/np_{\pm ANT}))/np$

---

[4]The definition of backward crossed composition given by Steedman (2000b) in example (31) on p. 63 must be generalized for his analysis to capture sentences such as *The referee gave today to the aggressive player a well-deserved red card.* In order to still block sentences such as (228b), the generalization must furthermore ensure that all categories in the rightward stack must have the value +SHIFT:

(a) $(Y/Z_{+SHIFT})/\$_{+SHIFT} \quad X\backslash Y \quad \Rightarrow_{\mathbf{B}} \quad X/Z_{+SHIFT}/\$_{+SHIFT}$ $(<\mathbf{B}_{\times})$

The modal analysis follows the same strategy as that for the ditransitive to allow leftward extraction while disallowing extraction to the right:

(234) *to* $\vdash \mathsf{pp}/_{\triangleleft}\mathsf{np}$

We have thus not lost any of the sensitivity of Steedman's analysis while at the same time making full use of the degrees of freedom permitted by the multi-modal setting. I now turn to the formulation of the forward crossed composition rule and a surprising result it brings to the analysis of subject extraction in English.

5.2.7 The Subject/Object Asymmetry Revisited

Another advantage of using lexically specified derivational control is that it allows us to bring forward crossed composition back into the grammar of English. §4.1.1 shows that it cannot apply globally since that would predict that phrases such as **team that John knew that would beat China* are grammatical. We can in fact enforce this behavior by once again using modalities on the appropriate categories. The modal formulation of forward crossed composition is the dual of backward crossed composition discussed earlier:

(235)
a. $\mathsf{X}/_{\times}\mathsf{Y} \quad \mathsf{Y}\backslash_{\times}\mathsf{Z} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X}\backslash_{\times}\mathsf{Z}$ $\qquad (>\mathbf{B}_{\times})$

b. $\mathsf{X}/_{\times}\mathsf{Y} \quad \mathsf{Y}\backslash_{\times\triangleright}\mathsf{Z}_{+ANT} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X}\backslash_{\times\triangleright}\mathsf{Z}_{+ANT}$ $\qquad (>\mathbf{B}_{\times}^{ant})$

c. $\mathsf{X}/_{\triangleleft\times}\mathsf{Y} \quad \mathsf{Y}\backslash_{\times}\mathsf{Z}_{+ANT} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X}\backslash_{\times}\mathsf{Z}_{+ANT}$ $\qquad (>\mathbf{B}_{\times}^{ant})$

d. $\mathsf{X}/_{\triangleleft\times}\mathsf{Y} \quad \mathsf{Y}\backslash_{\times\triangleright}\mathsf{Z}_{+ANT} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X}\backslash_{\times\triangleright}\mathsf{Z}_{+ANT}$ $\qquad (>\mathbf{B}_{\times}^{ant})$

Blocking the illicit subject-extraction example (125b) is simply a matter of using a harmonic modality on the category for the complemetizer *that*:

(236) *that* $\vdash \ddot{\mathsf{s}}_{c}/_{\diamond}\mathsf{s}_{i}$

(237) *team that John knew $[\text{that}]_{\ddot{\mathsf{s}}_{c}/_{\diamond}\mathsf{s}_{i}}$ $[\text{would defeat China}]_{\mathsf{s}_{i}\backslash\mathsf{np}}$

Not only do we regain the possibility of using forward crossed composition in the grammar for English, we can furthermore provide a single category (238) for sentential complement verbs like *knew* which allows embedded subject extraction when the complementizer is missing, as in (131).

(238) *knew* $\vdash (\mathsf{s}_{i}\backslash\mathsf{np})/_{\triangleleft}\mathsf{s}_{fin}$

The derivation for embedded subject extraction proceeds as follows:

(239)
$$
\begin{array}{llllll}
\textit{team} & \textit{that} & \textit{John} & \textit{knew} & \textit{would defeat China} \\
\hline
 & (\mathsf{n}\backslash_{\star}\mathsf{n})/_{\star}(\mathsf{s}_i\backslash\mathsf{np}_{\pm ANT}) & \mathsf{np}_{-ANT} & (\mathsf{s}_i\backslash\mathsf{np})/_{\triangleleft}\mathsf{s}_{fin} & \mathsf{s}_i\backslash\mathsf{np} \\
 & & \overline{}{>}\mathbf{T} & \multicolumn{2}{l}{\overline{}{>}\mathbf{B}^{ant}_{\times}} \\
 & & \mathsf{np}^{\uparrow}_{-ANT} & \multicolumn{2}{l}{(\mathsf{s}_i\backslash\mathsf{np})\backslash\mathsf{np}_{+ANT}} \\
 & & \multicolumn{3}{l}{\overline{}{>}\mathbf{B}^{ant}_{\times}} \\
 & & \multicolumn{3}{l}{\ddot{\mathsf{s}}_i\backslash\mathsf{np}_{+ANT}} \\
 & \multicolumn{4}{l}{\overline{}{>}} \\
 & \multicolumn{4}{l}{\mathsf{n}\backslash\mathsf{n}}
\end{array}
$$

The scrambled sentence *John Brazil knew would defeat China* is blocked because the outermost argument of the category derived for *knew would defeat China* is marked for antecedent government and therefore cannot consume a lexical noun phrase such as *Brazil*:

(240) *John [Brazil]$_{\mathsf{np}_{-ANT}}$ [knew would defeat China]$_{(\mathsf{s}_i\backslash\mathsf{np})\backslash\mathsf{np}_{+ANT}}$

Having brought forward crossed composition back into the English grammar merits pointing out that in some cases the same two categories can serve as input to both forward crossed composition and backward crossed composition. For example, in the standard CCG setting, the categories in (198) can in principle combine not only through backward crossed composition to produce the string *powerful by Rivaldo shot* as shown in (241), but also through forward crossed composition to produce *shot powerful by Rivaldo*, shown in (242). I have placed indexes on the atomic categories to make the results of combination clearer.

(241)
$$
\begin{array}{lll}
\textit{*powerful} & \textit{by Rivaldo} & \textit{shot} \\
\hline
\mathsf{n}_1/\mathsf{n}_2 & \mathsf{n}_3\backslash\mathsf{n}_4 & \mathsf{n} \\
\multicolumn{2}{l}{\overline{}{<}\mathbf{B}_{\times}} & \\
\multicolumn{2}{l}{\mathsf{n}_3/\mathsf{n}_2} & \\
\multicolumn{3}{l}{\overline{}{>}} \\
\multicolumn{3}{l}{\mathsf{n}_3}
\end{array}
$$

(242)
$$
\begin{array}{lll}
\textit{*shot} & \textit{powerful} & \textit{by Rivaldo} \\
\hline
\mathsf{n} & \mathsf{n}_1/\mathsf{n}_2 & \mathsf{n}_3\backslash\mathsf{n}_4 \\
 & \multicolumn{2}{l}{\overline{}{>}\mathbf{B}_{\times}} \\
 & \multicolumn{2}{l}{\mathsf{n}_1\backslash\mathsf{n}_4} \\
\multicolumn{3}{l}{\overline{}{<}} \\
\multicolumn{3}{l}{\mathsf{n}_1}
\end{array}
$$

This was not an issue in standard CCG since forward crossed composition was banned from English. However, even though forward crossed composition is assumed to be (always) active in MULTI-MODAL CCG, it does not engender illicit orders of this nature because the modally refined categories for *powerful* and *by* cannot serve as input to either of the crossed rule types:

(243) *powerful* $\vdash \mathsf{n}/_{\diamond}\mathsf{n}$

(244) *by* $\vdash$ $(\mathsf{n}\backslash_{\star}\mathsf{n})/_{\triangleleft}\mathsf{np}$

5.2.8 Extraction Out of Subjects

The categories we have assumed thus far incorrectly predict that extraction out of subjects is possible in English. Consider the following (adapted from Steedman (1996)):

(245) player$_i$ that I read a book about t_i

(246) *player$_i$ that a book about t_i astonished me

I will show in the following chapter how the inclusion of multiset categories in CCG can permit *a book* and *about* to combine via application. What is important here is that the category $\mathsf{np}/_{\triangleleft}\mathsf{np}$ can be derived for the string *a book about.* Clearly the object extraction can go through, but the subject extraction does as well:

(247) *player that [a book about]$_{\mathsf{np}/_{\times\triangleleft}\mathsf{np}}$ [astonished me]$_{\mathsf{s}\backslash\mathsf{np}}$

The category for *a book about* must allow limited backward crossed composition for leftward non-peripheral extractions, such as the following:

(248) player that I [read a book about]$_{(\mathsf{s}_i\backslash\mathsf{np})/_{\times\triangleleft}\mathsf{np}}$ [yesterday]$_{(\mathsf{s}\backslash\mathsf{np})\backslash(\mathsf{s}\backslash\mathsf{np})}$

We thus cannot block (246) using a less powerful mode for the preposition. Another option would be to place a more restricted modality on the slash for the subject:

(249) *astonished* $\vdash$ $\mathsf{s}_i\backslash_{\diamond}\mathsf{np}$

(250) *astonished* $\vdash$ $\mathsf{s}_i\backslash_{\triangleright}\mathsf{np}$

The first category (249) clearly blocks (246), but it will force me to abandon the analysis of the English subject/object asymmetry given earlier. For example, see (239), in which *knew* and *would defeat China* would no longer be able to combine through forward crossed composition. The second category (250) will not block (239), but could still be combined with *a book about* via the backward crossed composition rule (223d). However, that rule has not been found to be necessary thus far and was only provided for completeness, so it can perhaps be removed from the rule set. That would pleasantly reduce the rule set and disallow extraction out of subjects, but the consequences would need to be checked carefully.

Another option is to assume, following Steedman (2000b), that all noun phrases are type-raised, which immediately blocks all composition into them. Of course, this implies that objects may not be composed into. Though English does permit extraction out of objects, it is unique among most languages in doing so, and furthermore it is highly lexically restricted. Thus, there is a difference in acceptability between the following two sentences:

(251) Who did you paint a picture of?

(252) #Who did you lose a picture of?

I thus assume that where such extractions are allowed, it is due to verbs which subcategorize for type-raised noun phrases which are missing an argument. The next section also makes use of the assumption that all noun phrases are type-raised for capturing the crossing dependencies of Dutch.

5.3 Dutch

The previous section defines a multi-modal version of CCG that can cover the same English data handled under Steedman's analyses while requiring fewer categories for some constructions. Furthermore, there is no appeal to rule restrictions, which are often duplicated for different rules. In the multi-modal setting, this degree of freedom is eliminated and the burden is instead put on the choice of modalities on the slashes. This move represents a theoretical saving because we no longer have the freedom to arbitrarily place different restrictions on different rules. Instead, a category simply is or is not suitable as input to different rules, and we have thus far seen that this works effectively for a number of extraction contexts in English.

I also wish to proceed on the assumption that MULTI-MODAL CCG has an invariant rule component and thereby reduce this formerly parametric aspect of defining CCG grammars. It is therefore important to verify that this strong assumption can be maintained in the face of data from other languages. Steedman (2000b) provides an extensive analysis of Dutch that encompasses both main and subordinate clauses and deals with coordination, extraction, and control verbs. In this section, I demonstrate that MULTI-MODAL CCG is not only capable of receiving Steedman's analysis, but can substantially improve on it as well. The task will not be to provide a fundamentally different analysis from Steedman's, but to use it as a basis and reduce the level of stipulation used to capture the data.

5.3.1 Dutch Subordinate Clauses

I start with the cross-serial dependencies found in subordinate clauses. (All of my examples are adapted from Steedman (2000b)).

(253) ... omdat ik Cecilia de paarden zag voeren
... because I Cecilia the horses saw feed
... *because I saw Cecilia feed the horses*

(254) ... omdat ik Cecilia Henk de paarden zag helpen voeren
... because I Cecilia Henk the horses saw help feed
... *because I saw Cecilia help Henk feed the horses*

These sentences exhibit the intersecting dependencies for which Dutch is so famous in linguistic circles. This contrasts quite clearly with the German order, in which the dependencies are nested. Note also that the nested order is ungrammatical in Dutch:

(255) a. *... omdat ik Cecilia de paarden voeren zag

b. * ... omdat ik Cecilia Henk de paarden voeren helpen zag

Cross-serial dependencies provide the crucial evidence that natural language syntax demands greater than context-free power (Shieber, 1985; Huybregts, 1984). Unsurprisingly, the crossed composition rules, which are precisely the rules that boost CCG's power beyond context-free, are heavily implicated in the analysis of cross-serial dependencies.

The basic approach Steedman (2000b) takes is to assume that subordinate verbs in Dutch seek their verbal arguments to the right and their nominal arguments to the left. The verbal cluster is then formed through successive uses of the forward crossed composition rule. For example, consider the categories given below:

(256) a. *zag* $\vdash ((\mathsf{s}_i\backslash\mathsf{np})\backslash\mathsf{np})/_{\times}(\dot{\mathsf{s}}_v\backslash\mathsf{np})$

b. *helpen* $\vdash ((\mathsf{s}_v\backslash\mathsf{np})\backslash\mathsf{np})/_{\times}(\dot{\mathsf{s}}_v\backslash\mathsf{np})$

c. *voeren* $\vdash (\mathsf{s}_v\backslash\mathsf{np})\backslash\mathsf{np}$

In order to improve the readability of the derivations presented in this section, I assume the following abbreviation:

Convention 10 (Dutch verb phrase abbreviation). *The category of Dutch verb phrases,* $\mathsf{s}_v\backslash\mathsf{np}$*, will be abbreviated as* vp*. Features placed on the abbreviation* vp *indicate features on the* s_v *result of the full category, including the SPEC abbreviations* *of Convention 5. For example,* $\dot{\mathsf{vp}}_{inf}$ *abbreviates* $\dot{\mathsf{s}}_{v\,inf}\backslash\mathsf{np}$ *(which itself abbreviates* $\mathsf{s}_{v\,SPEC=-,\,VFORM=inf}\backslash\mathsf{n}_{SPEC=+}$*).*

With this convention, the categories for Dutch given above appear as follows:

(257) a. *zag* $\vdash ((\mathsf{s}_i\backslash\mathsf{np})\backslash\mathsf{np})/_{\times}\dot{\mathsf{vp}}$

b. *helpen* $\vdash (\mathsf{vp}\backslash\mathsf{np})/_{\times}\dot{\mathsf{vp}}$

c. *voeren* $\vdash \mathsf{vp}\backslash\mathsf{np}$

These categories enable derivations for both (253) and (254), given in (258) and (259), respectively. Note that the subscripts are used in these derivations only for readability and are not performing any grammatical work.

(258)

$$\begin{array}{cccccc}
\textit{dat} & \textit{ik} & \textit{Cecilia} & \textit{de paarden} & \textit{zag} & \textit{voeren} \\
 & \mathsf{np}_1 & \mathsf{np}_2 & \mathsf{np}_3 & ((\mathsf{s}_i\backslash\mathsf{np}_1)\backslash\mathsf{np}_2)/_{\times}\dot{\mathsf{vp}} & \mathsf{vp}\backslash\mathsf{np}_3 \\
 & & & & \multicolumn{2}{c}{\overline{((\mathsf{s}_i\backslash\mathsf{np}_1)\backslash\mathsf{np}_2)\backslash\mathsf{np}_3}\ {>}\mathbf{B}_{\times}} \\
 & & & \multicolumn{3}{c}{\overline{(\mathsf{s}_i\backslash\mathsf{np}_1)\backslash\mathsf{np}_2}\ {<}} \\
 & & \multicolumn{4}{c}{\overline{\mathsf{s}_i\backslash\mathsf{np}_1}\ {<}} \\
 & \multicolumn{5}{c}{\overline{\mathsf{s}_i}\ {<}}
\end{array}$$

(259)

$$\begin{array}{cccccccc}
\textit{dat} & \textit{ik} & \textit{Cecilia} & \textit{Henk} & \textit{de paarden} & \textit{zag} & \textit{helpen} & \textit{voeren} \\
 & \mathsf{np}_1 & \mathsf{np}_2 & \mathsf{np}_3 & \mathsf{np}_4 & ((\mathsf{s}_i\backslash\mathsf{np}_1)\backslash\mathsf{np}_2)/_{\times}\dot{\mathsf{vp}} & (\mathsf{vp}\backslash\mathsf{np}_3)/_{\times}\dot{\mathsf{vp}} & \mathsf{vp}\backslash\mathsf{np}_4 \\
 & & & & & & \multicolumn{2}{c}{\overline{(\mathsf{vp}\backslash\mathsf{np}_3)\backslash\mathsf{np}_4}\ {>}\mathbf{B}_{\times}} \\
 & & & & & \multicolumn{3}{c}{\overline{(((\mathsf{s}_i\backslash\mathsf{np}_1)\backslash\mathsf{np}_2)\backslash\mathsf{np}_3)\backslash\mathsf{np}_4}\ {>}\mathbf{B}^2_{\times}} \\
 & & & & \multicolumn{4}{c}{\overline{((\mathsf{s}_i\backslash\mathsf{np}_1)\backslash\mathsf{np}_2)\backslash\mathsf{np}_3}\ {<}} \\
 & & & \multicolumn{5}{c}{\overline{(\mathsf{s}_i\backslash\mathsf{np}_1)\backslash\mathsf{np}_2}\ {<}} \\
 & & \multicolumn{6}{c}{\overline{\mathsf{s}_i\backslash\mathsf{np}_1}\ {<}} \\
 & \multicolumn{7}{c}{\overline{\mathsf{s}_i}\ {<}}
\end{array}$$

With just these categories, the grammar does not permit the unacceptable nested orders (255). However, it does overgenerate subordinate clauses such as:

(260) ... *omdat ik Cecilia [zag]$_{((\mathsf{s}_i\backslash\mathsf{np})\backslash\mathsf{np})/_{\times}\dot{\mathsf{vp}}}$ [de paarden voeren]$_{\mathsf{vp}}$

To block such cases, I assume, like Steedman, that *all* noun complements are type-raised such that when they combine with a subordinate verb, they instantiate a feature that makes the resulting verb phrase incompatible with the verb phrase complement of the higher verb. For Steedman, the relevant feature is ±SUB, but I instead use the feature ±SPEC for this effect. Thus, in GB terms, the higher verb cannot take a verb phrase with a filled specifier as its complement. The universal form of type-raising defined in §3.3 ensures that after a type-raised argument has combined with its head, the head has a positive value for the SPEC attribute.

(261) Category for nominal complements of verb-final clauses:

$$\mathsf{Y}/_{\triangleleft}(\mathsf{Y}\backslash\mathsf{np}_{-ANT}) \equiv \ddot{\mathsf{s}}\$/_{\triangleleft}(\ddot{\mathsf{s}}\$\backslash\mathsf{np}_{-ANT})$$

As usual with type-raised categories, I will use the contextually appropriate result values in derivations and at times abbreviate them fully. Note that these complements cannot be marked for antecedent government, in line with Steedman's analysis and the earlier discussion on English.

The illicit subordinate clause (260) is now blocked as follows:

(262)

$$\begin{array}{llll}
\textit{(dat ik Cecilia)} & \textit{zag} & \textit{de paarden} & \textit{voeren} \\
 & \overline{((\mathsf{s}_i\backslash\mathsf{np}_1)\backslash\mathsf{np}_2)/_{\times}\dot{\mathsf{vp}}} & \overline{\ddot{\mathsf{vp}}/_{\triangleleft}(\ddot{\mathsf{vp}}\backslash\mathsf{np}_3)} & \overline{\mathsf{vp}\backslash\mathsf{np}_3} \\
 & & \multicolumn{2}{c}{\overline{\qquad\qquad\ddot{\mathsf{vp}}\qquad\qquad}>} \\
 & \multicolumn{3}{c}{\overline{\qquad\qquad\qquad\qquad\qquad\qquad}*}
\end{array}$$

Coordination of complement clusters works in the same manner as that discussed in Steedman (2000b). Importantly, the modalities assumed on the categories interact with the forward crossed composition rules given earlier to disallow scrambled orders in the complement cluster. For example, the sentence given in (263) does not have an interpretion in which *Cecilia* is the **Actor** argument of *zag*.

(263) ... dat Jan Cecilia zag zwemmen
... that Jan Cecilia saw swim
... that Jan saw Cecilia swim
**... that Cecilia saw Jan swim*

The unavailable interpretation can arise if *Cecilia* composes into the verbal cluster *zag zwemmen*. Steedman blocks this by restricting forward crossed composition suitably. This is unnecessary in the multi-modal setting because the categories thus far assumed and the rules we have already used in English conspire to disallow the scrambled derivation.

(264)

$$\begin{array}{lllll}
\textit{dat} & \textit{Jan} & \textit{Cecilia} & \textit{zag} & \textit{zwemmen} \\
 & \overline{\ddot{\mathsf{s}}_i/_{\triangleleft}(\ddot{\mathsf{s}}_i\backslash\mathsf{np}_{2,-ANT})} & \overline{\ddot{\mathsf{s}}_i/_{\triangleleft}(\ddot{\mathsf{s}}_i\backslash\mathsf{np}_{1,-ANT})} & \overline{((\mathsf{s}_i\backslash\mathsf{np}_1)\backslash\mathsf{np}_2)/_{\times}\dot{\mathsf{vp}}} & \overline{\mathsf{vp}} \\
 & & & \multicolumn{2}{c}{\overline{\qquad(\mathsf{s}_i\backslash\mathsf{np}_1)\backslash\mathsf{np}_2\qquad}>} \\
 & & \multicolumn{2}{c}{\overline{\qquad\qquad\ddot{\mathsf{s}}_i\backslash\mathsf{np}_{2,+ANT}\qquad\qquad}>\mathbf{B}_{\times}^{ant}} & \\
 & \multicolumn{2}{c}{\overline{\qquad\qquad\qquad\qquad\qquad}*} & &
\end{array}$$

Crucially, the fact that modalities on the primary slashes of type-raised categories have opposite directionality to their slashes requires the use of a version of forward crossed composition which leads to the creation of a category marked for antecedent government.

Though the scrambled order is blocked, the constituent formed from *Cecilia zag zwemmen* is still a suitable target for extraction, as in the following relative clause, which is ambiguous between two readings:

(265) ... [die]$_{(\mathsf{n}\backslash_{\star}\mathsf{n})/_{\star}(\mathsf{s}_i\backslash\mathsf{np}_{\pm ANT})}$ [Cecilia zag zwemmen]$_{\ddot{\mathsf{s}}_i\backslash\mathsf{np}_{2,+ANT}}$
... *who$_i$ Cecilia t_i saw swim*

(266) ... [die]$_{(\mathsf{n}\backslash_{\star}\mathsf{n})/_{\star}(\mathsf{s}_i\backslash\mathsf{np}_{\pm ANT})}$ [Cecilia zag zwemmen]$_{\ddot{\mathsf{s}}_i\backslash\mathsf{np}_{1,\pm ANT}}$
... *who$_i$ t_i saw Cecilia swim*

Thus, using the same rules that were assumed to be operative in English and without using restrictions on forward crossed composition, we capture the difference between an instance of crossed composition between two subordinate verbs and that between a complement and a verb in an extraction context.

Both the subject and object can be extracted from embedded sentences, unlike English.

(267) a. arts die ik denk dat het werk heeft gedaan
doctor who I think that the work has done
**doctor who I think that did the work*

b. werk dat ik denk dat ze heeft gedaan
work that I think that she has done
work that I think that she did

The category assignments (268) for *dat* and (269) for *denk* support analyses of both extractions, parallel to those given by Steedman (2000b).

(268) $\textit{dat} \vdash \ddot{\mathsf{s}}_c/_{\triangleleft}\mathsf{s}_i$

(269) $\textit{denk} \vdash (\mathsf{s}_i\backslash\mathsf{np})/_{\triangleleft}\ddot{\mathsf{s}}_c$

Interestingly, the category for *denk* is almost identical to the category $(\mathsf{s}_i\backslash\mathsf{np})/_{\triangleleft}\mathsf{s}_{fin}$ of words like *think* and *knew* in English (see (238)), apart from the fact that the latter category is underspecified for the MARKING feature. This is of course a welcome result given the common lineage of the two languages.

5.3.2 Dutch Equi Verbs

Equi verbs such as *proberen* 'to try' allow greater freedom of subordinate clause word order than others such as *zien* 'to see'. Specifically, verbal arguments of *proberen* and the arguments of those verbs can appear to its right.

(270) a. ... omdat ik Jan het lied probeer te leren (*te) zingen
... because I Jan the song try to teach (*to) sing
... *because I try to teach Jan to sing the song*

b. ... omdat ik probeer Jan het lied te leren (*te) zingen

c. ... omdat ik probeer Jan te leren het lied te zingen

d. ... omdat ik Jan probeer te leren het lied te zingen

e. ?... omdat ik Jan probeer het lied te leren (*te) zingen

The basic pattern can be captured by specifing the verb phrase argument of *probeer* to be rooted in the infinitival sentential category $\mathsf{s}_{v\,VFORM=inf}$, where *inf* is a subtype of *non−fin*. This category thus can unify with sententially rooted categories that have either a filled or unfilled specifier (i.e. ±SPEC). The category for *te* turns a verbal category from one rooted in $\mathsf{s}_{v\,VFORM=base}$ to $\mathsf{s}_{v\,VFORM=inf}$, just as *to* does in English.

(271) $\textsf{\textit{probeer}} \vdash (\mathsf{s}_i\backslash\mathsf{np})/_{\times}\mathsf{vp}_{inf}$

(272) $\textsf{\textit{te}} \vdash \mathsf{s}_{v\,inf}\$/_{\star}\mathsf{s}_{v\,base}\$$

(273) $\textsf{\textit{leren}} \vdash (\mathsf{vp}_{base}\backslash\mathsf{np})/_{\times}\dot{\mathsf{v}}\mathsf{p}$

By targeting verbal categories with *optionally* filled specifiers, the category for *probeer* allows it to combine with a verb phrase which has already combined with one or more of its arguments.

The order in (270a) clearly receives a derivation similar to the other Dutch subordinate clauses we have seen thus far. The order in (270b), in which *probeer* combines with a saturated verb phrase, proceeds as in (274). I will henceforth use the abbreviated form of type raised categories given in Convention 4 to save space in derivations. I will also ignore the finer grained distinctions given above for different kinds of s_v categories, focussing instead on the position of *probeer* within the clause.

(274) dat *ik* *probeer* *Jan* *het lied* *te leren* *zingen*

$$\begin{array}{llllll}
\mathsf{np}^{\uparrow}_{1} & (\mathsf{s}_i\backslash\mathsf{np}_1)/_{\times}\mathsf{vp} & \mathsf{np}^{\uparrow}_{2} & \mathsf{np}^{\uparrow}_{3} & (\mathsf{vp}\backslash\mathsf{np}_2)/_{\times}\dot{\mathsf{vp}} & \mathsf{vp}\backslash\mathsf{np}_3
\end{array}$$

$$\begin{array}{ll}
(\mathsf{vp}\backslash\mathsf{np}_2)\backslash\mathsf{np}_3 & {>}\mathbf{B}_{\times} \\
\ddot{\mathsf{vp}}\backslash\mathsf{np}_2 & > \\
\ddot{\mathsf{vp}} & > \\
\mathsf{s}_i\backslash\mathsf{np} & > \\
\ddot{\mathsf{s}}_i & >
\end{array}$$

The subtle alternation regarding the presence or absence of *te* exhibited in the sentences in (270) is not covered by the categories I have given. To cover the cases in which *te leren* applies to a verb which has already consumed an argument, we must provide a modalized version of the extra category which Steedman gives for *te*.

(275) *The Modalized Te Category Brute Force Stipulation*:
$te \vdash (\mathsf{vp}_{inf}\$_1/_{\star}\mathsf{vp}_{inf})/_{\star}(\mathsf{vp}_{base}\$_1/_{\times}\mathsf{vp}_{base})$

With this category, we can provide a derivation for the order in which the complement of *leren* combines with its argument before being consumed. The $\star$ modality in the result ensures that *te leren* cannot compose into its complement.

(276) dat *ik* *probeer* *Jan* *te leren* *het lied* *te zingen*

$$\begin{array}{llllll}
\mathsf{np}^{\uparrow} & (\mathsf{s}_i\backslash\mathsf{np})/_{\times}\mathsf{vp} & \mathsf{np}^{\uparrow} & (\mathsf{vp}\backslash\mathsf{np})/_{\star}\mathsf{vp} & \mathsf{np}^{\uparrow} & \mathsf{vp}\backslash\mathsf{np}
\end{array}$$

$$\begin{array}{ll}
\ddot{\mathsf{vp}} & > \\
\mathsf{vp}\backslash\mathsf{np} & > \\
\ddot{\mathsf{vp}} & > \\
\mathsf{s}_i\backslash\mathsf{np} & > \\
\ddot{\mathsf{s}}_i & >
\end{array}$$

The category employed by Steedman to handle the *te* distribution raises an interesting point with respect to the current enterprise. His category, translated into the present notation and conventions, is the following:

(277) $te \vdash (\mathsf{vp}_{inf}\$_1/\mathsf{vp}_{inf,+SUB})/(\mathsf{vp}_{base}\$_1/\mathsf{vp}_{-SUB})$

The string *te leren* thus has the additional category $(\mathsf{vp}_{inf}\backslash\mathsf{np})/\mathsf{vp}_{inf,+SUB}$, which captures (270c,d). It does not overgenerate *te* in front of *zingen* in (270a,b,e) because it cannot compose with *te zingen* due to the restriction Steedman places on forward crossed composition:

(278) *Dutch forward crossed composition* ($>\mathbf{B}_\times$)

$$\mathsf{X/Y} \quad \mathsf{Y\backslash Z} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X\backslash Z}$$

where $\mathsf{Y} = \mathsf{vp}_{-SUB}$

This restriction was actually put in place to control other aspects of the grammar. The extra category for *te* was thus created with *explicit* reference to the restriction made in the rule (278) so that its result could only be used with the forward application rule. Furthermore, the ±SUB feature is intended to distinguish matrix and subordinate clauses, but Steedman's use of the feature to control the combinatory rules actually forces him to abuse its intuitive meaning. In the present setting, control of the combinatory rules is achieved by principled usage of different modes with different behaviors.

It should be pointed out that an alternative to the use of the extra category for *te* would be to allow the verb phrase complement of *leren* to be rooted in $\mathsf{s}_{v\,non-fin}$ and to have an optionally filled specifier. This would permit all of the orders in (270), but would also overgenerate by allowing *te* in the locations where it is indicated as impermissible. Nonetheless, the impossibility of *te* appearing in those locations could be considered as a problem of a more phonological nature, as suggested by Steedman (p. 275, fn. 13).

A further advantage of the multi-modal setting is that it actually rules out some derivations which can only be blocked by stipulation on the restrictional account. Consider the following:

(279) ... dat Jan het lied probeert te zingen
... that Jan the song tries to sing
... *that Jan tries to sing the song*

(280) *... dat het lied Jan probeert te zingen

The scrambled order is not grammatical, but it can actually arise through harmonic composition of *Jan* and *probeert* in the standard CCG setting, which utilizes just one mode of grammatical composition.

(281)

$$\begin{array}{lcccc}
\textit{*dat} & \textit{het lied} & \textit{Jan} & \textit{probeert} & \textit{te zingen} \\
\cline{2-5}
 & \ddot{\mathsf{s}}_i/(\ddot{\mathsf{s}}_i\backslash\mathsf{np}_2) & \ddot{\mathsf{s}}_i/(\ddot{\mathsf{s}}_i\backslash\mathsf{np}_1) & (\mathsf{s}_i\backslash\mathsf{np}_1)/\mathsf{vp} & \mathsf{vp}\backslash\mathsf{np}_2 \\
\cline{3-4}
 & & \multicolumn{2}{c}{\ddot{\mathsf{s}}_i/\mathsf{vp}} \quad >\mathbf{B} & \\
\cline{3-5}
 & & \multicolumn{3}{c}{\ddot{\mathsf{s}}_i\backslash\mathsf{np}_2} \quad >\mathbf{B}_\times \\
\cline{2-5}
 & \multicolumn{4}{c}{\ddot{\mathsf{s}}_i} \quad > \\
\end{array}$$

To block this derivation, Steedman defines a modified form of forward harmonic composition and places a restriction on it (Steedman, 2000b, p. 154):[5]

(282) *Dutch forward harmonic composition* ($>\mathsf{B}$)

$\mathsf{X/Y} \quad \mathsf{Y/(Y\backslash Z)} \quad \Rightarrow_{\mathsf{B}} \quad \mathsf{X/(Y\backslash Z)}$

where $\mathsf{Y} = \mathsf{s}\backslash\$$

With MULTI-MODAL CCG, however, the unavailability of the scrambled order is predicted by the modes which we have already assumed for other reasons. The categories for verbs such as *probeert* simply do not have the appropriate slash type to serve as input to forward harmonic composition.

(283) *dat het lied [Jan]$_{\mathsf{s}_i/_{\triangleleft}(\mathsf{s}_i\backslash\mathsf{np})}$ [probeert]$_{(\mathsf{s}_i\backslash\mathsf{np})/_{\times}\mathsf{vp}}$ te zingen

The fact that *probeert* must forward cross compose with its verb phrase argument correctly predicts that it cannot be composed *into* harmonically. The account is less arbitrary than one which accounts for this data with separate restrictions on forward crossed composition and forward harmonic composition. The situation is even worse in a restrictional account because there is no *simple* restriction which can deal with the data — instead it is necessary to modify the actual forms of the rules as in (282), where the second input category is not that standardly assumed. Essentially, the form of (282) amounts to limiting forward composition to combining only type-raised arguments. This could lead to problems when extending the analysis to further data which might quite possibly require non-type-raised arguments, such as adjectives, to compose.

5.3.3 Dutch Main Clauses

The modal encoding of Steedman's analysis extends straightforwardly for Dutch main clauses and furthermore backs up the claim that the versions of backward crossed composition given earlier during the discussion on English extend to other languages without modification. Here is some example data:

(284) Het meisje at de appels.
the girl ate the apples
The girl ate the apples.

[5]The restriction given by Steedman, $\mathsf{Y}=\mathsf{s}$, is actually too strict and does not permit the rule to apply as desired. The statement of the restriction given in (282) corrects this problem.

(285) Hij gaf de politieman een bloem.
he gave the policeman a flower
He gave the policeman a flower.

Other orders are possible in which either the verb precedes all its dependents or one of the other dependents is fronted before the verb:

(286) a. At het meisje de appels?

b. De appels at het meisje.

(287) a. Gaf hij de politieman een bloem?

b. De politieman gaf hij een bloem.

c. Een bloem gaf hij de politieman.

Like Steedman, I assume that main clause verbs have separate categories that are related to the subordinate categories by a lexical rule, the details of which I will ignore here. In addition to covering the orders given above, we also wish to ensure that scrambled orders such as the following cannot be derived:

(288) *At de appels het meisje?
(for: *Did the girl eat the apples?*)

(289) a. *Gaf hij een bloem de politieman?

b. *Gaf de politieman een bloem hij?

(for: *Did he give the policeman a flower?*)

As with complements in subordinate contexts, complements in main clauses are assumed to be type-raised. The basic word order is assumed to be the verb-initial one, as reflected in the following categories.

(290) *at* $\vdash (\mathsf{s}_i/\mathsf{np}^{\mathbf{Pat}})/_{\triangleleft}\mathsf{np}^{\mathbf{Act}}$

(291) *gaf* $\vdash ((\mathsf{s}_i/\mathsf{np}^{\mathbf{Pat}})/\mathsf{np}^{\mathbf{Addr}})/_{\triangleleft}\mathsf{np}^{\mathbf{Act}}$

Scrambled sentences such as (288) and (289) are not derivable with these categories because type-raising followed by backward crossed composition is required to consume the inner argument and this results in the outer argument being marked for antecedent government. For example:

(292)
$$\begin{array}{ccc}
\textit{at} & \textit{de appels} & \textit{het meisje} \\
\hline
(\mathsf{s}_i/\mathsf{np}^{\mathbf{Pat}})/_{\triangleleft}\mathsf{np}^{\mathbf{Act}} & \ddot{\mathsf{s}}_i\backslash_{\triangleright}(\ddot{\mathsf{s}}_i/_{\triangleright}\mathsf{np}^{\mathbf{Pat}}_{-ANT}) & \ddot{\mathsf{s}}_i\backslash_{\triangleright}(\ddot{\mathsf{s}}_i/_{\triangleright}\mathsf{np}^{\mathbf{Act}}_{-ANT}) \\
\hline
\multicolumn{2}{c}{\ddot{\mathsf{s}}_i/\mathsf{np}^{\mathbf{Act}}_{+ANT} \quad {<}\mathbf{B}^{ant}_{\times}} & \\
\hline
\multicolumn{3}{r}{*}
\end{array}$$

To handle the sentences in which one of the elements is fronted, I assume, like Steedman, that Dutch has a category for fronted elements which is quite similar to that of topicalized elements in English and which embodies the GB hypothesis that these elements occupy SPEC of CP.

(293) $\ddot{\mathsf{s}}_c/_{\diamond}(\mathsf{s}_i/_{\cdot}\mathsf{np}_{\pm ANT})$

Sentences such as (285) and (287b,c) are then derived as follows:

(294)
$$
\begin{array}{ccccc}
\textit{Hij} & \textit{gaf} & \textit{de politieman} & \textit{een bloem} & \\
\hline
\ddot{\mathsf{s}}_c/_{\diamond}(\mathsf{s}_i/_{\cdot}\mathsf{np}^{\mathbf{Act}}_{\pm ANT}) & ((\mathsf{s}_i/\mathsf{np}^{\mathbf{Pat}})/\mathsf{np}^{\mathbf{Addr}})/_{\triangleleft}\mathsf{np}^{\mathbf{Act}} & \mathsf{np}^{\uparrow\mathbf{Addr}} & \mathsf{np}^{\uparrow\mathbf{Pat}} & \\
\cline{2-3}
 & \multicolumn{2}{c}{(\ddot{\mathsf{s}}_i/\mathsf{np}^{\mathbf{Pat}})/_{\triangleleft}\mathsf{np}^{\mathbf{Act}}_{+ANT}} & {<}\mathbf{B}^{ant}_{\times} & \\
\cline{2-4}
 & \multicolumn{3}{c}{\ddot{\mathsf{s}}_i/_{\triangleleft}\mathsf{np}^{\mathbf{Act}}_{+ANT}} & {<}\mathbf{B}^{ant}_{\times} \\
\cline{1-3}
\multicolumn{3}{c}{\ddot{\mathsf{s}}_c} & {>} &
\end{array}
$$

(295)
$$
\begin{array}{ccccc}
\textit{De politieman} & \textit{gaf} & \textit{hij} & \textit{een bloem} & \\
\hline
\ddot{\mathsf{s}}_c/_{\diamond}(\mathsf{s}_i/_{\cdot}\mathsf{np}^{\mathbf{Addr}}_{\pm ANT}) & ((\mathsf{s}_i/\mathsf{np}^{\mathbf{Pat}})/\mathsf{np}^{\mathbf{Addr}})/_{\triangleleft}\mathsf{np}^{\mathbf{Act}} & \mathsf{np}^{\uparrow\mathbf{Act}} & \mathsf{np}^{\uparrow\mathbf{Pat}} & \\
\cline{2-3}
 & \multicolumn{2}{c}{(\ddot{\mathsf{s}}_i/\mathsf{np}^{\mathbf{Pat}})/\mathsf{np}^{\mathbf{Addr}}} & {<} & \\
\cline{2-4}
 & \multicolumn{3}{c}{\ddot{\mathsf{s}}_i/\mathsf{np}^{\mathbf{Addr}}_{+ANT}} & {<}\mathbf{B}^{ant}_{\times} \\
\cline{1-3}
\multicolumn{3}{c}{\ddot{\mathsf{s}}_c} & {>} &
\end{array}
$$

(296)
$$
\begin{array}{ccccc}
\textit{Een bloem} & \textit{gaf} & \textit{hij} & \textit{de politieman} & \\
\hline
\ddot{\mathsf{s}}_c/_{\diamond}(\mathsf{s}_i/_{\cdot}\mathsf{np}^{\mathbf{Pat}}_{\pm ANT}) & ((\mathsf{s}_i/\mathsf{np}^{\mathbf{Pat}})/\mathsf{np}^{\mathbf{Addr}})/_{\triangleleft}\mathsf{np}^{\mathbf{Act}} & \mathsf{np}^{\uparrow\mathbf{Act}} & \mathsf{np}^{\uparrow\mathbf{Addr}} & \\
\cline{2-3}
 & \multicolumn{2}{c}{(\ddot{\mathsf{s}}_i/\mathsf{np}^{\mathbf{Pat}})/\mathsf{np}^{\mathbf{Addr}}} & {<} & \\
\cline{2-4}
 & \multicolumn{3}{c}{\ddot{\mathsf{s}}_i/\mathsf{np}^{\mathbf{Pat}}} & {<} \\
\cline{1-3}
\multicolumn{3}{c}{\ddot{\mathsf{s}}_c} & {>} &
\end{array}
$$

An incremental derivation is also available for all of these sentences if the category for the fronted element is generalized as follows:

(297) $\ddot{\mathsf{s}}_c\$_1/_{\diamond}(\mathsf{s}_i\$_1/_{\cdot}\mathsf{np}_{\pm ANT})$

Doubly topicalized main clauses are not permitted by the grammar because of the CP/IP distinction encoded in the categories:

(298) $*[\text{Jan}]_{\ddot{\mathsf{s}}_c/_{\diamond}(\mathsf{s}_i/_{\cdot}\mathsf{np}^{\mathbf{Pat}}_{\pm ANT})}$ $[\text{appels at}]_{\ddot{\mathsf{s}}_c/\mathsf{np}^{\mathbf{Pat}}}$

Of course, this sequence does have a derivation as a subordinate clause with the category $\ddot{\mathsf{s}}_i$.

An interesting situation arises with adverb placement in main clauses. Adverbs, such as *gisteren* 'yesterday', may not come between the verb and its subject:

(299) a. Jan at gisteren appels.

b. *Appels at gisteren Jan.

(300) a. At Jan gisteren appels?

b. *At gisteren Jan appels?

(301) a. Ik gaf gisteren het meisje de appels.

b. Ik gaf het meisje gisteren de appels.

(302) a. *Het meisje gaf gisteren ik de appels.

b. Het meisje gaf ik gisteren de appels.

Notice the similarity of (302a) with the inability in English for an adverb to come between the verb and the **Addressee** in ditransitive sentences, as in (228b). This was controlled in English by assuming that the slash for the **Addressee** had a modality with directionality opposite to that of the slash itself. The category for the Dutch ditransitive has the same slash on its **Actor** argument, which in combination with the following category for *gisteren* conspires to capture the adverb placement data given above.

(303) $\mathit{gisteren} \vdash \mathsf{s}\backslash\mathsf{s}$

These examples highlight an important difference between Steedman's use of the ±SHIFT feature to control permutativity and the multi-modal approach. In order to block scrambled orders such as (289a) and ungrammatical adverb placement such as (302a), Steedman requires both the **Actor** and the **Addressee** to be marked as −SHIFT.

(304) $\mathit{gaf} \vdash ((\mathsf{s}_i/\mathsf{np}^{\mathbf{Pat}})/\mathsf{np}^{\mathbf{Addr}}_{-SHIFT})/\mathsf{np}^{\mathbf{Act}}_{-SHIFT}$

This category, however, also blocks the grammatical sentence (301a), and there is no simple way to capture that sentence without permitting scrambling of the **Addressee** with respect to the **Patient**.

I assume the same analysis as Steedman for main clause verbs which subcategorize for verbal arguments, such as *zal* 'shall', *heeft* 'has' and *zag* 'saw'. The translation into the present setting is transparent, and the predictions are essentially the same.

5.3.4 Topicalization Out of Subordinate Clauses

The topicalized element in a Dutch sentence need not necessarily be an argument of the matrix verb, as seen in the following sentence:

(305) Appels denk ik dat Jan heeft gegeten
Apples think I that Jan has eaten
Apples, I think that Jan has eaten.

For these extractions, we need the additional topicalization category (306) as well as categories for *denk* (307) and *dat* (308).

(306) $\ddot{\mathsf{s}}_c /_{\diamond} (\mathsf{s}_i \backslash_{\cdot} \mathsf{np}_{\pm ANT})$

(307) $\textit{denk} \vdash (\mathsf{s}_i /_{\triangleleft} \ddot{\mathsf{s}}_c) /_{\triangleleft} \mathsf{n}$

(308) $\textit{dat} \vdash \ddot{\mathsf{s}}_c /_{\times} \mathsf{s}_i$

(309)
$$
\begin{array}{cccc}
\textit{Appels} & \textit{denk ik} & \textit{dat} & \textit{Jan heeft gegeten} \\
\hline
\ddot{\mathsf{s}}_c /_{\diamond} (\mathsf{s}_i \backslash_{\cdot} \mathsf{np}_{\pm ANT}) & \ddot{\mathsf{s}}_i / \ddot{\mathsf{s}}_c & \ddot{\mathsf{s}}_c / \mathsf{s}_i & \ddot{\mathsf{s}}_i \backslash \mathsf{np}_{+ANT} \\
\cline{3-4}
 & & \multicolumn{2}{c}{\ddot{\mathsf{s}}_c \backslash \mathsf{np}_{+ANT}} \quad {>}\mathbf{B}_{\times} \\
\cline{2-4}
 & \multicolumn{3}{c}{\ddot{\mathsf{s}}_i \backslash \mathsf{np}_{+ANT}} \quad {>}\mathbf{B}_{\times} \\
\hline
\multicolumn{4}{c}{\ddot{\mathsf{s}}_c} \quad {>}
\end{array}
$$

The grammar as it stands does overgenerate in an interesting way — it predicts that a verb like *denk* can take a main clause sentence as its argument when the complementizer is absent:

(310) *[Ik denk]$_{\mathsf{s}_i/\ddot{\mathsf{s}}_c}$ [Jan heeft appels gegeten]$_{\ddot{\mathsf{s}}_c}$

Though this does not appear to be grammatical in Dutch, it is possible for certain verbs in other verb-second languages, as seen in the following German example from (Iatridou and Kroch, 1992):

(311) a. Er sagte, daß er kommen würde.
he said that he come would

b. Er sagte, er würde kommen.
he said he would come

Another way in which the grammar overgenerates without a distinction between matrix and subordinate contexts is less interesting and certainly forces the need for such a distinction. The topicalization category (306) can apply to a subordinate clause verb directly to produce the category $\ddot{\mathsf{s}}_c$ as follows:

(312) [Appels]$_{\ddot{\mathsf{s}}_c /_{\diamond} (\mathsf{s}_i \backslash_{\cdot} \mathsf{np}_{\pm ANT})}$ [Jan at]$_{\ddot{\mathsf{s}}_i \backslash \mathsf{np}_{+ANT}}$

These examples indicate that we do need to encode some sort of distinction between main clauses and subordinate clauses in Dutch, akin to the ±SUB feature used by Steedman (2000b). Nonetheless, the present analysis would utilize the feature only within its intended meaning rather than depending on it to control the crossed composition rules.

5.4 Substitution

The use of modalities to license harmonic and crossed composition generalizes quite readily to harmonic and crossed substitution rules. Since substitution is syntactically a generalized form of composition, we can consider using the same modalities employed for composition. For example, the harmonic substitution rules are defined as follows:

(313) $(\mathsf{X}/_{\diamond}\mathsf{Y})/_{\diamond}\mathsf{Z} \quad \mathsf{Y}/_{\diamond}\mathsf{Z} \;\Rightarrow_{\mathbf{S}}\; \mathsf{X}/_{\diamond}\mathsf{Z}$ $(>\mathbf{S})$

(314) $\mathsf{Y}\backslash_{\diamond}\mathsf{Z} \quad (\mathsf{X}\backslash_{\diamond}\mathsf{Y})\backslash_{\diamond}\mathsf{Z} \;\Rightarrow_{\mathbf{S}}\; \mathsf{X}\backslash_{\diamond}\mathsf{Z}$ $(<\mathbf{S})$

The forward substitution rule (313) will clearly permit the following along a standard CCG analysis (Steedman, 1996):

(315) a team which I [persuaded every detractor of]$_{((\mathsf{s}\backslash\mathsf{np})/(\mathsf{s}\backslash\mathsf{np}))/_{\triangleleft}\mathsf{np}}$ [to cheer for]$_{(\mathsf{s}\backslash\mathsf{np})/_{\triangleleft}\mathsf{np}}$

The crossed substitution rules are defined analogously.

(316) $(\mathsf{X}/_{\times}\mathsf{Y})\backslash_{\times}\mathsf{Z} \quad \mathsf{Y}\backslash_{\times}\mathsf{Z} \;\Rightarrow_{\mathbf{S}}\; \mathsf{X}\backslash_{\times}\mathsf{Z}$ $(>\mathbf{S}_{\times})$

(317) $\mathsf{Y}/_{\times}\mathsf{Z} \quad (\mathsf{X}\backslash_{\times}\mathsf{Y})/_{\times}\mathsf{Z} \;\Rightarrow_{\mathbf{S}}\; \mathsf{X}/_{\times}\mathsf{Z}$ $(<\mathbf{S}_{\times})$

(318) John [watched]$_{(\mathsf{s}\backslash\mathsf{np})/\mathsf{np}}$ [without enjoying]$_{((\mathsf{s}\backslash\mathsf{np})\backslash(\mathsf{s}\backslash\mathsf{np}))/\mathsf{np}}$ the game between Germany and Paraguay.

Once again, we may dispense with the usual restrictions placed on $<\mathbf{S}_{\times}$ and leave the control up to the lexicon. Furthermore, based on categories that we have defined already in other contexts, we predict that some potential cases of substitution are in fact not grammatical:

(319) *John [supported]$_{(\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np}}$ [without]$_{((\mathsf{s}\backslash\mathsf{np})\backslash(\mathsf{s}\backslash\mathsf{np}))/_{\triangleleft}\mathsf{np}}$ the team.

(320) *a [picture of]$_{\mathsf{n}/_{\triangleleft}\mathsf{np}}$ [for]$_{(\mathsf{n}\backslash_{\star}\mathsf{n})/_{\triangleleft}\mathsf{np}}$ Ronaldo

The fact that modal restrictions put in place in crossed composition contexts extend naturally to substitution contexts is unsurprising given that the CTL rules defined by Kruijff and Baldridge (2000) to simulate substitution depend on the structural rules for associativity and permutivity. To simulate any one of the CCG substitution rules, we need one of the four CTL structural rules for composition – Right Association (RA), Left Association (LA), Mixed Right Permutation (MRP), and Mixed Left Permutation(MLP) – and one of the two rules for substitution: Right Substitution (RS) and Left Substitution (LS). For example, to simulate $>\mathbf{S}$, we need RA and RS, and for $<\mathbf{S}_{\times}$ we need MRP and RS. Since RA and MRP independently simulate $>\mathbf{B}$ and backward crossed composition respectively, this implies that we could not have a simulation of a CCG grammar which included $>\mathbf{B}$, $<\mathbf{B}_{\times}$, and $>\mathbf{S}$, but lacked $<\mathbf{S}_{\times}$. The following table summarizes how the six CTL structural rules interact to simulate the eight combinatory rules.

CTL rules	simulate	CCG rules
RA + RS	$\Rightarrow$	$>\mathbf{B}$ + $>\mathbf{S}$
MRP + RS	$\Rightarrow$	$<\mathbf{B}_{\times}$ + $<\mathbf{S}_{\times}$
LA + LS	$\Rightarrow$	$<\mathbf{B}$ + $<\mathbf{S}$
MLP + LS	$\Rightarrow$	$>\mathbf{B}_{\times}$ + $>\mathbf{S}_{\times}$

This predicts that a language which uses a given substitution rule will necessarily exploit the corresponding composition rule. Upon preliminary investigation of English, Portuguese, German and Dutch, this prediction appears to hold. In the present multi-modal setting, English has lexical entries which license the rules $>\mathbf{B}$, $<\mathbf{B}$, $<\mathbf{B}_{\times}$, $>\mathbf{S}$, and $<\mathbf{S}_{\times}$, which can be simulated with RA, LA, MRP, and RS. The same appears to be true of Portuguese. Exemplar sentences and abbreviated derivations are given for each of the crossed and substitution rules. I use vp as an abbreviation for $\mathsf{s}\backslash_{.}\mathsf{np}$ and $\mathsf{s}/_{.}\mathsf{np}$.

(321) English

a. John noticed suddenly the man with the big black briefcase.
   $<\mathbf{B}_{\times}$: [noticed]$_{\mathsf{vp/np}}$ [suddenly]$_{\mathsf{vp\backslash vp}}$

b. Which articles did you file without reading?
   $<\mathbf{S}_{\times}$: [file]$_{\mathsf{vp\backslash np}}$ [without reading]$_{\mathsf{(vp\backslash vp)/np}}$

c. He is the man I will persuade every friend of to vote for.
   $>\mathbf{S}$: [persuade every friend of]$_{\mathsf{(vp/vp)/_{\triangleleft}np}}$ [to vote for]$_{\mathsf{vp/np}}$

(322) Portuguese

a. João registrou rapidamente os arquivos.
John filed quickly the articles
John quickly filed the articles.

$<\mathbf{B}_{\times}$: $[\text{registrou}]_{\mathsf{vp/np}}$ $[\text{rapidamente}]_{\mathsf{vp\backslash vp}}$

b. Que arquivos você registrou sem ter lido?
which articles you filed without having read
Which articles did you file without reading?

$<\mathbf{S}_{\times}$: $[\text{registrou}]_{\mathsf{vp/np}}$ $[\text{sem ter lido}]_{\mathsf{(vp\backslash vp)/np}}$

c. Eu sem verbalizar prometi ao João que eu faria o trabalho.
I without mentioning promised to John that I do-FUT the work
I without mentioning (that I would do the work) promised John that I would do the work.

$>\mathbf{S}$: $[\text{sem verbalizar}]_{\mathsf{(vp/vp)/\ddot{s}}_c}$ $[\text{prometi ao João}]_{\mathsf{vp/\ddot{s}}_c}$

German and Dutch exemplify the complement of these relationships in English and Portuguese by having lexicons which license $>\mathbf{B}_{\times}$, $<\mathbf{S}$, and $>\mathbf{S}_{\times}$.

(323) German

a. den Hund den ich fütterte
the dog that I fed
$>\mathbf{B}_{\times}$: $[\text{ich}]_{\mathsf{s}/_{\triangleleft}\mathsf{(s\backslash np)}}$ $[\text{fütterte}]_{\mathsf{(s\backslash np)\backslash np}}$

b. Welche Artikel hast du ohne zu lesen abgelegt?
which article have you without reading away-put
$>\mathbf{S}_{\times}$: $[\text{ohne zu lesen}]_{\mathsf{(vp/vp)\backslash np}}$ $[\text{abgelegt}]_{\mathsf{vp\backslash np}}$

c. Welche Artikel hast du abgelegt ohne zu lesen?
which article have you away-put without reading
$<\mathbf{S}$: $[\text{abgelegt}]_{\mathsf{vp\backslash np}}$ $[\text{ohne zu lesen}]_{\mathsf{(vp\backslash vp)\backslash np}}$

(324) Dutch

a. ...omdat ik Cecilia de paarden zag voeren
that I Cecilia the horses saw feed
...that I saw Cecilia feed the horses

$>\mathbf{B}_{\times}$: $[\text{zag}]_{\mathsf{((s\backslash np)\backslash np)}/_{\times}\mathsf{vp}}$ $[\text{voeren}]_{\mathsf{vp\backslash np}}$

b. Welke boeken heb je zonder te lezen weggezet?
which books have you without reading away-put
$>\mathbf{S}_{\times}$: $[\text{zonder te lezen}]_{\mathsf{(vp/vp)\backslash np}}$ $[\text{weggezet}]_{\mathsf{vp\backslash np}}$

c. Welke boeken heb je weggezet zonder te lezen?
which books have you away-put without reading
$<\mathbf{S}$: $[\text{weggezet}]_{\mathsf{vp\backslash np}}$ $[\text{zonder te lezen}]_{\mathsf{(vp\backslash vp)\backslash np}}$

With MULTI-MODAL CCG, each language's grammar has the same rule set, but the lexicons will exploit that set in potentially different ways. The prediction is the same as that made by Kruijff and Baldridge (2000) — given the definitions of backward crossed composition in (200b) and backward crossed substitution in (317), if a language has the categories to exploit the latter, it will *necessarily* exploit the former.

5.5 Generative Power

Given the significance attached to generative power in the CCG tradition, it is important to consider whether the modifications proposed in this chapter increase the strength of the system. The answer is that the formulation provided here does not have greater power than standard CCG. This can be easily shown by encoding MULTI-MODAL CCG as a standard CCG system with special rule restrictions. Since MULTI-MODAL CCG does not utilize any new rules of combination, a MULTI-MODAL CCG grammar can be simulated with standard CCG by translating the modes as features on the ultimate targets of (possibly complex) categories and then formulating the rules with restrictions that reference those features. For example, the multi-modal category (325) would be converted into (326).

(325) $(\mathsf{s}\backslash_{\cdot}\mathsf{np})/_{\diamond}(\mathsf{s}\backslash_{\cdot}\mathsf{np})$

(326) $(\mathsf{s}\backslash\mathsf{np}_{mode=\cdot})/(\mathsf{s}_{mode=\diamond}\backslash\mathsf{np}_{mode=\cdot})$

Forward composition would then be formulated as in (327). The other rules would receive similar definitions, with restrictions keyed to the appropriate modes.

(327) *Forward composition* ($>\mathbf{B}$)

$\mathsf{X/Y} \quad \mathsf{Y/Z} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X/Z}$

where $\mathsf{Y} = \mathsf{a}_{mode=\diamond}\$_1$ and $\mathsf{Z} = \mathsf{a}_{mode=\diamond}\$_2$

Using restrictions in this manner in standard CCG would technically provide the same advantages as MULTI-MODAL CCG, but is a less clean formulation that loses the close connection with structural control in CTL. Furthermore, by using modally decorated slashes to enforce these effects, we have a clear separation between specifications that control category combining operations and more standard features which encode distinctions such as number, gender, verbal voice, tense, etc.

Essentially, MULTI-MODAL CCG provides a principled stratification of the rules already assumed by CCG. Thus, an extremely permissive lexicon which uses only the most general slash types $\backslash_{.}$ and $/_{.}$ essentially becomes a standard CCG lexicon. On the other hand, a lexicon which avails itself of *only* the non-associative slashes $\backslash_{\star}$ and $/_{\star}$ will be reduced to an AB lexicon (see §2.1) and therefore be effectively context-free like AB (Bar-Hillel *et al.*, 1964). A lexicon which uses the harmonic slash types $\backslash_{\diamond}$ and $/_{\diamond}$ will exhibit associativity yet remain context-free due to the fact that harmonic composition is a theorem of the Lambek calculus and the Lambek calculus is context-free (Pentus, 1997). What the modal refinement of CCG thus amounts to is the ability to utilize less powerful systems, in a *principled* manner, within the overall mildly context-sensitive apparatus of standard CCG. This is similar to the manner in which the CTL approach allows weaker logical systems to be embedded within more powerful ones (Kurtonina and Moortgat, 1997). As this chapter has demonstrated, natural language grammars demand this capability.

In the event that the mildly context-sensitive limitation of CCG proves untenable in the face of new linguistic phenomena, the resource-management capabilities of MULTI-MODAL CCG provides excellent prospects for scaling up the generative power of CCG. More powerful operations (rules) can be utilized for exactly the phenomena which require greater power while leaving the bulk of linguistic analyses intact at the weaker level currently provided by CCG. We will thus not need to rethink everything if presented with such a situation.

As an example of a new rule that would increase the generative power of the system, consider a modalized formulation of the rule of Generalized Weak Permutation (Briscoe, 1997) in which we employ a new mode $\otimes$ to permit some element to permute with others:

(328) $\mathsf{X}/_{\otimes}\mathsf{Y}\$ \Rightarrow_{\mathbf{P}} \mathsf{X}\$/_{\otimes}\mathsf{Y}$ $(>\mathbf{P})$

This rule corresponds to the combinator $\mathbf{C}$, (which is the Cardinal in Smullyan's combinatory fable (Smullyan, 1985)). It is defined as follows:

(329) $\mathbf{C}f \equiv \lambda x.\lambda y.(fy)x$

Defined in this way, Generalized Weak Permutation allows an argument with the permuting slash type to shift not only with respect to all other arguments in the category as in (330), but also with arguments of a category that its category has composed into (331).

(330) $((\mathsf{s}/_{\otimes}\mathsf{n}_1)/_{\star}\mathsf{n}_2)/_{\diamond}\mathsf{s} \quad \mathsf{n}_1$

$$\begin{array}{l}
((\mathsf{s}/_{\otimes}\mathsf{n}_1)/_{\star}\mathsf{n}_2)/_{\diamond}\mathsf{s} \quad \mathsf{n}_1 \\
\hline \quad >\mathbf{P} \\
((\mathsf{s}/_{\star}\mathsf{n}_2)/_{\diamond}\mathsf{s})/_{\otimes}\mathsf{n}_1 \\
\hline \quad > \\
(\mathsf{s}/_{\star}\mathsf{n}_2)/_{\diamond}\mathsf{s}
\end{array}$$

(331) $((\mathsf{s}/_{\otimes}\mathsf{n}_1)/_{\star}\mathsf{n}_2)/_{\diamond}\mathsf{s} \quad \mathsf{s}/_{\diamond}\mathsf{n}_3 \quad \mathsf{n}_1 \quad \mathsf{n}_3$

$$\begin{array}{l}
((\mathsf{s}/_{\otimes}\mathsf{n}_1)/_{\star}\mathsf{n}_2)/_{\diamond}\mathsf{s} \quad \mathsf{s}/_{\diamond}\mathsf{n}_3 \quad \mathsf{n}_1 \quad \mathsf{n}_3 \\
\hline \quad >\mathbf{B} \\
((\mathsf{s}/_{\otimes}\mathsf{n}_1)/_{\star}\mathsf{n}_2)/_{\diamond}\mathsf{n}_3 \\
\hline \quad >\mathbf{P} \\
((\mathsf{s}/_{\star}\mathsf{n}_2)/_{\diamond}\mathsf{n}_3)/_{\otimes}\mathsf{n}_1 \\
\hline \quad > \\
(\mathsf{s}/_{\star}\mathsf{n}_2)/_{\diamond}\mathsf{n}_3 \\
\hline \quad > \\
\mathsf{s}/_{\star}\mathsf{n}_2
\end{array}$$

The fact that an argument of a matrix category can permute with the arguments of a subordinate category parallels the mixing of multisets via composition in MULTISET-CCG, as discussed in §4.2.2. Since MULTISET-CCG has greater than mildly context-sensitive power (Hoffman, 1995), the rule of Generalized Weak Permutation is likely to have similar effects if added to the CCG rule set. Nonetheless, the multi-modal approach would allow the rule to apply only when the categories license it.

Though such possibilities exist, I have not augmented the CCG rule set in this manner for the purposes of this dissertation. Thus, the system that I employ is mildly context-sensitive, like CCG. New rules that increase generative power should be added only with great skepticism and after careful consideration — using the more restricted apparatus of CCG limits the space of possible analyses and therefore gives it greater predictive power than a system which allows more powerful operations. See page 48 for further discussion on the linguistic significance of generative power.

5.6 Summary of Multi-Modal CCG Rules

This section groups all of the MULTI-MODAL CCG rules in one place for easy reference. The rules for composition can of course be (boundedly) generalized.

(332) *Forward application*

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

(333) *Backward application*

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

(334) *Forward type-raising*

$$\mathsf{X} \quad \Rightarrow_{\mathbf{T}} \quad \mathsf{Y}/_i(\mathsf{Y}\backslash_i\mathsf{X}) \qquad (>\mathbf{T})$$

(335) *Backward type-raising*

$\mathsf{X} \;\Rightarrow_{\mathbf{T}}\; \mathsf{Y}\backslash_i(\mathsf{Y}/_i\mathsf{X})$ $(<\mathbf{T})$

(336) *Forward harmonic composition*

$\mathsf{X}/_\diamond\mathsf{Y} \;\; \mathsf{Y}/_\diamond\mathsf{Z} \;\Rightarrow_{\mathbf{B}}\; \mathsf{X}/_\diamond\mathsf{Z}$ $(>\mathbf{B})$

(337) *Backward harmonic composition*

$\mathsf{Y}\backslash_\diamond\mathsf{Z} \;\; \mathsf{X}\backslash_\diamond\mathsf{Y} \;\Rightarrow_{\mathbf{B}}\; \mathsf{X}\backslash_\diamond\mathsf{Z}$ $(<\mathbf{B})$

(338) *Forward crossed composition*

a. $\mathsf{X}/_\times\mathsf{Y} \;\; \mathsf{Y}\backslash_\times\mathsf{Z} \;\Rightarrow_{\mathbf{B}}\; \mathsf{X}\backslash_\times\mathsf{Z}$ $(>\mathbf{B}_\times)$

b. $\mathsf{X}/_\times\mathsf{Y} \;\; \mathsf{Y}\backslash_{\times\triangleright}\mathsf{Z}_{+ANT} \;\Rightarrow_{\mathbf{B}}\; \mathsf{X}\backslash_{\times\triangleright}\mathsf{Z}_{+ANT}$ $(>\mathbf{B}_\times^{ant})$

c. $\mathsf{X}/_{\triangleleft\times}\mathsf{Y} \;\; \mathsf{Y}\backslash_\times\mathsf{Z}_{+ANT} \;\Rightarrow_{\mathbf{B}}\; \mathsf{X}\backslash_\times\mathsf{Z}_{+ANT}$ $(>\mathbf{B}_\times^{ant})$

d. $\mathsf{X}/_{\triangleleft\times}\mathsf{Y} \;\; \mathsf{Y}\backslash_{\times\triangleright}\mathsf{Z}_{+ANT} \;\Rightarrow_{\mathbf{B}}\; \mathsf{X}\backslash_{\times\triangleright}\mathsf{Z}_{+ANT}$ $(>\mathbf{B}_\times^{ant})$

(339) *Backward crossed composition*

a. $\mathsf{Y}/_\times\mathsf{Z} \;\; \mathsf{X}\backslash_\times\mathsf{Y} \;\Rightarrow_{\mathbf{B}}\; \mathsf{X}/_\times\mathsf{Z}$ $(<\mathbf{B}_\times)$

b. $\mathsf{Y}/_\times\mathsf{Z}_{+ANT} \;\; \mathsf{X}\backslash_{\times\triangleright}\mathsf{Y} \;\Rightarrow_{\mathbf{B}}\; \mathsf{X}/_\times\mathsf{Z}_{+ANT}$ $(<\mathbf{B}_\times^{ant})$

c. $\mathsf{Y}/_{\triangleleft\times}\mathsf{Z}_{+ANT} \;\; \mathsf{X}\backslash_\times\mathsf{Y} \;\Rightarrow_{\mathbf{B}}\; \mathsf{X}/_{\triangleleft\times}\mathsf{Z}_{+ANT}$ $(<\mathbf{B}_\times^{ant})$

d. $\mathsf{Y}/_{\triangleleft\times}\mathsf{Z}_{+ANT} \;\; \mathsf{X}\backslash_{\times\triangleright}\mathsf{Y} \;\Rightarrow_{\mathbf{B}}\; \mathsf{X}/_{\triangleleft\times}\mathsf{Z}_{+ANT}$ $(<\mathbf{B}_\times^{ant})$

(340) *Forward harmonic substitution*

$(\mathsf{X}/_\diamond\mathsf{Y})/_\diamond\mathsf{Z} \;\; \mathsf{Y}/_\diamond\mathsf{Z} \;\Rightarrow_{\mathbf{S}}\; \mathsf{X}/_\diamond\mathsf{Z}$ $(>\mathbf{S})$

(341) *Backward harmonic substitution*

$\mathsf{Y}\backslash_\diamond\mathsf{Z} \;\; (\mathsf{X}\backslash_\diamond\mathsf{Y})\backslash_\diamond\mathsf{Z} \;\Rightarrow_{\mathbf{S}}\; \mathsf{X}\backslash_\diamond\mathsf{Z}$ $(<\mathbf{S})$

(342) *Forward crossed substitution*

$(\mathsf{X}/_\times\mathsf{Y})\backslash_\times\mathsf{Z} \;\; \mathsf{Y}\backslash_\times\mathsf{Z} \;\Rightarrow_{\mathbf{S}}\; \mathsf{X}\backslash_\times\mathsf{Z}$ $(>\mathbf{S}_\times)$

(343) *Backward crossed substitution*

$\mathsf{Y}/_\times\mathsf{Z} \;\; (\mathsf{X}\backslash_\times\mathsf{Y})/_\times\mathsf{Z} \;\Rightarrow_{\mathbf{S}}\; \mathsf{X}/_\times\mathsf{Z}$ $(<\mathbf{S}_\times)$

5.7 Discussion of the CTL Basis of Multi-Modal CCG

Each MULTI-MODAL CCG rule has a corresponding CTL structural rule which provides the basis of the modalities that appear on the categories of the rules. I have listed only some of these explicitly thus far in the chapter, and in this section I provide a summary of all the structural rules, discuss other rules which are theorems of the CTL system but are not included in MULTI-MODAL CCG, and show how the hierarchy of modalities Figure 5.1 can be simulated in CTL.

The following rules of *Right Association* and *Left Association* underly the MULTI-MODAL CCG rules of forward harmonic composition and backward harmonic composition.

(344) *Right Association*

$$\frac{(\Delta_a \ \circ_{\diamond} \ (\Delta_b \ \circ_{\diamond} \ \Delta_c)) \vdash \mathsf{X}}{((\Delta_a \ \circ_{\diamond} \ \Delta_b) \ \circ_{\diamond} \ \Delta_c) \vdash \mathsf{X}}\,[RA]$$

(345) *Left Association*

$$\frac{((\Delta_a \ \circ_{\diamond} \ \Delta_b) \ \circ_{\diamond} \ \Delta_c) \vdash \mathsf{X}}{(\Delta_a \ \circ_{\diamond} \ (\Delta_b \ \circ_{\diamond} \ \Delta_c)) \vdash \mathsf{X}}\,[LA]$$

There are several structural rules to handle the behavior of crossed composition and its interaction with antecedent government. The following two groups of rules for *Left Permutation* and *Right Permutation* provide the basis for the forward crossed composition and backward crossed composition rules, respectively.

(346) a. *Left Permutation*

$$\frac{(\Delta_a \ \circ_{\times\!\triangleright} \ (\Delta_b \ \circ_{\triangleleft\!\times} \ \Delta_c)) \vdash \mathsf{X}}{(\Delta_b \ \circ_{\triangleleft\!\times} \ (\Delta_a \ \circ_{\times\!\triangleright} \ \Delta_c)) \vdash \mathsf{X}}\,[LP]$$

b. *Antecedent-Governed Left Permutation I*

$$\frac{(\Delta_a \ \circ_{\times\!\triangleright} \ (\langle\Delta_b\rangle^{ant} \ \circ_{\times\!\triangleright} \ \Delta_c)) \vdash \mathsf{X}}{(\langle\Delta_b\rangle^{ant} \ \circ_{\times\!\triangleright} \ (\Delta_a \ \circ_{\times\!\triangleright} \ \Delta_c)) \vdash \mathsf{X}}\,[AGLP{-}I]$$

c. *Antecedent-Governed Left Permutation II*

$$\frac{(\Delta_a \ \circ_{\triangleleft\!\times} \ (\langle\Delta_b\rangle^{ant} \ \circ_{\times\!\triangleright} \ \Delta_c)) \vdash \mathsf{X}}{(\langle\Delta_b\rangle^{ant} \ \circ_{\times\!\triangleright} \ (\Delta_a \ \circ_{\triangleleft\!\times} \ \Delta_c)) \vdash \mathsf{X}}\,[AGLP{-}II]$$

d. *Antecedent-Governed Left Permutation III*

$$\frac{(\Delta_a \ \circ_{\triangleleft\!\times} \ (\langle\Delta_b\rangle^{ant} \ \circ_{\triangleleft\!\times} \ \Delta_c)) \vdash \mathsf{X}}{(\langle\Delta_b\rangle^{ant} \ \circ_{\triangleleft\!\times} \ (\Delta_a \ \circ_{\triangleleft\!\times} \ \Delta_c)) \vdash \mathsf{X}}\,[AGLP{-}III]$$

(347) a. *Right Permutation*

$$\frac{((\Delta_a \ \circ_{\times\!\triangleright} \ \Delta_b) \ \circ_{\triangleleft\!\times} \ \Delta_c) \vdash \mathsf{X}}{((\Delta_a \ \circ_{\triangleleft\!\times} \ \Delta_c) \ \circ_{\times\!\triangleright} \ \Delta_b) \vdash \mathsf{X}}\,[RP]$$

b. *Antecedent-Governed Right Permutation I*

$$\frac{((\Delta_a \ \circ_{\times\triangleright} \langle\Delta_b\rangle^{ant}) \ \circ_{\times\triangleright} \Delta_c) \vdash \mathsf{X}}{((\Delta_a \ \circ_{\times\triangleright} \Delta_c) \ \circ_{\times\triangleright} \langle\Delta_b\rangle^{ant}) \vdash \mathsf{X}}\,[AGRP{-}I]$$

c. *Antecedent-Governed Right Permutation II*

$$\frac{((\Delta_a \ \circ_{\triangleleft\times} \langle\Delta_b\rangle^{ant}) \ \circ_{\times\triangleright} \Delta_c) \vdash \mathsf{X}}{((\Delta_a \ \circ_{\times\triangleright} \Delta_c) \ \circ_{\triangleleft\times} \langle\Delta_b\rangle^{ant}) \vdash \mathsf{X}}\,[AGRP{-}II]$$

d. *Antecedent-Governed Right Permutation III*

$$\frac{((\Delta_a \ \circ_{\triangleleft\times} \langle\Delta_b\rangle^{ant}) \ \circ_{\triangleleft\times} \Delta_c) \vdash \mathsf{X}}{((\Delta_a \ \circ_{\triangleleft\times} \Delta_c) \ \circ_{\triangleleft\times} \langle\Delta_b\rangle^{ant}) \vdash \mathsf{X}}\,[AGRP{-}III]$$

Kruijff and Baldridge (2000) provide just two substitution rules that interact with the structural rules for composition to simulate the behavior of all the substitution rules (see §5.4). However, the CTL system defined there uses only two modes, as described in §5.1, and the greater discrimination provided by the modes employed in the structural rules above means that we need four structural rules in the present setting, one for each of the MULTI-MODAL CCG substitution rules.

The following rules of *Right Substitution* and *Left Substitution* underly the rules of forward harmonic substitution and backward harmonic substitution, respectively.

(348) *Right Substitution*

$$\frac{((\Delta_a \ \circ_{\diamond} \Delta_c) \ \circ_{\diamond} (\Delta_b \ \circ_{\diamond} \Delta_c)) \ \vdash \mathsf{X}}{((\Delta_a \ \circ_{\diamond} \Delta_b) \ \circ_{\diamond} \Delta_c) \vdash \mathsf{X}}\,[RS]$$

(349) *Left Substitution*

$$\frac{((\Delta_a \ \circ_{\diamond} \Delta_b) \ \circ_{\diamond} (\Delta_a \ \circ_{\diamond} \Delta_c)) \ \vdash \mathsf{X}}{(\Delta_a \ \circ_{\diamond} (\Delta_b \ \circ_{\diamond} \Delta_c)) \vdash \mathsf{X}}\,[LS]$$

The following rules of *Left Crossed Substitution* and *Right Crossed Substitution* provide the basis for the MULTI-MODAL CCG rules of forward crossed substitution and backward crossed substitution, respectively.

(350) *Left Crossed Substitution*

$$\frac{((\Delta_a \ \circ_{\triangleleft\times} \Delta_b) \ \circ_{\times\triangleright} (\Delta_a \ \circ_{\triangleleft\times} \Delta_c)) \ \vdash \mathsf{X}}{(\Delta_a \ \circ_{\triangleleft\times} (\Delta_b \ \circ_{\times\triangleright} \Delta_c)) \vdash \mathsf{X}}\,[LCS]$$

(351) *Right Crossed Substitution*

$$\frac{((\Delta_a \ \circ_{\times\triangleright} \Delta_c) \ \circ_{\triangleleft\times} (\Delta_b \ \circ_{\times\triangleright} \Delta_c)) \ \vdash \mathsf{X}}{((\Delta_a \ \circ_{\triangleleft\times} \Delta_b) \ \circ_{\times\triangleright} \Delta_c) \vdash \mathsf{X}}\,[RCS]$$

With all of the above structural rules, we can of course generate more rules than just those used in MULTI-MODAL CCG. For example, the Division rule (352) (Lambek, 1958) follows from just the base logic and the rule of Right Association, as shown in (353).

(352) *Forward division*

$$\mathsf{X}/_{\diamond}\mathsf{Y} \;\Rightarrow\; (\mathsf{X}/_{\diamond}\mathsf{Z})/_{\diamond}(\mathsf{Y}/_{\diamond}\mathsf{Z}) \qquad (\mathit{Division})$$

(353)

$$\dfrac{\dfrac{\dfrac{\dfrac{\Gamma \vdash \mathsf{X}/_{\diamond}\mathsf{Y} \qquad \dfrac{[x_2 \vdash \mathsf{Y}/_{\diamond}\mathsf{Z}]^{\ddagger} \qquad [x_1 \vdash \mathsf{Z}]^{\dagger}}{(x_2 \;\circ_{\diamond}\; x_1) \vdash \mathsf{Y}}\,[/_{\diamond}E]}{(\Gamma \;\circ_{\diamond}\; (x_2 \;\circ_{\diamond}\; x_1)) \vdash \mathsf{X}}\,[/_{\diamond}E]}{((\Gamma \;\circ_{\diamond}\; x_2) \;\circ_{\diamond}\; x_1) \vdash \mathsf{X}}\,[RA]}{(\Gamma \;\circ_{\diamond}\; x_2) \vdash \mathsf{X}/_{\diamond}\mathsf{Z}}\,[/_{\diamond}I]^{\dagger}}{\Gamma \vdash (\mathsf{X}/_{\diamond}\mathsf{Z})/_{\diamond}(\mathsf{Y}/_{\diamond}\mathsf{Z})}\,[/_{\diamond}I]^{\ddagger}$$

There are thus further rules which are theorems of the base logic and the structural rules that I have defined, but they have not been utilized in MULTI-MODAL CCG so that it retains the combinatory basis assumed by CCG and because it is not immediately clear that they are even necessary. However, the issue of including further rules which are theorems of the CTL basis of MULTI-MODAL CCG merits future investigation.

The hierarchy of Figure 5.1 which I have utilized to define the modalities is not a standardly assumed device by CTL practitioners, but it can be trivially simulated in a CTL system. The most obvious way to do so would be to use *linkage* rules (Hepple, 1997) that relate a more powerful modality to its less powerful subtypes. For example, we could express the fact that $\triangleright$ is both associative and permutative with the following two linkage rules:

(354) a. $\dfrac{(\Delta_a \;\circ_{\triangleright}\; \Delta_b) \vdash \mathsf{X}}{(\Delta_a \;\circ_{\diamond}\; \Delta_b) \vdash \mathsf{X}}$

b. $\dfrac{(\Delta_a \;\circ_{\triangleright}\; \Delta_b) \vdash \mathsf{X}}{(\Delta_a \;\circ_{\times\triangleright}\; \Delta_b) \vdash \mathsf{X}}$

The resulting structures can then be used as input to the structural rules needed to back up the MULTI-MODAL CCG rules, such as (195). However, this simulation runs into a problem: once the linkage rule has applied, the resulting modality is either associative or permutative, and can no longer be both for future purposes. For example, this would mean that the result of applying (195) in a proof that composes the categories $\mathsf{a/b}$ and $\mathsf{b/c}$ would be $\mathsf{a/_{\diamond}c}$ and not $\mathsf{a/c}$.

A strategy which gives the desired result is to simply enumerate the structural rules so that in addition to (195), we have the following:

(355) $$\frac{(\Delta_a \circ_{\triangleright} (\Delta_b \circ_{\triangleright} \Delta_c)) \vdash \mathsf{X}}{((\Delta_a \circ_{\triangleright} \Delta_b) \circ_{\triangleright} \Delta_c) \vdash \mathsf{X}}$$

(356) $$\frac{(\Delta_a \circ_{\diamond} (\Delta_b \circ_{\triangleright} \Delta_c)) \vdash \mathsf{X}}{((\Delta_a \circ_{\diamond} \Delta_b) \circ_{\triangleright} \Delta_c) \vdash \mathsf{X}}$$

(357) $$\frac{(\Delta_a \circ_{\triangleright} (\Delta_b \circ_{\diamond} \Delta_c)) \vdash \mathsf{X}}{((\Delta_a \circ_{\triangleright} \Delta_b) \circ_{\diamond} \Delta_c) \vdash \mathsf{X}}$$

Under this schema, the set of rules needed to cover the modality $\cdot$ is quite large indeed, but it should be a trivial matter to create them automatically from the hierarchy in Figure 5.1.

5.8 Discussion

The resource-sensitive approach permits the grammatical system to make more discriminating use of the linguistic signs that it must combine. Rather than restricting and banning the rules of the CCG system, we can instead *engineer* categories in a manner that makes them applicable in only the contexts that we require them to be active. This has very real implications for how linguistic analyses are created on the basis of the system since it is quite common in CCG analyses to restrict or even ban rules. A further example is the analysis of Tzotzil by Trechsel (2000), who uses the feature ±FC to restrict the rule of forward composition in Tzotzil. The use of modes to define different slash types allows us to model these constraints in the categories of the lexicon, without involving the CCG rule set or *ad hoc* features.

This strategy allows me to dispense with the parameterized view of the CCG rule component advocated by Steedman throughout the history of CCG. This move brings the MULTI-MODAL CCG formulation more in line with the perspective predominant in TAG that the rules of the grammatical system can always be utilized. We could nonetheless consider a version of MULTI-MODAL CCG that allows different rules to be created using the original combinatory rules as schemas which can be refined with a different set of modalities with differing side-effects. This latter approach would be more in line with the view on the parametricity of rules in CTL.

Kruijff (2001) provides an architecture for reducing the overall parametricity while allowing variation in the rule base by providing multilingual networks of structural rule packages, and a similar architecture could be envisioned for MULTI-MODAL CCG.

As an example of a further potential rule that I have not exploited here, we could consider adding a version of harmonic forward composition that results in a category marked for antecedent government:

(358) $\mathsf{X}/_{\diamond}\mathsf{Y} \quad \mathsf{Y}/_{\times}\mathsf{Z}_{+ANT} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X}/_{\times}\mathsf{Z}_{+ANT}$

The slashes would thus remain harmonic, but the modalities would not permit the combination to go forward without inducing the +ANT value.

The set of modalities I have employed here is small, yet it is quite effective in handling a wide range of data from English and Dutch in this chapter and Turkish, Tagalog, and Toba Batak in later chapters. It would nonetheless be particularly interesting to investigate a set of modalities based on heads and dependents such as those proposed by Moortgat and Morrill (1991), Hepple (1997), and Kruijff (2001). It could also be interesting in this respect to consider modalities which would allow composition into other categories but block composition into their own category, or vice versa.

CTL is a very attractive system, but there are several reasons why I do not use it outright. Methodologically speaking, I believe that CCG allows one to concentrate more on the actual linguistics and less on minute modeling of structural control. As shown at the beginning of this chapter, CCG collapses CTL proofs into hypothesis-free derivations, and this is a good thing since there are fewer things to keep track of for a linguist who is creating or reading an analysis. Theoretically speaking, the fact that CCG's competence grammar and performance grammar are one and the same is appealing and distinctive. CTL grammars are normally transformed into some other form for efficient parsing purposes (Hepple, 1999). Also, CCG makes stronger claims about what is linguistically possible by essentially handcuffing itself with limited generative capacity. MULTI-MODAL CCG has equivalent generative power to CCG, and thus these properties hold for it as well.

One way of looking at the relationship between CTL and MULTI-MODAL CCG is that the CCG system represents the spelling out of a given CTL system into a string calculus. Under this view, the proofs available in CTL would appear as hypothesis-free deductive rules in CCG. Although this would lead to rules that do not correspond to any of the combinators and therefore not really be CCG, the relationship between MULTI-MODAL CCG and the CTL system behind it could point to a way of transforming CTL grammars into CCG-based grammars for parsing.

Chapter 6

A Restricted Approach for Argument Scrambling

Chapter 4 discussed scrambling in Turkish and Tagalog and explicated some of the main issues that arise in providing a categorial account of the phenomenon. Some possible approaches were explored within the standard CCG framework, including lexical ambiguity and the use of crossing composition, but these approaches were rejected for being either *ad hoc* or inadequate. The MULTISET-CCG extension of CCG (Hoffman, 1995) provided answers to these problems, but at the cost of incurring extra generative capacity. This chapter extends the multi-modal formulation of CCG with multisets in a way that keeps the system mildly context-sensitive. This chapter thus completes the formal developments needed to support accounts of the linguistic phenomena discussed in Chapter 4 with greater explanatory adequacy. Demonstrations of its adequacy are given for local scrambling in §6.1, long distance scrambling in §6.2 and for extraction asymmetries in Chapter 7.

6.1 Local Scrambling in Multi-Modal CCG

As discussed in Chapter 2 and Chapter 4, MULTISET-CCG incorporates multisets into the rules and categories of CCG in order to directly capture scrambled word orders with a single category. However, along with Rambow's highly related Vector-TAG, MULTISET-CCG attains higher generative power than CCG and other mildly context-sensitive formalisms (Vijay-Shanker and Weir, 1994) such as TAG.

This section augments MULTI-MODAL CCG with multisets, providing definitions for categories and combinatory rules and demonstrating its application to local scrambling in Turkish. Like MULTISET-CCG, MULTI-MODAL CCG can accept scrambled orders with a single category assignment. However, unlike MULTISET-CCG, MULTI-MODAL CCG categories and rules retain a much closer relationship to their standard CCG counterparts. With respect to scrambling, MULTI-MODAL CCG is thus a *de-extension* of MULTISET-CCG; however, it also incorporates the developments of Chapter 5 to permit the use of modally specified categories which license certain combinatory rules but not others.

Based on the definitions of categories for MULTISET-CCG given in Definition 2 and for CTL in Definition 3, the categories of MULTI-MODAL CCG are defined as follows.

Definition 4 (Multi-Modal CCG Categories). *Given $\mathcal{A}$, a finite set of atomic categories and $\mathcal{M}$, a finite set of modalities, the set of categories $\mathcal{C}$ is defined as:*

- $\mathcal{A} \subseteq \mathcal{C}$
- $\mathsf{X_0}\{|^{1}_{\mathsf{m_1}}\mathsf{X_1}, \ldots, |^{\mathsf{n}}_{\mathsf{m_n}}\mathsf{X_n}\} \in \mathcal{C}$,
  if $\mathsf{X_i} \in \mathcal{C}$, $|^{\mathsf{i}} \in \{\backslash, /, |\}$, *and* $\mathsf{m_i} \in \mathcal{M}$, *for* $1 \leq i \leq n$

I assume that $\mathcal{A}$ is defined by the hierarchy given in Figure 3.1. The set of modalities $\mathcal{M}$ is that used in Chapter 5 and shown in Figure 5.1. As before, I will use X, Y and Z as variables for categories and the Greek letters α, β, and γ as variables for sets.

In order to simplify the presentation of MULTI-MODAL CCG derivations, I use two notational conventions:

Convention 11 (Common Slash Types). *When all of the arguments in a multiset have the same slash type, the slash is placed before the entire multiset instead of before the individual arguments. For example,* $\mathsf{s}\{\backslash_{\times}\mathsf{np}_{nom}, \backslash_{\times}\mathsf{np}_{acc}\}$ *is written as* $\mathsf{s}\backslash_{\times}\{\mathsf{np}_{nom}, \mathsf{np}_{acc}\}$. *Also, a multiset or dollar variable preceded by a slash type is constrained so that all members of the multiset or schematized list are compatible with that slash type. Moreover, feature values placed on a multiset or dollar variable indicate that all member categories must be compatible with the feature.*

Convention 12 (Singleton set abbreviation). *Singleton sets containing only a single category are written without brackets; for example,* $\mathsf{s}\{\backslash_{\diamond}\mathsf{n}\}$ *is abbreviated as* $\mathsf{s}\backslash_{\diamond}\mathsf{n}$ *and* $\mathsf{s}\{\backslash_{\triangleright}(\mathsf{s}\{\backslash\mathsf{n}\})\}$ *as* $\mathsf{s}\backslash_{\triangleright}(\mathsf{s}\backslash\mathsf{n})$.

With this convention, a grammar which has only singleton multisets reduces in appearance (as well as functionality) to the system introduced in Chapter 5.

The application rules of MULTI-MODAL CCG are basically modalized variants of those of MULTISET-CCG and appear as follows:

(359) *Multiset-based functional application*

a. $\mathsf{X}(\alpha \uplus \{/_\star \mathsf{Y}\}) \quad \mathsf{Y} \quad \Rightarrow \quad \mathsf{X}\alpha$ (>)

b. $\mathsf{Y} \quad \mathsf{X}(\alpha \uplus \{\backslash_\star \mathsf{Y}\}) \quad \Rightarrow \quad \mathsf{X}\alpha$ (<)

I use multiset union '$\uplus$' in the rule definitions instead of regular set union '$\cup$' to ensure that, for example, $\mathsf{Y} \notin \alpha$, which would lead to the possibility of α containing unboundedly many copies of Y.[1] The set α in the following rules can be empty, and, following Hoffman (1995), we assume a general cleanup rule that removes empty set arguments, so that the result category $\mathsf{X}\emptyset$ reduces to just X.

With rules defined in this manner, the use of multiset categories amounts to an extremely local manner of licensing associativity and permutativity. The backward application rule (359b) in conjunction with a verbal category which puts two or more arguments in the same set handles local scrambling in a manner very similar to MULTISET-CCG. For example, we can assign the Turkish transitive verb *okuyor* 'read' the single category (361a), which permits derivations for the two variants of (153), repeated here as (360), while obtaining a common logical form.

(360) a. Ayse kitabi okuyor
Ayse-NOM book-ACC read-PROG

b. Kitabi Ayse okuyor

Ayse reads the book.

(361) a. *okuyor* $\vdash \mathsf{s}_i\backslash_\cdot\{\mathsf{np}_{nom}, \mathsf{np}_{acc}\}$

b. *Ayse* $\vdash \mathsf{np}_{nom}$

c. *kitabi* $\vdash \mathsf{np}_{acc}$

Equipped with the mini-lexicon (361) and the rule of backward application (359b), we can give the following derivations of (153a) and (153b).

(362) a.

$$\dfrac{\dfrac{Ayse}{\mathsf{np}_{nom}} \quad \dfrac{\dfrac{kitabi}{\mathsf{np}_{acc}} \quad \dfrac{okuyor}{\mathsf{s}_i\backslash_\cdot\{\mathsf{np}_{nom}, \mathsf{np}_{acc}\}}}{\mathsf{s}_i\backslash_\cdot\mathsf{np}_{nom}}<}{\mathsf{s}_i}<$$

[1] I am grateful to Geert-Jan Kruijff for pointing this out.

b.
$$
\begin{array}{ccc}
\textit{kitabi} & \textit{Ayse} & \textit{okuyor} \\
\hline
\mathsf{np}_{acc} & \mathsf{np}_{nom} & \mathsf{s}_i \backslash_{\cdot} \{\mathsf{np}_{nom}, \mathsf{np}_{acc}\} \\
\cline{2-3}
 & \multicolumn{2}{c}{\mathsf{s}_i \backslash_{\cdot} \mathsf{np}_{acc}} \; < \\
\hline
\multicolumn{3}{c}{\mathsf{s}_i} \; <
\end{array}
$$

In order to derive argument cluster coordinations such as those in (160), repeated here as (363), it is necessary to bring in the rules for type-raising and harmonic composition.

(363) a. Ayse kitabi, Fatma da gazeteyi okuyor
Ayse book-Acc, Fatma too newspaper-Acc read-Prog
Ayse is reading the book, and Fatma the newspaper.

b. Kitabi Ayse, gazeteyi de Fatma okuyor
book-Acc Ayse, newspaper-Acc too Fatma read-Prog
As for the book, Ayse is reading it, and the newspaper, Fatma.

Type-raising is essentially the same as the previous modal formulation (196) except that it obeys the definition of categories given in Definition 4.

(364) *Multiset-based type-raising*

a. $\mathsf{X} \Rightarrow_{\mathbf{T}} \mathsf{Y}\{/_i(\mathsf{Y}\{\backslash_i \mathsf{X}\})\}$ $(>\mathbf{T})$

b. $\mathsf{X} \Rightarrow_{\mathbf{T}} \mathsf{Y}\{\backslash_i(\mathsf{Y}\{/_i \mathsf{X}\})\}$ $(<\mathbf{T})$

where $i \in \mathcal{M}$

Of course, with the abbreviations given in Convention 11 and Convention 12, these rules reduce to exactly the same form as (196), so we never actually have to work in practice with the rather unreadable result categories seen in the above rules. Furthermore, the abbreviation given in Convention 4 (page 28) for type-raised categories is still in effect.

The rules for harmonic composition require that all of the members of the set β have the harmonic slash type, a factor which is crucial for keeping resource-management in the multiset based formulation robust.

(365) *Multiset-based harmonic composition*

a. $\mathsf{X}(\alpha \uplus \{/_\diamond \mathsf{Y}\}) \;\; \mathsf{Y}/_\diamond\beta \Rightarrow \mathsf{X}\alpha/_\diamond\beta$ $(>\mathbf{B})$

b. $\mathsf{Y}\backslash_\diamond\beta \;\; \mathsf{X}(\alpha \uplus \{\backslash_\diamond \mathsf{Y}\}) \Rightarrow \mathsf{X}\alpha\backslash_\diamond\beta$ $(<\mathbf{B})$

With the rules of forward harmonic composition and forward type-raising as defined above, we can derive argument cluster coordinations in the manner given in Steedman (1985) and Dowty (1988).

(366) 'Ayse is reading the book, and Fatma the newspaper.'

$$
\begin{array}{cccccc}
\textit{Ayse} & \textit{kitabi} & , & \textit{Fatma} & \textit{gazeteyi} & \textit{okuyor} \\
\hline
\mathsf{np}_n & \mathsf{np}_a & \mathsf{X}\backslash_{\star}\mathsf{X}/_{\star}\mathsf{X} & \mathsf{np}_n & \mathsf{np}_a & \mathsf{s}_i\backslash_{\cdot}\{\mathsf{np}_n, \mathsf{np}_a\} \\
\xrightarrow{\quad}{>}\mathbf{T} & \xrightarrow{\quad}{>}\mathbf{T} & & \xrightarrow{\quad}{>}\mathbf{T} & \xrightarrow{\quad}{>}\mathbf{T} & \\
\ddot{\mathsf{s}}_i/_{\cdot}(\ddot{\mathsf{s}}_i\backslash_{\cdot}\mathsf{np}_n) & (\ddot{\mathsf{s}}_i\backslash_{\cdot}\mathsf{np}_n)/_{\cdot}((\ddot{\mathsf{s}}_i\backslash_{\cdot}\mathsf{np}_n)\backslash_{\cdot}\mathsf{np}_a) & & \mathsf{np}_n^{\uparrow} & \mathsf{np}_a^{\uparrow} & \\
\multicolumn{2}{c}{\xrightarrow{\qquad\qquad}{>}\mathbf{B}} & & \multicolumn{2}{c}{\xrightarrow{\qquad}{>}\mathbf{B}} & \\
\multicolumn{2}{c}{\ddot{\mathsf{s}}_i/_{\cdot}((\ddot{\mathsf{s}}_i\backslash_{\cdot}\mathsf{np}_n)\backslash_{\cdot}\mathsf{np}_a)} & & \multicolumn{2}{c}{\ddot{\mathsf{s}}_i/_{\cdot}((\ddot{\mathsf{s}}_i\backslash_{\cdot}\mathsf{np}_n)\backslash_{\cdot}\mathsf{np}_a)} & \\
\multicolumn{5}{c}{\xrightarrow{\qquad\qquad\qquad\qquad}{<}\Phi{>}} & \\
\multicolumn{5}{c}{\ddot{\mathsf{s}}_i/_{\cdot}((\ddot{\mathsf{s}}_i\backslash_{\cdot}\mathsf{np}_n)\backslash_{\cdot}\mathsf{np}_a)} & \\
\multicolumn{6}{c}{\xrightarrow{\qquad\qquad\qquad\qquad\qquad}{>}} \\
\multicolumn{6}{c}{\ddot{\mathsf{s}}_i}
\end{array}
$$

Note that I have abbreviated the type-raised categories on the right conjunct. Also, notice that the last step requires the categories $(\ddot{\mathsf{s}}_i\backslash_{\cdot}\mathsf{np}_n)\backslash_{\cdot}\mathsf{np}_a$ and $\mathsf{s}_i\backslash_{\cdot}\{\mathsf{np}_n, \mathsf{np}_a\}$ to unify. Unification for multiset categories is thus defined so that a flexible category like the latter can unify with more rigid categories like the former which specify one of the orders made possible by the flexible one.

As was the case with English, MULTI-MODAL CCG allows the grammar of Turkish to cast off rule restrictions such as that which Hoffman places on backward composition (172) to avoid overgenerations such as (170c), repeated here as (367).

(367) *Siyah geldi kedi.
black come-PAST cat
(for: *The black cat came in.*)

In MULTI-MODAL CCG the rules are left untouched, and the ungrammatical sentence is blocked by assigning *siyah* a category which cannot serve as input to either of the crossed composition rules (369):

(368) *siyah* $\vdash \mathsf{n}_x/_{\diamond}\mathsf{n}_x$

(369) *Multiset-based crossed composition rules*

a. $\mathsf{X}(\alpha \uplus \{/_{\times}\mathsf{Y}\}) \quad \mathsf{Y}\backslash_{\times}\beta \quad \Rightarrow \quad \mathsf{X}\alpha\backslash_{\times}\beta$ $\quad (>\mathbf{B}_{\times})$

b. $\mathsf{Y}/_{\times}\beta \quad \mathsf{X}(\alpha \uplus \{\backslash_{\times}\mathsf{Y}\}) \quad \Rightarrow \quad \mathsf{X}\alpha/_{\times}\beta$ $\quad (<\mathbf{B}_{\times})$

The fact that it is possible to constrain the arguments of a set to conform to a common slash type – in contrast to MULTISET-CCG – is crucial for permitting a distinction between harmonic and crossed composition rules. Note that I have omitted the rules for antecedent government parallel to (235b-d) and (223b-d), but they should be assumed to be active in the grammar.

The ability to specify nominally rooted categories which *do* allow permutation of their arguments is a necessary degree of freedom. Hoffman (1995) gives the following data regarding discontinuous possessive noun phrases in Turkish (Hoffman's (30)).

(370) a. Ben [evin kapısını] boyadim.
I [house-GEN door-POSS-ACC] paint-PAST
I painted the house's door.

b. Ben [kapısınıevin] boyadim.
I [door-POSS-ACC house-GEN] paint-PAST
I painted its, the house's, door.

c. Evin$_i$ ben [t$_i$ kapısını] boyadim.
house-GEN$_i$ I [t$_i$ door-POSS-ACC] paint-PAST
As for the house, I painted its door.

d. Ben [t$_i$ kapısını] boyadim evin$_i$.
I [t$_i$ door-POSS-ACC] paint-PAST house-GEN$_i$
I painted its door, the house's.

The lexical entry for *kapısını* is given in (371). It permits not only the basic orders in (370a,b), but also the discontinuous orders of (370c,d). For example, the derivation of (370d) is given in (372).

(371) *kapısını* $\vdash \mathsf{np}_{acc}|.\mathsf{np}_{gen}$

(372)

$$\cfrac{\cfrac{\textit{ben}}{\mathsf{np}_{nom}} \quad \cfrac{\cfrac{\cfrac{\textit{kapısını}}{\mathsf{np}_{acc}|.\mathsf{np}_{gen}} \quad \cfrac{\textit{boyadim}}{\mathsf{s}_i\backslash.\{\mathsf{np}_{nom}, \mathsf{np}_{acc}\}}}{(\mathsf{s}_i\backslash.\mathsf{np}_{nom})|.\mathsf{np}_{gen}}{<}\mathbf{B}_\times \quad \cfrac{\textit{evin}}{\mathsf{np}_{gen}}}{\mathsf{s}_i\backslash.\mathsf{np}_{nom}}{>}}{\mathsf{s}_i}{<}$$

6.2 Long Distance Scrambling in Multi-Modal CCG

In this section, I demonstrate that MULTI-MODAL CCG is capable of capturing bounded long distance scrambling. Since MULTI-MODAL CCG is mildly context-sensitive (see §6.5), I thus counter and Rambow's claims that we need greater generative power to provide a satifactory account of long distance scrambling. This relates to the analysis of Joshi *et al.* (2000), which demonstrates a formulation of Tree-Adjoining Grammar called Tree-local Multi-component TAG that is able to handle the extent of long distance scrambling fow which native speaker judgements can be reliably obtained.

Unsurprisingly, the permutation-inducing crossed composition rules make it possible to capture long distance scrambled orders. For example, to derive (165b), repeated here as (373), the subject *Fatma* of the matrix verb must cross compose into the verbal cluster in order for the derivation to proceed, as shown in (374).

(373) *Kitabi*$_i$ Fatma [Esra'nın t_i okudugunu] biliyor.
book-ACC$_i$ Fatma [Esra-GEN t_i read-GER-ACC] know-PROG
As for the book, Fatma knows that Esra read it.

(374) *Kitabi* *Fatma* *Esra'nın* *okudugunu* *biliyor*

$$
\begin{array}{ccccc}
\underline{\textit{Kitabi}} & \underline{\textit{Fatma}} & \underline{\textit{Esra'nın}} & \underline{\textit{okudugunu}} & \underline{\textit{biliyor}} \\
\mathsf{np}_a & \mathsf{np}_n & \mathsf{np}_g & \mathsf{s}_{v\,acc}\backslash_{\cdot}\{\mathsf{np}_g, \mathsf{np}_a\} & \mathsf{s}_i\backslash_{\cdot}\{\mathsf{np}_n, \mathsf{s}_{v\,acc}\} \\
 & \overline{\ddot{\mathsf{s}}_i/_{\cdot}(\ddot{\mathsf{s}}_i\backslash_{\cdot}\mathsf{np}_n)}^{>\mathbf{T}} & \multicolumn{2}{c}{\overline{\mathsf{s}_{v\,acc}\backslash_{\cdot}\mathsf{np}_a}^{<}} & \\
 & & \multicolumn{3}{c}{\overline{(\mathsf{s}_i\backslash_{\cdot}\mathsf{np}_n)\backslash_{\cdot}\mathsf{np}_a}^{<\mathbf{B}}} \\
 & \multicolumn{4}{c}{\overline{\ddot{\mathsf{s}}_i\backslash_{\cdot}\mathsf{np}_a}^{>\mathbf{B}_\times}} \\
\multicolumn{5}{c}{\overline{\ddot{\mathsf{s}}_i}^{<}}
\end{array}
$$

Arguments may also scramble to the right, as evidenced by (165c) and (168), repeated here as (375) and (376), respectively.

(375) Fatma [Esra'nın t_i okudugunu] biliyor *kitabi*$_i$.
Fatma [Esra-GEN t_i read-GER-ACC] know-PROG *book*-ACC$_i$
Fatma knows that Esra read it, the book.

(376) Bu kitabi$_j$ Fatma [t_i [t_j okumak] istedigimi] biliyor benim$_i$
This book-ACC$_j$ Fatma [t_i [t_j read-INF] want-GER-ACC know-PROG I-GEN$_i$
As for this book, Fatma knows that I want to read it.

Because I have assumed that the categories of Turkish verbs seek their arguments to the left, something more is needed to allow arguments to appear on the right. Following Bozsahin (2002), I assume the contraposed category given in (377) for such arguments.

(377) $\ddot{\mathsf{s}}_c\backslash_{\diamond}(\mathsf{s}\backslash_{\cdot}\mathsf{np}_{+ANT})$

With this category, an MULTI-MODAL CCG analysis of (376) proceeds as follows, again using crossed composition to allow the matrix subject to dig into the verbal cluster and thereby let the verb reach the other arguments.

(378)

$$
\begin{array}{cccccc}
\textit{Bu kitabi} & \textit{Fatma} & \textit{okumak} & \textit{istedigimi} & \textit{biliyor} & \textit{benim} \\
\hline
\mathsf{np}_a & \mathsf{np}_n & \mathsf{s}_{v\,inf}\backslash_{\cdot}\{\mathsf{np}_g, \mathsf{np}_a\} & \mathsf{s}_{v\,acc}\backslash_{\cdot}\mathsf{s}_{v\,inf} & \mathsf{s}_i\backslash_{\cdot}\{\mathsf{np}_n, \mathsf{s}_{v\,acc}\} & \ddot{\mathsf{s}}_c\backslash_{\diamond}(\mathsf{s}\backslash_{\cdot}\mathsf{np}_g) \\
 & \overline{\mathsf{np}_n^{\uparrow}}^{>\mathbf{T}} & \multicolumn{2}{c}{\overline{\mathsf{s}_{v\,acc}\backslash_{\cdot}\{\mathsf{np}_g, \mathsf{np}_a\}}^{<\mathbf{B}}} & & \\
 & & \multicolumn{3}{c}{\overline{(\mathsf{s}_i\backslash_{\cdot}\mathsf{np}_n)\backslash_{\cdot}\{\mathsf{np}_g, \mathsf{np}_a\}}^{<\mathbf{B}}} & \\
 & \multicolumn{4}{c}{\overline{\ddot{\mathsf{s}}_i\backslash_{\cdot}\{\mathsf{np}_g, \mathsf{np}_a\}}^{>\mathbf{B}^2_{\times}}} & \\
\multicolumn{5}{c}{\overline{\mathsf{s}_i\backslash_{\cdot}\mathsf{np}_g}^{<}} & \\
\multicolumn{6}{c}{\overline{\ddot{\mathsf{s}}_c}^{<}}
\end{array}
$$

In addition to having recourse to modalized categories to control access to permutative rules such as forward crossed composition and backward crossed composition, the restricted power of the MULTI-MODAL CCG system saves it from overgenerating in certain cases, such as the following (Hoffman's example (50)):

(379) *Ahmet [benim t_i okudugumu] Fatma'ya *kitabı*$_i$ soyledi.
*Ahmet [I-GEN t_i read-GER-ACC] Fatma-DAT *book*-ACC say
(for: Ahmet told Fatma that I read the BOOK.)

(380)

$$
\begin{array}{cccccc}
\textit{*Ahmet} & \textit{benim} & \textit{okudugumu} & \textit{Fatma'ya} & \textit{kitabi} & \textit{soyledi} \\
\hline
\mathsf{np}_n & \mathsf{np}_g & \mathsf{s}_{v\,ger}\backslash_{\cdot}\{\mathsf{np}_g, \mathsf{np}_a\} & \mathsf{np}_{dat} & \mathsf{np}_a & \mathsf{s}_i\backslash_{\cdot}\{\mathsf{np}_n, \mathsf{np}_{dat}, \mathsf{s}_{v\,ger}\} \\
 & \multicolumn{2}{c}{\overline{\mathsf{s}_{v\,ger}\backslash_{\cdot}\mathsf{np}_a}^{<}} & & &
\end{array}
$$

Hoffman (1995) shows how the MULTISET-CCG rules allow *kitabi* to type-raise and compose with *soyledi* and then progressively apply to the elements to their left. She blocks the derivation by placing restrictions on the composition rules. With MULTI-MODAL CCG, the derivation is stuck at the point indicated in derivation (380) since there are no operations available which allow *okudugumu* to reach *kitabi* over *Fatma'ya* nor allow *soyledi* to reach *Fatma'ya*.

MULTI-MODAL CCG also does not accept sentences of the form SO & OSV, whereas MULTISET-CCG does. Hoffman states that such sequences are not grammatical in Turkish, but that they have been reported to be grammatical in other languages, such as German. She suggests that the unacceptability of such coordinations in Turkish may be due to pragmatic factors. It is not clear then, whether being able to derive a coordination like SO & OSV is desirable, but if such strings are ultimately determined to need an analysis, we can do it through the decomposition rule, as shown by Karamanis (2001), rather than as a core property of the grammar.

One piece of evidence that would favor the more powerful composition rule of MULTISET-CCG over that of MULTI-MODAL CCG would be if it were possible to coordinate noun clusters with a scrambled element inside, such as the following example, modeled after (373), in which the accusative noun phrase has scrambled out of the clause of the subordinate verb (v_2):

(381) $[\mathsf{np}_{acc}\mathsf{np}_{nom}\mathsf{np}_{gen}\&\mathsf{np}_{acc}\mathsf{np}_{nom}\mathsf{np}_{gen}]$ v_2 v_1

Of course, if more power ultimately does prove necessary, we can scale up the rules to work in a manner like Hoffman's while maintaining the resource-sensitivity available under the present proposal. See §6.5 for more discussion on this.

6.3 Multiset Categories in English

Despite the general rigidness of English word order, some analyses can be simplified by placing the arguments of some categories in a multiset. As discussed in §2.3, this strategy can be employed for phrasal verbs to capture alternations such as (71), repeated here as (382).

(382) a. Marcos picked up the ball.

b. Marcos picked the ball up.

This alternation was captured by placing the categories for the **Patient** argument and the particle in a multiset. However, the need for modal control appears again in these contexts, as evidenced by the following adverb placement data, in which we see that adverbs cannot come between the verb and the particle.

(383) a. Marcos picked up quickly the ball that the defender had passed back.

b. *Marcos picked quickly up the ball that the defender had passed back.

(384) a. *Marcos picked quickly the ball up.

b. *Marcos picked the ball quickly up.

In the restrictional setting, the adverb can by default combine with the verb via backward crossed composition in all of these cases, so restrictions would need to be placed on rules in order to capture this distribution. Having different modes of composition provides a much simpler solution: the slash for the particle is non-associative, while that of the **Patient** argument is the usual one assumed for English direct objects.

(385) *picked* $\vdash (\mathsf{s}_i\backslash\mathsf{np})\{/\mathsf{np}, /_{\star}\mathsf{prt}\}$

Because the MULTI-MODAL CCG rule for backward crossed composition (369b) requires all members of the set β to have a slash compatible with $/_{\times}$, the adverb cannot compose with the verb until the particle is consumed.

(386) a. Marcos [picked up]$_{(\mathsf{s}_i\backslash\mathsf{np})/\mathsf{np}}$ [quickly]$_{(\mathsf{s}\backslash\mathsf{np})\backslash(\mathsf{s}\backslash\mathsf{np})}$ the ball... .

b. *Marcos [picked]$_{(\mathsf{s}_i\backslash\mathsf{np})\{/\mathsf{np},/_{\star}\mathsf{prt}\}}$ [quickly]$_{(\mathsf{s}\backslash\mathsf{np})\backslash(\mathsf{s}\backslash\mathsf{np})}$ up the ball... .

(387) a. *Marcos [picked]$_{(\mathsf{s}_i\backslash\mathsf{np})\{/\mathsf{np},/_{\star}\mathsf{prt}\}}$ [quickly]$_{(\mathsf{s}\backslash\mathsf{np})\backslash(\mathsf{s}\backslash\mathsf{np})}$ the ball... up.

b. *Marcos [picked the ball...]$_{(\mathsf{s}_i\backslash\mathsf{np})/_{\star}\mathsf{prt}}$ [quickly]$_{(\mathsf{s}\backslash\mathsf{np})\backslash(\mathsf{s}\backslash\mathsf{np})}$ up.

Another example of where multisets are useful in English are prepositions. The category of prepositions which modify nouns is generally assumed to be $(\mathsf{n}\backslash_{\star}\mathsf{n})/_{\triangleleft}\mathsf{np}$. Such a category must consume its object noun phrase on the right before combining with the noun that it modifies on the left, leading to an apparent problem for providing an analysis of the following coordination:

(388) The fans saw the foul on, and the shot by, their team's most prolific goal scorer.

The only way of combining *foul* and *on* would be to provide an extra category for bare nouns that took the nominal modifier as an argument, i.e. $\mathsf{n}/_{\diamond}(\mathsf{n}\backslash_{\star}\mathsf{n})$. However, we can instead place the arguments of the category for prepositions in a multiset, as shown in (389), which then permits derivations for strings such as *the foul on* as shown in (390).

(389) *on, by* $\vdash \mathsf{n}\{\backslash_{\star}\mathsf{n}, /_{\triangleleft}\mathsf{np}\}$

(390)
$$
\begin{array}{ccc}
\textit{the} & \textit{foul} & \textit{on} \\
\overline{\mathsf{np}/_{\diamond}\mathsf{n}} & \overline{\mathsf{n}} & \overline{\mathsf{n}\{\backslash_{\star}\mathsf{n}, /_{\triangleleft}\mathsf{np}\}} \\
 & \multicolumn{2}{c}{\underline{\hspace{10em}}<} \\
 & \multicolumn{2}{c}{\mathsf{n}/_{\triangleleft}\mathsf{np}} \\
\multicolumn{3}{c}{\underline{\hspace{16em}}>\mathbf{B}} \\
\multicolumn{3}{c}{\mathsf{np}/_{\triangleleft}\mathsf{np}}
\end{array}
$$

Other categories which have two arguments that are found on opposite sides, such as those for coordination and relativization, can likewise utilize a multiset for their arguments.

6.4 Inherent Limits on Scrambling

Rambow (1994) claims that scrambling is doubly unbounded in that there is no bound over which each element can scramble and that there is no bound on the number of elements that can scramble in one sentence. Nonetheless, there are limits on the amount of scrambling which native speakers find acceptable. However, both Rambow and Hoffman claim that the unacceptability of scrambled sentences increases gradually and that there is no particular cut-off point beyond which all orders are ungrammatical (Rambow, 1994; Hoffman, 1995), just as it is the case for center-embedding in English.

(391) a. The player scored.

b. The player [the referee penalized] scored.

c. The player [the referee [the fans jeered] penalized] scored.

d. The player [the referee [the fans [the camera showed ...] jeered] penalized] scored.

Even at just two levels of embedding (391c), sentences become extremely difficult to process and require considerable meta-analysis to pick apart. However, this is assumed to arise not because the competence grammar does not have sufficient power to provide an analysis, but because of processing factors.

Rambow shows that power greater than mild context-sensitivity is necessary to capture unbounded scrambling, and therefore concludes that limits on the amount of scrambling arise due to processing factors, like center-embedding. However, Joshi *et al.* (2000) demonstrate that one of the TAG variants, Tree-local Multicomponent TAG, has the interesting property that it can capture up to two levels of scrambling and no more. This leads to the surprising conclusion that we can assume neither (i) that the competence grammar must deal with the fully unbounded case nor (ii) that processing limitations keep humans from finding exceedingly scrambled sentences as unacceptable.

MULTI-MODAL CCG does not have the same *inherent* restriction – stopping at two levels of scrambling – as Tree-local Multi-component TAG, but it is nonetheless restricted by the limit on level of composition permitted by the grammar. Joshi *et al.* (2000) show that Tree-local Multi-component TAG cannot provide a derivation with the correct dependencies for an abstract string such as (392), where each verb_i subcategorizes for arg_i and verb_{i-1} (except, of course, for verb_4, which subcategorizes only for arg_4).

(392) arg_2 arg_4 arg_3 arg_1 verb_4 verb_3 verb_2 verb_1

MULTI-MODAL CCG (and standard CCG as well) can derive this string if forward crossed composition is bounded at least at three, as shown in (393). I assume categories that are similar to those needed for German, the language from which Joshi *et al.*'s examples are drawn.

(393)

$$
\begin{array}{cccccccc}
\mathit{arg}_2 & \mathit{arg}_4 & \mathit{arg}_3 & \mathit{arg}_1 & \mathit{verb}_4 & \mathit{verb}_3 & \mathit{verb}_2 & \mathit{verb}_1 \\
\hline
\mathsf{np}_2 & \mathsf{np}_4 & \mathsf{np}_3 & \mathsf{np}_1 & \mathsf{s}_4\backslash_{.}\mathsf{np}_4 & (\mathsf{s}_3\backslash_{.}\mathsf{np}_3)\backslash_{.}\mathsf{s}_4 & (\mathsf{s}_2\backslash_{.}\mathsf{np}_2)\backslash_{.}\mathsf{s}_3 & (\mathsf{s}_1\backslash_{.}\mathsf{np}_1)\backslash_{.}\mathsf{s}_2
\end{array}
$$

$$\mathsf{np}_1 \Rightarrow_{>\mathbf{T}} \mathsf{s}_1/_{.}(\mathsf{s}_1\backslash_{.}\mathsf{np}_1)$$

$$(\mathsf{s}_2\backslash_{.}\mathsf{np}_2)\backslash_{.}\mathsf{s}_3 \quad (\mathsf{s}_1\backslash_{.}\mathsf{np}_1)\backslash_{.}\mathsf{s}_2 \Rightarrow_{<\mathbf{B}^2} ((\mathsf{s}_1\backslash_{.}\mathsf{np}_1)\backslash_{.}\mathsf{np}_2)\backslash_{.}\mathsf{s}_3$$

$$(\mathsf{s}_3\backslash_{.}\mathsf{np}_3)\backslash_{.}\mathsf{s}_4 \quad ((\mathsf{s}_1\backslash_{.}\mathsf{np}_1)\backslash_{.}\mathsf{np}_2)\backslash_{.}\mathsf{s}_3 \Rightarrow_{<\mathbf{B}^2} (((\mathsf{s}_1\backslash_{.}\mathsf{np}_1)\backslash_{.}\mathsf{np}_2)\backslash_{.}\mathsf{np}_3)\backslash_{.}\mathsf{s}_4$$

$$\mathsf{s}_4\backslash_{.}\mathsf{np}_4 \quad (((\mathsf{s}_1\backslash_{.}\mathsf{np}_1)\backslash_{.}\mathsf{np}_2)\backslash_{.}\mathsf{np}_3)\backslash_{.}\mathsf{s}_4 \Rightarrow_{<\mathbf{B}} (((\mathsf{s}_1\backslash_{.}\mathsf{np}_1)\backslash_{.}\mathsf{np}_2)\backslash_{.}\mathsf{np}_3)\backslash_{.}\mathsf{n}_4$$

$$\mathsf{s}_1/_{.}(\mathsf{s}_1\backslash_{.}\mathsf{np}_1) \quad (((\mathsf{s}_1\backslash_{.}\mathsf{np}_1)\backslash_{.}\mathsf{np}_2)\backslash_{.}\mathsf{np}_3)\backslash_{.}\mathsf{n}_4 \Rightarrow_{>\mathbf{B}^3_{\times}} ((\mathsf{s}_1\backslash_{.}\mathsf{np}_2)\backslash_{.}\mathsf{np}_3)\backslash_{.}\mathsf{n}_4$$

Though it is possible to permit higher (finite) levels of composition in CCG, it is generally assumed that an upper bound of three is sufficient for natural language grammars (Steedman, 2000b). It thus appears that MULTI-MODAL CCG (also, CCG) provides the generative power necessary to capture the scrambled sentences for which native speaker judgements can be reliably obtained. MULTI-MODAL CCG is not alone in having this property — Kulick (2000) presents a variant of TAG called Segmented Tree Adjoining Grammar that has the same gradient restriction on the level of scrambling as MULTI-MODAL CCG.

This issue is interesting to consider with respect to learning and processing CCG grammars. Instead of looking at the definitions of the rules as the rules which are actually used in processing, we can treat the definitions as schemas which are instantiated to the levels mandated by the learning data during the course of acquiring a lexicon. We might then find a stratification of the use of each kind of rule in which the higher levels of the rule type would be somehow less accessible than the more commonly utilized lower level instantiations of the rule type. Whether this is represented as a last-resort strategy to allow a derivation to proceed or as statistical preferences for the application of different instantiations of the rule, the processing is likely to get more difficult as the level increases. I do not wish to claim that this is the actual state of affairs for human language acquisition and processing, but such a scenario would have the interesting property of explaining the gradual scaling of unacceptability in scrambled sentences discussed by Rambow (1994) and Hoffman (1995).

6.5 Generative Power of Multi-Modal CCG with Multisets

Since this chapter has proposed a change in the way that CCG categories are defined and in the form of the rules that combine them, it is important to consider whether these modification incur an increase in generative capacity. The answer, as with the modifications in the previous chapter, is that it does not.

Clearly, MULTI-MODAL CCG with multisets is at least as powerful as it is without them. By restricting all multisets to contain only one element, we obtain the original system.

To show that MULTI-MODAL CCG with multisets is not more powerful is more involved, but trivial. First, note that we can simulate a multiset category with multiple standard categories to achieve all the variation in orders provided by the multiset category. For example, the category $\mathsf{s}\{/_{\times}\mathsf{n}_1, /_{\diamond}\mathsf{n}_2, /_{\star}\mathsf{n}_3\}$ expands into the follow set of categories:

(394) $$\left\{ \begin{array}{lll} ((\mathsf{s}/_{\times}\mathsf{n}_1)/_{\diamond}\mathsf{n}_2)/_{\star}\mathsf{n}_3, & ((\mathsf{s}/_{\times}\mathsf{n}_1)/_{\star}\mathsf{n}_3)/_{\diamond}\mathsf{n}_2, & ((\mathsf{s}/_{\diamond}\mathsf{n}_2)/_{\times}\mathsf{n}_1)/_{\star}\mathsf{n}_3, \\ ((\mathsf{s}/_{\star}\mathsf{n}_3)/_{\times}\mathsf{n}_1)/_{\diamond}\mathsf{n}_2, & ((\mathsf{s}/_{\diamond}\mathsf{n}_2)/_{\star}\mathsf{n}_3)/_{\times}\mathsf{n}_1, & ((\mathsf{s}/_{\star}\mathsf{n}_3)/_{\diamond}\mathsf{n}_2)/_{\times}\mathsf{n}_1 \end{array} \right\}$$

If we take any lexicon with multisets and expand all of its categories in this manner, we can use the rules which are not based on multisets to combine these categories and thereby simulate every combination that was possible in the multiset system. The two most important aspects of the use of multisets in MULTI-MODAL CCG that make this possible are the following:

1. All arguments of multisets of the non-functor categories in the composition rules must have the same directionality (e.g. $\mathsf{Y}\backslash_{\times}\beta$ of forward crossed composition).

2. Multisets can never increase in size due to the nature of the multiset-based combinatory rules. That is, no combinatory rule joins the multisets of its input categories.

The first point simply guarantees that we can distinguish between harmonic and crossed varieties of the combinatory rules. The second point is more important for generative power and is best illustrated by comparing the forward composition rules of both MULTI-MODAL CCG and MULTISET-CCG:

(395) $\mathsf{X}(\alpha \uplus \{/_{\diamond}\mathsf{Y}\}) \quad \mathsf{Y}/_{\diamond}\beta \quad \Rightarrow \quad \mathsf{X}\alpha/_{\diamond}\beta$ ($>$**B**, MULTI-MODAL CCG)

(396) $\mathsf{X}(\alpha \uplus \{/\mathsf{Y}\}) \quad \mathsf{Y}\beta \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X}(\alpha \uplus \beta)$ ($>$**B**, MULTISET-CCG)

The MULTISET-CCG rule joins the remainder set of the functor category with the set of the argument category. Hoffman (1995) shows that the fact that MULTISET-CCG allows sets to grow in this manner is a major factor in increasing the parsing complexity of the formalism (though MULTISET-CCG is still parsable in polynomial time). MULTI-MODAL CCG avoids this problem and remains mildly context-sensitive, like CCG.

In some ways, the addition of multisets in the manner proposed in this chapter amounts to schematization over a set of rigidly ordered categories, and the permutative effects of this modification are thus only seen in local contexts. The original CCG system already provides the mechanisms necessary for handling long distance scrambling, and MULTI-MODAL CCG uses this same capability for long distance scrambling. Nonetheless, if we did wish to use composition rules that mix multisets such as those that Hoffman proposes, the multi-modal setting will allow us to do so in a controlled fashion. Thus, we might use the mode discussed in §5.5 to permit *clause union* (Evers, 1975) as follows:

(397) *Clause union composition*

$$\mathsf{X}(\alpha \uplus \{/_{\otimes}\mathsf{Y}\}) \quad \mathsf{Y}\beta \quad \Rightarrow \quad \mathsf{X}(\alpha \uplus \beta)$$

This is very much like the HPSG operation of *domain union* of (Reape, 1994) which merges a daughter node's domain into its mother's domain. Adding such a rule will give the overall system the same power as MULTISET-CCG, but we have the major advantage that a given grammar can exploit that power *only* when it *needs* it, as we would expect from our explicit resource-management strategy.

6.6 Summary

This chapter completes the formal modifications which I argue are necessary to handle scrambling phenomena in CCG. It also incorporates the modal approach discussed in Chapter 5 which allows us to keep a tight grip on the permutative capabilities of the system. The next chapter shows how an MULTI-MODAL CCG based analysis can explain the syntactic extraction asymmetries and word order of Tagalog and Toba Batak without conflating the two phenomena as some previous accounts have done.

Chapter 7

Syntactic Extraction Asymmetries in Tagalog and Toba Batak

Chapter 6 completed the formal developments which this dissertation proposes. The multi-modal formulation of CCG provided in Chapter 5 was shown to be desirable for modeling the limitations of CCG rules in particular grammars without placing restrictions or entire bans on any given rule. The augmentations given in Chapter 6 answered the limitations of standard CCG for accounting for scrambling phenomena while showing that scrambling can be handled with a mildly context-sensitive system.

§4.1.1 shows that the subject/object asymmetry of English arises because the grammar of English limits the applicability of the forward crossed composition rule. This limitation is modeled within MULTI-MODAL CCG by assuming that the category $\ddot{\mathsf{s}}_c/_{\diamond}\mathsf{s}_i$ of complementizers such as *that* does not license the use of forward crossed composition, as shown in §5.2.7. Given that English asymmetries stem from this limitation, we can then ask whether any other languages might exhibit asymmetries because of similar limitations on either of the *harmonic* composition rules. This chapter shows that this degree of freedom permitted by the MULTI-MODAL CCG formulation is exactly what is necessary for modeling and explaining the syntactic asymmetries of Tagalog. It furthermore demonstrates that Toba Batak, while being a typologically similar language cousin of Tagalog, exhibits asymmetries because of limitations on crossed composition, like English.

In addition to validating the MULTI-MODAL CCG formulation, the analysis given in this chapter provides the most extensive coverage of Tagalog's famed asymmetries to date, which it achieves while using far more constrained grammatical machinery than the analysis of Nakamura (1998) discussed in §4.1.2. In doing so, the analysis demonstrates how a categorial system can exhibit the configurationality which Guilfoyle *et al.* (1992) claim is necessary for capturing Tagalog asymmetries. However, instead of employing a phrase structural notion, configurationality in the present account refers to the way in which resources are structured in the CTL system that underlies MULTI-MODAL CCG. Combined with the brief discussion of asymmetries in Toba Batak given in §7.2, the analysis furthermore illustrates how the few degrees of freedom permitted by the modes of MULTI-MODAL CCG provide the means to characterize asymmetries cross-linguistically.

7.1 Tagalog Asymmetries

The analysis provided in this section is an extension and evolution of that provided in Baldridge (1998). The specific asymmetries of Tagalog which this section explores are those found in headless relative clause (HRC) formation, relativization, *wh*-extraction, and *ay*-inversion. I begin by providing a discussion and account of Tagalog verbs.

7.1.1 Basic Verbal Categories

Tagalog has a case-marking system that interacts with an intricate voice system, which is briefly described in §4.1.2. Following Kroeger (1993), I assume the following correspondences between the markers and case values:

	NOM	GEN	DAT
Common noun markers	ang	ng	sa
Personal name markers	si	ni	kay

Table 7.1: Morphological expression of case in Tagalog

However, I use this correspondence as a matter of convenience and do not take a stance on the debate of whether Tagalog is a nominative/accusative type language or an ergative/unaccusative one. See Maclachlan and Nakamura (1997) for a discussion of this issue. What matters for the present analysis is that the markers map to some set of *distinct* cases.

Tagalog is a verb-initial language, as we can see in simple intransitive (398a), transitive (398b), and ditransitive (398c) sentences. The **Actor** arguments of Active Voice verbs such as those in (398) receive nominative case.

(398) a. Tumakbo ang lalaki.
ran-AV NOM-man
The man ran.

b. Bumili ang titser ng libro.
buy-AV NOM-teacher GEN-book
The teacher bought a book.

c. Nagbigay ang lalaki ng libro sa babae.
give-AV NOM-man GEN-book DAT-woman
The man gave a book to the woman.

As mentioned in Chapter 4, arguments may scramble within the clause, as evidenced by the alternative order of (398b) given in (399) and the five alternative orders of (398c) given in (152).

(399) Bumili ng libro ang titser.

This variability plus the case marking suggests the following categories for these verbs:

(400) a. *tumakbo* $\vdash \mathsf{s}_i/\mathsf{np}_n$

b. *bumili* $\vdash \mathsf{s}_i\{/_\times\mathsf{np}_{gen}, /\mathsf{np}_n\}$

c. *nagbigay* $\vdash \mathsf{s}_i\{/_\times\mathsf{np}_d, /_\times\mathsf{np}_g, /\mathsf{np}_n\}$

The transitive and ditransitive categories allow their arguments to be consumed in any order to the right, but the nominative argument is given the more powerful associative and permutative slash (/) than the permutative but non-associative slashes of the other arguments ($/_\times$). This differentiation is forced by data revealed later in this chapter and is in fact a major factor in modeling the asymmetries of Tagalog.

As pointed out in §4.1.2, Tagalog verbs can receive different voice affixes which assign nominative case on particular arguments. This process is often called *focussing* or *topicalization* in the Austronesian literature, but due to the ambiguity inherent in both of these terms with other linguistic processes, I will refer to this as *distinguishing*. The data in (398) shows verbs in the Active Voice, which distinguish the syntactic argument corresponding to the **Actor** argument of the verbal predicate. In (401), we find the root verb *bili* 'buy' in its Objective Voice form *binili*, which distinguishes the syntactic argument corresponding to the **Patient** argument.

(401) Binili ng titser ang libro.
buy-OV GEN-teacher NOM-book
A teacher bought the book.

Again, the order of the arguments may be permuted, which leads us to assign the following category to *binili*.

(402) *binili* $\vdash \mathsf{s}_i\{/\mathsf{np}_{nom}, /_{\times}\mathsf{np}_g\}$

It is possible to use an underspecified category for root verbs which then marks the syntactic category of the distinguished dependent with nominative case and the more powerful slash / when the morphological operation of voice affixation occurs. However, this issue is orthogonal to the present analysis, so I do not discuss the details of this process here.

The categories of (400b) and (402) may appear to be essentially the same; however, when the dependents are made explicit as in (403), the difference is clear — the nominative argument of (400b)/(403a) is the **Actor** whereas that of (402)/(403b) is the **Patient**.

(403) a. *bumili* $\vdash \mathsf{s}_i\{/_{\times}\mathsf{np}_g{}^{\mathbf{Pat}}, /\mathsf{np}_n{}^{\mathbf{Act}}\}$

b. *binili* $\vdash \mathsf{s}_i\{/\mathsf{np}_n{}^{\mathbf{Pat}}, /_{\times}\mathsf{np}_g{}^{\mathbf{Act}}\}$

I omit semantic annotations throughout this chapter so that the focus can remain entirely on the syntactic system and note that the categories and derivations given here will return the correct dependencies between elements of Tagalog sentences.

Given these verbal categories, it is now necessary to consider how nominal arguments such as *ang titser* and *ng libro* come to have the categories np_n and np_g, respectively.

7.1.2 Case Markers and Bare Nouns

Determining the internal structure of the noun phrase is crucial to understanding Tagalog's asymmetries. If bare nouns have the category n, then the case markers must be functions with a syntactic type of the form np/n. Upon closer examination, however, there is compelling evidence that nouns bear the syntactic type of an intransitive verb, transparent to the semantic type $<\mathsf{e},\mathsf{t}>$ which Montague (1973) assumes for bare nouns in English. The case markers must then receive categories which take intransitive verbs as arguments and produce case marked nouns.

Our first indication that nouns have the syntactic type of a predicate comes from nominal predicative sentences and adjectival predicative sentences. Tagalog does not have a copula, as can be seen in (404).

(404) a. Artista ang babae.
actress Nom woman
The woman is an actress.

b. Matalino ang babae.
smart Nom woman
The woman is smart.

Adhering to the Principle of Adjacency (Steedman, 2000b) bars us from appealing to empty categories such as null copulas, so if we are to derive a sentence from (404a), one of the lexical items must be a function into s. A first temptation would be to say that *ang* is actually a copula, and that in addition to a standard prenominal category, it has category like $(\mathsf{s}\backslash\mathsf{n})/\mathsf{n}$. However, this approach will require yet further categories for sentences which have nominal predicates taking headless relative clauses as arguments (409a,c), adjectival predicative sentences (404b), and others, making *ang* a very heavily loaded lexical item. A better option is to assume that *ang* and the second noun of a nominal predicative sentence combine to produce np_n and then let the first noun be a function from np_n into an non-finite s, represented by the category type s_v in the hierarchy of Figure 3.1. Carrying the above reasoning to adjectival predicates as well, the derivations for (404a,b) would then be as in (405).

(405) a.
$$\dfrac{\dfrac{\textit{artista}}{\mathsf{s}_v/\mathsf{np}_n} \quad \dfrac{\textit{ang babae}}{\mathsf{np}_n}}{\mathsf{s}_v}>$$

b.
$$\dfrac{\dfrac{\textit{matalino}}{\mathsf{s}_v/\mathsf{np}_n} \quad \dfrac{\textit{ang babae}}{\mathsf{np}_n}}{\mathsf{s}_v}>$$

This approach treats bare nouns as properties, which is rather unsurprising in the case of predicate nominals—especially in light of the Montogovian tradition. It furthermore sits naturally within the approach advocated by Komagata (2002) for coordination of supposedly unlike categories, as in (406) from Sag *et al.* (1985):

(406) Pat is a republican and proud of it.

Komagata argues that such cases should be treated as instances of *like* coordination. Specifically, he uses a category pe, or *predicative element*, that is assigned to noun phrases, common nouns, and adjectival elements. Essentially, pe is an atomic encoding of the category $\mathsf{s}|\mathsf{n}$ which may only act as an argument for small clause and copular sentences. The difference in Tagalog, then, is that these predicative elements are active both as arguments and as functors and may thus consume their argument to form a sentence as in (405a).[1]

We might consider at this point the possibility that nouns have the two categories n and $\mathsf{s}_v/\mathsf{np}_n$, the former pertaining to bare nouns and the latter to predicative sentences. However, we may also opt to omit the category n entirely for bare nouns and use just $\mathsf{s}_v/\mathsf{np}_n$ for all occurrences. This accords with Gil (1993), who argues that one of the major typological differences between Tagalog and languages like English is that it makes no syntactic or semantic distinction between nouns, verbs, and adjectives. He provides many more examples like those in (404) which support this hypothesis.

As mentioned previously, a consequence of this move is that the marker categories must be functions from unary predicates to nominal arguments, leading to the following categories:

(407) a. *ang* $\vdash \mathsf{np}_n/_{\star}(\mathsf{s}/_{\diamond}\mathsf{np})$

b. *ng* $\vdash \mathsf{np}_g/_{\star}(\mathsf{s}/_{\diamond}\mathsf{np})$

c. *sa* $\vdash \mathsf{np}_d/_{\star}(\mathsf{s}/_{\diamond}\mathsf{np})$

With these categories in hand, we can give the following derivation for sentence (139).

(408)

$$
\begin{array}{ccccc}
\textit{bumili} & \textit{ang} & \textit{titser} & \textit{ng} & \textit{libro} \\
\hline
\mathsf{s}_i\{/_{\times}\mathsf{np}_{gen}, /\mathsf{np}_n\} & \mathsf{np}_n/_{\star}(\mathsf{s}/_{\diamond}\mathsf{np}) & \mathsf{s}_v/\mathsf{np}_n & \mathsf{np}_g/_{\star}(\mathsf{s}/_{\diamond}\mathsf{np}) & \mathsf{s}_v/\mathsf{np}_n \\
 & \multicolumn{2}{c}{\underline{\qquad\qquad\qquad}>} & \multicolumn{2}{c}{\underline{\qquad\qquad\qquad}>} \\
 & \multicolumn{2}{c}{\mathsf{np}_n} & \multicolumn{2}{c}{\mathsf{np}_g} \\
\multicolumn{3}{c}{\underline{\qquad\qquad\qquad\qquad\qquad}>} & & \\
\multicolumn{3}{c}{\mathsf{s}_i/_{\times}\mathsf{np}_g} & & \\
\multicolumn{5}{c}{\underline{\qquad\qquad\qquad\qquad\qquad\qquad\qquad}>} \\
\multicolumn{5}{c}{\mathsf{s}_i}
\end{array}
$$

Note that the modality $\diamond$ is a subtype of $\triangleright$, so the slashes $/_{\diamond}$ and $/$ therefore unify.

[1] An interesting possibility exists to bring unary modalities into MULTI-MODAL CCG and use them to effectively block categories from being the principal functor in the application of *any* of the combinatory rules, e.g. $\Box^{\downarrow}(\mathsf{s}\backslash\mathsf{n})$. Then the use of pe can be discarded in favor of a category that encodes the syntactic reflection of the semantics of predicative elements more faithfully, without the risk of generating non-sentences such as *John smart*.

These categories look more complicated than the more standard $\mathsf{np}/_{\diamond}\mathsf{n}$ and n, but an immediate payoff is that with no further assumptions they provide an analysis of headless relative clauses (HRCs) in Tagalog, which we turn to next.

7.1.3 Headless Relative Clause Formation

Based on the sentences (398b) and (401), the possibilities for using them as HRCs in nominal predicative sentences are those in (409).

(409) a. Titser ang bumili ng libro.
teacher NOM buy-AV GEN-book
The one who bought a book is a teacher.

b. *Libro ang bumili ang titser.
book NOM buy-AV NOM-teacher
(for: *The thing the teacher bought is a book.*)

c. Libro ang binili ng titser.
book NOM buy-OV GEN-teacher
The thing a teacher bought is a book.

d. *Titser ang binili ang libro.
teacher NOM buy-OV NOM-book
(for: *The one who bought the book is a teacher.*)

As can be seen from the impossibility of (409b,d), only the nominative arguments of the sentences on which they are based may be absent in HRCs. The role of *ang* in such sentences is often considered to be that of a complementizer (Guilfoyle *et al.*, 1992; Nakamura, 1994; Richards, 2000). The category given for *ang* in (407) immediately permits an analysis in which *ang* in such sentences is treated no differently than it is in pre-nominal positions, in agreement with Kroeger (1993). Sentences like those in (409a,c)—in addition to the nominal predicative sentence (404)—can be derived with the lexicon developed thus far. The derivation of (409a) is given below.

(410)
$$\begin{array}{cccc}
\textit{titser} & \textit{ang} & \textit{bumili} & \textit{ng libro} \\
\hline
\mathsf{s}_v/\mathsf{np}_n & \mathsf{np}_n/_{\star}(\mathsf{s}/_{\diamond}\mathsf{np}) & \mathsf{s}_i\{/_{\times}\mathsf{np}_{gen}, /\mathsf{np}_n\} & \mathsf{np}_g \\
\cline{3-4}
 & & \multicolumn{2}{c}{\mathsf{s}_i/\mathsf{np}_n \quad >} \\
\cline{2-4}
 & \multicolumn{3}{c}{\mathsf{np}_n \quad >} \\
\cline{1-4}
\multicolumn{4}{c}{\mathsf{s}_v \quad >}
\end{array}$$

Sentences like (409a,c) are thus syntactically similar to more simple nominal predicative sentence like (404a), and their analysis follows immediately from the proposed categories for nouns and markers. Furthermore, only the nominative argument can be the missing argument in an HRC since it is the only one which is decorated with the $\triangleright$ mode compatible with *ang*'s argument. We see this clearly in the following blocked derivation of (409d).

(411)
$$\begin{array}{cccc}
\textit{titser} & \textit{ang} & \textit{binili} & \textit{ang libro} \\
\hline
\mathsf{s}_v/\mathsf{np}_n & \mathsf{np}_n/_{\star}(\mathsf{s}/_{\diamond}\mathsf{np}) & \mathsf{s}_i\{/\mathsf{np}_{nom}, /_{\times}\mathsf{np}_g\} & \mathsf{np}_n \\
\cline{3-4}
 & & \multicolumn{2}{c}{\mathsf{s}_i/_{\times}\mathsf{np}_g \quad >} \\
\cline{2-4}
 & \multicolumn{3}{c}{*}
\end{array}$$

The lexicon developed so far thus produces the subject/object asymmetry of (409) by *modeling* in the verbal categories the fact that distinguished arguments in Tagalog are extractable because they have an associative-enabled slash that is compatible with the slash required by the extracting type category of *ang*, $\mathsf{np}_n/_{\star}(\mathsf{s}/_{\diamond}\mathsf{np})$. Ultimately, the facts of long distance *wh*-extraction considered in §7.1.5 will show that these assumptions about the Tagalog lexicon are forced and therefore *explain*, as well as model, such asymmetries.

Another advantage of the analysis is that it needs only one category for *ang*, unlike Guilfoyle *et al.* (1992), Nakamura (1998), and Richards (2000), who take *ang* to be a complementizer in sentences like (409a,c) as well as a case-marker/determiner in prenominal positions.

Given the proposed category for *ang*, an intransitive verb preceded by a marker should be able to act as an argument in a sentence. This correctly predicts that a sentence such as (412) is grammatical. The derivation proceeds in a manner very similar to that in (408).

(412) Bumili ang yumaman ng libro.
buy-AV NOM-became_rich GEN-book
The one who got rich bought a book.

Transitive and ditransitive verbs which have consumed all but their nominative argument may be used in this way as well. Derivation (410) shows that strings such as *bumili ng libro* can be case marked and used as nominal arguments. The use of a nominative argument such as *ang bumili ng libro* is not restricted to just nominal predicative sentences—it may be used in a sentence such as (413), which has the derivation in (414).

(413) Nagbigay ang bumili ng libro ng isda sa babae.
gave-AV NOM buy-AV GEN-book GEN-fish DAT-woman
The one who bought a book gave some fish to the woman.

(414)

nagbigay	*ang*	*bumili ng libro*	*ng isda*	*sa babae*
$\mathsf{s}_i\{/_{\times}\mathsf{np}_d, /_{\times}\mathsf{np}_g, /\mathsf{np}_n\}$	$\mathsf{np}_n/_{\star}(\mathsf{s}/_{\diamond}\mathsf{np})$	$\mathsf{s}_i/\mathsf{np}_n$	np_g	np_d

np_n (>)

$\mathsf{s}_i/_{\times}\{\mathsf{np}_d, \mathsf{np}_g\}$ (>)

$\mathsf{s}_i/_{\times}\mathsf{np}_d$ (>)

s_i (>)

Adjectives such as *matalino* may also be *ang*-marked, as seen in (415), which also has a derivation similar to those already seen.

(415) bumili ang matalino ng libro.
buy-AV NOM-smart GEN-book
The smart one bought a book.

By assuming that nouns have only the predicational category $\mathsf{s}_v/\mathsf{np}_n$, we have forced the markers to have categories which correctly predict the restriction on HRC formation. We now turn to more evidence in support of the semantically transparent category $\mathsf{s}_v/\mathsf{np}_n$ for nouns and adjectives, and examine further its role in its conspiracy with other functional categories in bringing about the asymmetries found in Tagalog.

7.1.4 Adjectival Modification and Relativization

The linker *na/ng* occurs often and in many seemingly different roles in Tagalog sentences.[2] We will only consider its role in adjectival modification and in relativization, with the result that the linker does not function differently for these two constructions—similar to what was found for the behavior of *ang* in conjunction with nouns, adjectives, verbs, and HRCs. It should be noted that the marker *ng* and the linker *ng* are separate, phonetically distinct lexical items which share the same orthography.

Adjectives may precede (416a) or follow (416b) the noun which they modify, although it is apparently more common and preferable for the adjective to precede

[2]Whether *na* or *ng* is the form of the linker used depends on the phonological properties of the last segment of the word preceding it (e.g. a consonant as in ...*Manuel na*.... and or a vowel as in ...*libro-ng*...).

the noun. When two or more adjectives are required, they are coordinated as in (416c).

(416) a. bago-ng libro
new-LNK book
new book

b. libro-ng bago
book-LNK new
new book

c. bago at maganda-ng libro
new and beautiful-LNK book
beautiful new book

Since bare nouns are unary predicates, we can see that adjectival modification in Tagalog involves the combination of several properties. These adjectivally modified constructions are properties similar to the others we have seen, and we must produce the category $\mathsf{s}_v/\mathsf{np}_n$ for them. Since both nouns and adjectives share the category $\mathsf{s}_v/\mathsf{np}_n$, I assign *na/ng* the following lexical entry:

(417) $\textit{na, ng} \vdash (\mathsf{s}_v/\mathsf{np}_n)\{/_\star(\mathsf{s}/_\diamond\mathsf{np}), \backslash_\star(\mathsf{s}_v/_\diamond\mathsf{np})\}$

This category may appear somewhat complex and imposing at first, but it is actually not very different from the more simple looking category for noun coordination in English, $\mathsf{n}\{\backslash_\star\mathsf{n}, /_\star\mathsf{n}\}$. Because it is a coordinating-like category, it receives the application-only mode ($\star$) on its slashes to limit its combinatory potential, similar to the category given for English coordination discussed in §5.2. The derivation of (416a) is given in (418).

(418)
$$\dfrac{\dfrac{\dfrac{\textit{bago}}{\mathsf{s}_v/\mathsf{np}_n} \quad \dfrac{\textit{-ng}}{(\mathsf{s}_v/\mathsf{np}_n)\{/_\star(\mathsf{s}/_\diamond\mathsf{np}), \backslash_\star(\mathsf{s}_v/_\diamond\mathsf{np})\}}}{(\mathsf{s}_v/\mathsf{np}_n)/_\star(\mathsf{s}/_\diamond\mathsf{np})}< \quad \dfrac{\textit{libro}}{\mathsf{s}_v/\mathsf{np}_n}}{\mathsf{s}_v/\mathsf{np}_n}>$$

Both orderings (416a) and (416b) are acceptable since both arguments of the category of *na/ng* are compatible with categories rooted in s_v. The fact that the rightward argument of *na/ng*'s category is rooted in s allows *na/ng* to act as a relativizer, as discussed below. The category of the coordinator *at* is almost the same except that it can only coordinate verbal functions rooted in s_v:

(419) $\textit{at} \vdash (\mathsf{s}_v/\mathsf{np}_n)\{/_\star(\mathsf{s}_v/_\diamond\mathsf{np}), \backslash_\star(\mathsf{s}_v/_\diamond\mathsf{np})\}$

When preceded by *ang*, *ng*, or *sa*, the derivation is just like that with the bare noun.

(420)
$$\cfrac{\cfrac{\textit{ng}}{\mathsf{np}_g/_\star(\mathsf{s}/_\diamond\mathsf{np})} \qquad \cfrac{\textit{bagong libro}}{\mathsf{s}_v/\mathsf{np}_n}}{\mathsf{np}_g}\,>$$

Interestingly, relativized sentences also involve the linker *na/ng*. We find the subject/object asymmetry operative again — only the nominative argument can be relativized.

(421) a. lalaki-ng nagbigay ng libro sa babae
man-LNK give-AV GEN-book DAT-woman
man that gave a book to the woman

b. *lalaki-ng ibinigay ang libro sa babae
man-LNK give-IV NOM-book DAT-woman
(for: *man that gave the book to the woman'*

c. *libro-ng nagbigay ang lalaki sa babae
book-LNK give-AV NOM-man DAT-woman
(for: *book that the man gave to the woman'*

d. libro-ng ibinigay ng lalaki sa babae
book-LNK give-IV GEN-man DAT-woman
book that a man gave to the woman

Given the previous analysis of HRC formation, this is remarkably similar to adjectival modification. In fact, the proposed category set predicts that relativization will work in this way—the derivation of (421a) falls out of the categories already assumed.

(422)
$$\cfrac{\cfrac{\cfrac{\textit{lalaki}}{\mathsf{s}_v/\mathsf{np}_n} \quad \cfrac{\textit{-ng}}{(\mathsf{s}_v/\mathsf{np}_n)\{/_\star(\mathsf{s}/_\diamond\mathsf{np}),\ \backslash_\star(\mathsf{s}_v/_\diamond\mathsf{np})\}}}{(\mathsf{s}_v/\mathsf{np}_n)/_\star(\mathsf{s}/_\diamond\mathsf{np})}\,< \qquad \cfrac{\cfrac{\cfrac{\textit{nagbigay}}{\mathsf{s}_i\{/_\times\mathsf{np}_d,\ /_\times\mathsf{np}_g,\ /\mathsf{np}_n\}} \quad \cfrac{\textit{ng libro}}{\mathsf{np}_g}}{\mathsf{s}_i\{/_\times\mathsf{np}_d,\ /\mathsf{np}_n\}}\,> \quad \cfrac{\textit{sa babae}}{\mathsf{np}_d}}{\mathsf{s}_i/\mathsf{np}_n}\,>}{\mathsf{s}_v/\mathsf{np}_n}\,>$$

The resulting category in (422) should look very familiar by now. Importantly, ungrammatical attempts at relativization will not go through.

(423) a. *[libro-ng]$_{(\mathsf{s}_v/\mathsf{np}_n)/_\star(\mathsf{s}/_\diamond\mathsf{np})}$ [nagbigay ang lalaki sa babae]$_{\mathsf{s}_i/_\times\mathsf{np}_g}$

b. *[lalaki-ng]$_{(\mathsf{s}_v/\mathsf{np}_n)/_\star(\mathsf{s}/_\diamond\mathsf{np})}$ [ibinigay ang libro sa babae]$_{\mathsf{s}_i/_\times\mathsf{np}_g}$

Another asymmetry thus manifests itself, due to the consistency of functions such as nouns, adjectives, and verbs in seeking arguments with harmonic associative slashes and of case markers and linkers in seeking these functions. We turn next to *wh*-extraction of arguments, yet another syntactic phenomenon which manifests the subject/object asymmetry. It is with long distance *wh*-extraction that we see the modes doing explicit grammatical work, as opposed to their basically specificational use thus far in the analysis.

7.1.5 Local Argument *Wh*-extraction

The primary data for argument *wh*-extraction appears quite similar to that for predicate nominals with HRC arguments (409), and in fact I argue that *wh*-sentences are no different, in contrast with Kroeger (1993) and Richards (2000), who regard *wh*-sentences as clefts. In most clause types, the extraction site must correspond to the *ang*-marked noun, as seen in (424).

(424) a. Sino ang bumili ng libro?
who NOM buy-AV GEN-book
Who bought a book?

b. *Ano ang bumili ang titser?
what NOM buy-AV NOM-teacher
(for: *What did the teacher buy?*)

c. Ano ang binili ng titser?
what NOM buy-OV GEN-teacher
What did a teacher buy?

d. *Sino ang binili ang libro?
who NOM buy-OV NOM-book
(for: *Who bought the book?*)

Strings like *ang bumili ng libro* should by now be quite familiar as producing the category np_n. Therefore, if we assume that *wh*-words are predicates with the category $\ddot{\mathsf{s}}_c/\mathsf{np}_n$, the above asymmetry immediately falls out.[3] The *wh*-item requires an np_n, and we already know that *ang* can referentialize predicates which are missing their associatively enabled arguments. The derivation of (424a) thus succeeds, as shown in (425), and that of (424b), shown in (426), fails. Similar success and failure occurs for (424c,d).

[3]Technically, this category is $\ddot{\mathsf{s}}_{c\,VFORM=non-fin}/\mathsf{np}_n$ since the clause created by it is not actually tensed. Since this is peripheral to the present discussion, I omit this detail from the categories.

(425)

$$\begin{array}{cccc}
\textit{sino} & \textit{ang} & \textit{bumili} & \textit{ng libro} \\
\ddot{\mathsf{s}}_c/\mathsf{np}_n & \mathsf{np}_n/_{\star}(\mathsf{s}/_{\diamond}\mathsf{np}) & \mathsf{s}_i\{/_{\times}\mathsf{np}_{gen}, /\mathsf{np}_n\} & \mathsf{np}_g
\end{array}$$

$$\textit{bumili ng libro}: \; \mathsf{s}_i/\mathsf{np}_n \; (>)$$
$$\textit{ang bumili ng libro}: \; \mathsf{np}_n \; (>)$$
$$\textit{sino ang bumili ng libro}: \; \ddot{\mathsf{s}}_c \; (>)$$

(426)

$$\begin{array}{cccc}
\textit{*ano} & \textit{ang} & \textit{bumili} & \textit{ang titser} \\
\ddot{\mathsf{s}}_c/\mathsf{np}_n & \mathsf{np}_n/_{\star}(\mathsf{s}/_{\diamond}\mathsf{np}) & \mathsf{s}_i\{/_{\times}\mathsf{np}_{gen}, /\mathsf{np}_n\} & \mathsf{np}_n
\end{array}$$

$$\textit{bumili ang titser}: \; \mathsf{s}_i/_{\times}\mathsf{np}_g \; (>)$$
$$\textit{ang bumili ang titser}: \; * $$

Note that the category given for *wh*-items allows them to be case marked, thus allowing strings like *ang sino* and *ng ano* to be nominal arguments which function as in situ *wh*-phrases. Resty Cena (p.c.) indicates that phrases such as *ng ano* are quite common, but Richards (2000) claims that this is not possible. To exclude *wh*-items from being case marked, I assume that the case markers take arguments which bear the value *unmarked* for the feature MARKING.

This category for *wh*-elements accords with their use as modifiers also. Schachter and Otanes (1972) give (427) as an example of this.

(427) Sino -ng babae ang pinakamaganda?
who LNK woman NOM-prettiest
Which woman is the prettiest?

The derivation is straightforward, proceeding along similar lines to those seen before with the assumption of an additional category for the linker. The linker combines the *wh*-predicate with *babae* and then applies to the nominal argument.

(428)

$$\begin{array}{cccc}
\textit{sino} & \textit{-ng} & \textit{babae} & \textit{ang pinakamaganda} \\
\ddot{\mathsf{s}}_c/\mathsf{np}_n & (\ddot{\mathsf{s}}_c/\mathsf{np}_n)\{/_{\star}(\mathsf{s}_v/\mathsf{np}_n), \backslash_{\star}(\ddot{\mathsf{s}}_c/\mathsf{np}_n)\} & \mathsf{s}_v/\mathsf{np}_n & \mathsf{np}_n
\end{array}$$

$$\textit{sino -ng}: \; (\ddot{\mathsf{s}}_c/\mathsf{np}_n)/_{\star}(\mathsf{s}_v/\mathsf{np}_n) \; (<)$$
$$\textit{sino -ng babae}: \; \ddot{\mathsf{s}}_c/\mathsf{np}_n \; (>)$$
$$\textit{sino -ng babae ang pinakamaganda}: \; \ddot{\mathsf{s}}_c \; (>)$$

This extra category for the linker is only marginally different from the one already used.

Extraction from a matrix clause and the asymmetry that comes with it thus follows trivially from the analysis of HRC referentialization. We will now look at extraction from verbs with sentential complements, which will provide further evidence the categories we have assigned are appropriate and explanatory.

If we alternate the voice of the matrix verb and the subordinate verb, we have the sentences in (429) (from Nakamura 1998).

(429) a. Nagsabi si Pedro na bumili si Linda ng kotse.
say-AV Nom-Pedro Comp buy-AV Nom-Linda Gen-car
Pedro said that Linda bought a car.

b. Nagsabi si Pedro na binili ni Linda ang kotse.
say-AV Nom-Pedro Comp buy-OV Gen-Linda Nom-car
Pedro said that Linda bought the car.

c. Sinabi ni Pedro na bumili si Linda ng kotse.
say-OV Gen-Pedro Comp buy-AV Nom-Linda Gen-car
Pedro said that Linda bought a car.

d. Sinabi ni Pedro na binili ni Linda ang kotse.
say-OV Gen-Pedro Comp buy-OV Gen-Linda Nom-car
Pedro said that Linda bought the car.

Kroeger (1993) and Nakamura (1998) assume that the matrix subject of sentences such as (429c,d) is in fact the embedded clause. In present terms, the sentence is distinguished like a nominative noun phrase, though it does not receive overt morphological marking. Evidence that this is a correct assumption comes from the ability to both *wh*-extract and *ay*-invert the sentence. *Ay*-inversion is discussed in §7.1.9, but what matters presently is that inverted noun phrases must be nominative. The sentences in (430) show that the sentential complement of *sinabi* can be questioned and *ay*-inverted.

(430) a. Ano ang sinabi ni Pedro?
what Nom said-OV Gen-Pedro
What did Pedro say?

b. Na tumakbo ang lalaki ay sinabi ni Pedro.
Comp run-AV Nom-man Ay say-OV Gen-Pedro
Pedro said that the man runs.

If the matrix verb is in Active Voice (*nagsabi* rather than *sinabi*), such extraction and inversion is not possible. The data in (429,430) suggest the following categories and interpretations for *nagsabi* and *sinabi*.

(431) a. *nagsabi* $\vdash \mathsf{s}_i\{/_{\times}\ddot{\mathsf{s}}_{c_{gen}}, /\mathsf{np}_n\}$

b. *sinabi* $\vdash \mathsf{s}_i\{/\ddot{\mathsf{s}}_{c_{nom}}, /_{\times}\mathsf{np}_g\}$

We need to assign *na* the usual complementizer category, $\ddot{\mathsf{s}}_c/_{\diamond}\mathsf{s}_i$. Also, the sentential extraction in (430a) indicates that the we must assign *ang* the category $\mathsf{np}_n/_{\star}(\mathsf{s}/_{\diamond}\ddot{\mathsf{s}}_c)$ in addition to $\mathsf{np}_n/_{\star}(\mathsf{s}/_{\diamond}\mathsf{np})$.[4]

Extraction of the matrix noun phrase from the sentences in (429) follows the pattern we have seen for other matrix clauses. It can be extracted when the matrix verb is in Active Voice (432a), but not when the verb is in Objective Voice (432b).

(432) a. Sino ang nagsabi na binili ni Linda ang kotse?
who Nom say-AV Comp buy-OV Gen-Linda Nom-car
Who said that Linda bought the car?

b. *Sino ang sinabi na binili ni Linda ang kotse?
who Nom say-OV Comp buy-OV Gen-Linda Nom-car
(for: *Who said that Linda bought the car?*)

Given the categories for *nagsabi* and *sinabi* in (431), this asymmetry follows automatically.

7.1.6 Long Distance *Wh*-extraction

The impact of the modes is fully felt when we examine long distance extraction in Tagalog. An argument may only be extracted from an embedded clause if the clause itself is the nominative argument of the matrix clause. The result of attempting to extract the **Patient** argument of the embedded clauses in (429) is given in (433).

(433) a. *Ano ang nagsabi si Pedro na bumili si Linda?
what Nom say-AV Nom-Pedro Comp buy-AV Nom-Linda
(for: *What is the thing that Pedro said that Linda bought?*)

b. *Ano ang nagsabi si Pedro na binili ni Linda?
what Nom say-AV Nom-Pedro Comp buy-OV Gen-Linda
(for: *What is the thing that Pedro said that Linda bought?*)

c. *Ano ang sinabi ni Pedro na bumili si Linda?
what Nom say-OV Gen-Pedro Comp buy-AV Nom-Linda
(for: *What is the thing that Pedro said that Linda bought?*)

d. Ano ang sinabi ni Pedro na binili ni Linda?
what Nom say-OV Gen-Pedro Comp buy-OV Gen-Linda
What is the thing that Pedro said that Linda bought?

[4] Note that I could use the underspecified category $\mathsf{np}_n/_{\star}(\mathsf{s}/_{\diamond}\ddot{\mathsf{a}})$, a function over predicates which are missing an argument which is licensed for associative extraction. This category will then cover both sentential and nominal extractions, but I will use the relevant appropriately specified category in derivations.

The status of each of the sentences (433) follows immediately from what has been said already. The derivation of (433d) is given in (434).

(434)
$$\begin{array}{ccccc}
\textit{ano} & \textit{ang} & \textit{sinabi} & \textit{ni Pedro} & \textit{na binili ni Linda}\\
\hline
\ddot{s}_c/np_n & np_n/_{\star}(s/_{\diamond}np) & s_i\{/\ddot{s}_{c\,nom}, /_{\times}np_g\} & np_g & \ddot{s}_c/np_n
\end{array}$$
$$\dfrac{\ddot{s}_c/np_n \quad \dfrac{np_n/_{\star}(s/_{\diamond}np) \quad \dfrac{\dfrac{s_i\{/\ddot{s}_{c\,nom}, /_{\times}np_g\} \quad np_g}{s_i/\ddot{s}_{cn}}> \quad \ddot{s}_c/np_n}{s_i/np_n}>\mathbf{B}}{np_n}>}{\ddot{s}_c}>$$

The fact that the nominative arguments of both *sinabi* and *binili* are labeled with the $\triangleright$ modality allows them to combine through forward composition ($>\mathbf{B}$). In CTL terms, both verbs allow the extracted argument to shift to the right periphery through the use the Right Association (RA) rule (344) and not remain locked inside the CTL structured antecedent, as shown in the following (abbreviated) proof.

(435)
$$\dfrac{\dfrac{\dfrac{(\textit{sinabi ni Pedro}) \vdash s_i/\ddot{s}_{cn} \qquad \dfrac{(\textit{na binili ni Linda}) \vdash \ddot{s}_c/np_n \qquad [x_1 \vdash n]^1}{((\textit{na binili ni Linda}) \triangleright x_1) \vdash \ddot{s}_c}[/E]}{((\textit{sinabi ni Pedro}) \triangleright ((\textit{na binili ni Linda}) \triangleright x_1)) \vdash s_i}[/E]}{(((\textit{sinabi ni Pedro}) \triangleright (\textit{na binili ni Linda})) \triangleright x_1) \vdash s_i}[RA]}{((\textit{sinabi ni Pedro}) \triangleright (\textit{na binili ni Linda})) \vdash s_i/np_n}[/I]^1$$

Derivations of the other three sentences of (433) are blocked because of the failure of one of the verbs to license $>\mathbf{B}$.

(436) a. *Ano ang [nagsabi si Pedro]$_{s_i/_{\times}\ddot{s}_{cg}}$ [na binili ni Linda]$_{\ddot{s}_c/np_n}$?

b. *Ano ang nagsabi si Pedro [na]$_{\ddot{s}_c/s_i}$ [bumili si Linda]$_{s_i/_{\times}np_g}$?

c. *Ano ang sinabi ni Pedro [na]$_{\ddot{s}_c/s_i}$ [bumili si Linda]$_{s_i/_{\times}np_g}$?

The long distance extraction asymmetry thus arises because the lower clause, with category $s_i/_{\times}np_g$, constrains its unconsumed argument from escaping (436b,c) or because the matrix verb refuses to allow its subordinate clause or any arguments of its subordinate clause to escape (436a). We thus see how the local configurations responsible for producing the bounded asymmetries conspire to create a global pattern.

7.1.7 *Wh*-extraction from Voiceless Verbs

The categories assigned to Tagalog verbs in (400) have played a large role in producing the asymmetric behavior discussed thus far. Most importantly, they single out one of their arguments to receive nominative case and the $\triangleright$ modality. Nonetheless, the MULTI-MODAL CCG system admits of the possibility of a verb labeling all of its arguments with the $\triangleright$ modality. This proves to be a necessary degree of freedom, as data regarding verbs that lack voice morphology demonstrates.

Verbs in the recent past lack voice morphology, and none of the arguments are nominative, as in (437). Reduplication and the morpheme *ka* put the verb *bili* in the recent past form.

(437) Kabibili lang ni Juan ng tela.
buy-RECPAST just GEN-Juan GEN-cloth
Juan has just bought some cloth.

Here we see that there are no nominative arguments, and we discover that all of the arguments may be *wh*-extracted. In (438a), the **Actor** has been extracted, while in (438b) it is the **Patient**.

(438) a. Sino ang kabibili lang ng tela?
who NOM buy-RECPAST just GEN-cloth
Who has just bought some cloth?

b. Ano ang kabibili lang ni Juan?
what NOM buy-RECPAST just GEN-Juan
What has Juan just bought?

Relativization of both arguments is also grammatical. The new category we must add to the lexicon is (439):

(439) *kabibili* $\vdash \mathsf{s}_i/\{\mathsf{np}_g, \mathsf{np}_g\}$

In addition to licensing both arguments for extraction, this category allows these arguments to scramble within the clause, but since there is no unique case marking, it will permit ambiguous readings (for example, (437) would receive the interpretation corresponding to *Some cloth has just bought Juan* in addition to the gloss of (437)).

With these categories, the extractions in (438) are derived as in (440). The arguments of *kabibili* are labeled as **Actor** and **Patient** to aid readability and to emphasize that the logical form built is different for the two derivations.

(440) a. **Actor** extraction (438a):

$$
\begin{array}{ccccc}
\textit{sino} & \textit{ang} & \textit{kabibili} & \textit{lang} & \textit{ng tela} \\
\overline{\ddot{\mathsf{s}}_c/\mathsf{np}_n} & \overline{\mathsf{np}_n/_\star(\mathsf{s}/_\diamond\mathsf{np})} & \overline{\mathsf{s}_i/\{\mathsf{np}_g{}^{\mathbf{Pat}}, \mathsf{np}_g{}^{\mathbf{Act}}\}} & \overline{\mathsf{s}\backslash\mathsf{s}} & \overline{\mathsf{np}_g}
\end{array}
$$

$$\mathsf{s}_i/\{\mathsf{np}_g{}^{\mathbf{Pat}}, \mathsf{np}_g{}^{\mathbf{Act}}\} \quad {<}\mathbf{B}^2_\times$$

$$\mathsf{s}_i/\mathsf{np}_g{}^{\mathbf{Act}} \quad {>}$$

$$\mathsf{np}_n \quad {>}$$

$$\ddot{\mathsf{s}}_c \quad {>}$$

b. **Patient** extraction (438b):

$$
\begin{array}{ccccc}
\textit{ano} & \textit{ang} & \textit{kabibili} & \textit{lang} & \textit{ni Juan} \\
\overline{\ddot{\mathsf{s}}_c/\mathsf{np}_n} & \overline{\mathsf{np}_n/_\star(\mathsf{s}/_\diamond\mathsf{np})} & \overline{\mathsf{s}_i/\{\mathsf{np}_g{}^{\mathbf{Pat}}, \mathsf{np}_g{}^{\mathbf{Act}}\}} & \overline{\mathsf{s}\backslash\mathsf{s}} & \overline{\mathsf{np}_g}
\end{array}
$$

$$\mathsf{s}_i/\{\mathsf{np}_g{}^{\mathbf{Pat}}, \mathsf{np}_g{}^{\mathbf{Act}}\} \quad {<}\mathbf{B}^2_\times$$

$$\mathsf{s}_i/\mathsf{np}_g{}^{\mathbf{Pat}} \quad {>}$$

$$\mathsf{np}_n \quad {>}$$

$$\ddot{\mathsf{s}}_c \quad {>}$$

There is of course an inherent ambiguity induced by the category assigned to *kabibili* since its two arguments are in the same set but have the same case. I take the resolution of this ambiguity to be mediated by reasoning about the plausibility of different arguments filling different roles, e.g. *tela* 'cloth' is more probable as a **Patient** and *Juan* is the more likely **Actor**.

Kroeger (1993) accounts for Tagalog's asymmetries in Lexical Functional Grammar (Kaplan and Bresnan, 1982) by arguing that nominative arguments are subjects and then formulating a functional structure pathway that is sensitive to the grammatical relation SUBJ. On his own admission, his account fails to permit the non-**Actor** arguments of verbs like *kabibili* to extract since they can in no way be construed to be subjects.

Also relevant to extraction of non-nominative arguments are the comitative and comparative constructions. The following examples are from Cena (1979), cited in Nakamura (1998).

(441) a. Kasama ni Juan ang tao.
with GEN-Juan NOM-man
Juan is with the man.

b. Kasingtaas ni Juan ang tao.
as-tall-as GEN-Juan NOM-man
Juan is as tall as the man.

In addition to being able to extract the *ang*-marked NP, we can extract the *ng*-marked one as well.

(442) a. Sino ang kasama ang tao?
who NOM with NOM-man
Who is with the man?

b. Sino ang kasingtaas ang tao?
who NOM as-tall-as NOM-man
Who is as tall as the man?

Like the recent past, these verbs lack voice morphology. We thus assume the categories in (443) and the extractions follow in the same manner as those in (440) without further specification.

(443) *kasama, kasingtaas* $\vdash \mathsf{s}_i/\{\mathsf{np}_n, \mathsf{np}_g\}$

7.1.8 Adjunct *Wh*-extraction

Adjunct extraction is less sensitive than argument extraction to the effects of voice-marking affixes. *Wh*-items like *sino* 'who' and *ano* 'what' are unary predicates which take an n argument, whereas the *wh*-item *saan* 'where' takes an s argument. Thus, the marker *ang* does not take part in adjunct extraction in the manner which it does for argument extraction. This can be seen in (444), in which the adjunct extraction is permissible for both voices of the verb *bili* 'buy'.

(444) a. Saan bumili si Juan ng damit.
where buy-AV NOM-Juan GEN-dress
Where did Juan buy a dress?

b. Saan binili ni Juan ang damit.
where buy-OV GEN-Juan NOM-dress
Where did Juan buy the dress?

Adjunct *wh*-words like *saan* thus take the category $\ddot{\mathsf{s}}_c/_{\star}\mathsf{s}_i$, which is the same as that of English adjunct extracting categories.

One the the other hand, long distance extraction of adjuncts provides another asymmetry in Tagalog. If the embedded sentence is not the subject, adjuncts cannot be extracted. Thus, in the following sentences from Nakamura (1998), the question can only be about the location of the saying event in (445a,b), while it can be about the location of either the saying event or the buying event in (445c,d).

(445) a. Saan nagsabi si Pedro na bumili ang tao ng damit.
where say-AV NOM-Pedro COMP buy-AV NOM-man GEN-dress
Where$_i$ did Pedro say t_i that the man bought a dress?

b. Saan nagsabi si Pedro na binili ng tao ang damit.
where say-AV NOM-Pedro COMP buy-OV GEN-man NOM-dress
Where$_i$ did Pedro say t_i that the man bought a dress?

c. Saan sinabi ni Pedro na bumili ang tao ng damit.
where say-OV GEN-Pedro COMP buy-AV NOM-man GEN-dress
Where$_i$ did Pedro say t_i that the man bought a dress?

Where$_i$ did Pedro say that the man bought a dress t_i?

d. Saan sinabi ni Pedro na binili ng tao ang damit.
where say-OV GEN-Pedro COMP buy-OV GEN-man NOM-dress
Where$_i$ did Pedro say t_i that the man bought a dress?

Where$_i$ did Pedro say that the man bought a dress t_i?

The matrix adjuncts extract in the same manner as (444) and are thus insensitive to any effects of verbal voice. The asymmetry in subordinate clause adjunct extraction between (445a,b) and (445c,d) arises if we assume the following additional category for *saan*.

(446) *saan* $\vdash (\ddot{s}_c/_\diamond\ddot{s}_c)/_\star(s_i/_\diamond\ddot{s}_c)$

This category embodies a phrase structure analysis in which the adjunct trace has adjoined at the lower clause in the tree and moved out. It would be redundant given a fully specified theory of the lexicon and the $\ddot{s}_c/_\star s_i$ category of *saan*.

The successful subordinate clause adjunct extraction of (445c) is shown in derivation (447), while the inability to extract the lower clause of (445a) is demonstrated in (448).

(447)
$$\begin{array}{cccc}
\textit{saan} & \textit{sinabi} & \textit{ni Pedro} & \textit{na bumili ang tao ng damit} \\
(\ddot{s}_c/_\diamond\ddot{s}_c)/_\star(s_i/_\diamond\ddot{s}_c) & s_i\{/\ddot{s}_{c\,nom}, /_\times np_g\} & np_g & \ddot{s}_c \\
 & \multicolumn{2}{c}{s_i/\ddot{s}_{c\,gen}} \; > & \\
\multicolumn{3}{c}{\ddot{s}_c/_\diamond\ddot{s}_c} \; > & \\
\multicolumn{4}{c}{\ddot{s}_c} \; >
\end{array}$$

(448)
$$\begin{array}{cccc}
\textit{saan} & \textit{nagsabi} & \textit{si Pedro} & \textit{na bumili ang tao ng damit} \\
(\ddot{s}_c/_\diamond\ddot{s}_c)/_\star(s_i/_\diamond\ddot{s}_c) & s_i\{/_\times\ddot{s}_{c\,gen}, /np_n\} & np_n & \ddot{s}_c \\
 & \multicolumn{2}{c}{s_i/_\times\ddot{s}_{c\,gen}} \; > & \\
\multicolumn{3}{c}{} \; * &
\end{array}$$

Ajdunct extraction over further distances will meet with the same restrictions due to the limitations which the modes place on the combinatory rules, just as was the case for long distance argument extraction in (436).

7.1.9 *Ay*-inversion

With the data covered so far, the analysis keeps pace with the coverage of Nakamura (1998) and surpasses that of Kroeger (1993). In this section, I go beyond Nakamura and consider the subject/object asymmetry present in *ay*-inversion. Furthermore, I show that restrictions on long distance *ay*-inversion follow for the same grammatical reasons as they did for both argument and adjunct extraction and that Nakamura's analysis of extraction will overgenerate if applied to *ay*-inversion.

In *ay*-inversion, the nominative argument of the clause is fronted and the word *ay* comes between the argument and the verb, as demonstrated in the following *ay*-inverted sentences.

(449) a. Ang lalaki ay tumakbo.
NOM-man AY ran-AV
The man ran.

b. Ang titser ay bumili ng libro.
NOM-teacher AY buy-AV GEN-book
The teacher bought a book.

c. Ang lalaki ay nagbigay ng libro sa babae.
NOM-man AY gave-AV GEN-book DAT-woman
The man gave a book to the woman.

Sentences which are subjects may also be *ay*-inverted, as in (430). Non-nominative arguments, however, may not be inverted:

(450) *Ng titser ay bumili ang titser.
GEN-book AY buy-AV NOM-teacher
The teacher bought a book.

Even though both arguments of the recent past construction (437) may be *wh*-extracted and relativized, neither may be *ay*-inverted, as shown in (451).

(451) a. *Ni Juan ay kabibili lang ng tela?
GEN-Juan AY buy-RECPAST just GEN-cloth
(for: *Juan just bought cloth.*)

b. *Ng tela ay kabibili lang ni Juan?
GEN-cloth AY buy-RECPAST just GEN-Juan
(for: *Juan just bought cloth.*)

The behavior of *ay*-inversion has other dimensions not presented here—adjuncts such as adverbials, gerunds, time adverbs, and subordinate clauses may also be inverted (Schachter and Otanes 1972). Kroeger (1993) suggests that inverted nominative arguments function as topics, while all other inverted elements are focussed. This indicates that we can deal with these cases differently, and I will hence only discuss inversion of nominative arguments as it is the case which is relevant to the present discussion.

The straightforward solution is to give *ay* a category which reverses the direction in which the verb seeks its nominative argument. We thus assign the following categories to *ay*:[5]

(452) $\mathit{ay} \vdash (\mathsf{s}\backslash_{\star}\mathsf{np}_n)/_{\star}(\mathsf{s}/_{\diamond}\mathsf{np}_n)$

(453) $\mathit{ay} \vdash (\mathsf{s}\backslash_{\star}\mathsf{s}_n)/_{\star}(\mathsf{s}/_{\diamond}\mathsf{s}_n)$

The result s of the overall category and result s of the first arguments of these categories are co-indexed so that the former receives the feature values of the latter after rule application, as seen in the derivation (454) below.

The argument which *ay* seeks is thus essentially the same as the one which the markers *ang*, *ng*, and *sa* and the linker *na/ng* seek. The only difference is that it specifies that the extracted argument must be nominative, thereby blocking *ay*-inversion with the recent past (451). The derivation of (449a) is given in (454).

(454)
$$
\begin{array}{cccc}
\mathit{ang} & \mathit{lalaki} & \mathit{ay} & \mathit{tumakbo} \\
\hline
\mathsf{np}_n/_{\star}(\mathsf{s}/_{\diamond}\mathsf{np}) & \mathsf{s}_v/\mathsf{np}_n & (\mathsf{s}\backslash_{\star}\mathsf{np}_n)/_{\star}(\mathsf{s}/_{\diamond}\mathsf{np}_n) & \mathsf{s}_i/\mathsf{np}_n \\
\hline
\multicolumn{2}{c}{\mathsf{np}_n \quad >} & \multicolumn{2}{c}{\mathsf{s}_i\backslash_{\star}\mathsf{np}_n \quad >} \\
\hline
\multicolumn{4}{c}{\mathsf{s}_i \quad <}
\end{array}
$$

The category given for *ay* predicts that *ay*-inversion can occur in embedded clauses since the inverted derivation produces a category for the clause that is indistinguishable from that in the non-inverted derivation. More importantly, it predicts that long distance *ay*-inversion can occur, proceeding in a manner similar to long distance extraction. Both of these are correct predictions, as the following examples from Kroeger (1993) demonstrate.

[5]Note that these two categories can be collapsed to the following underspecified category: $(\mathsf{s}\backslash_{\star}\ddot{a}_n)/_{\star}(\mathsf{s}/_{\diamond}\ddot{a}_n)$.

(455) a. Sinabi ni Pedro na ninakaw ni Josie ang kotse.
say-OV Gen-Pedro Comp steal-OV Gen-Josie Nom-car
Pedro said that Josie stole the car.

b. Sinabi ni Pedro na ang kotse ay ninakaw ni Josie.
say-OV Gen-Pedro Comp Nom-car Ay steal-OV Gen-Josie
Pedro said that the car, Josie stole.

c. ang kotse ay sinabi ni Pedro na ninakaw ni Josie.
Nom-car Ay say-OV Gen-Pedro Comp steal-OV Gen-Josie
The car, Pedro said that Josie stole.

Given (454), the derivation for (455b) is obvious. The derivation for (455c), given in (456), demonstrates the syntactic similarity to long distance *wh*-extraction, in which the matrix verb composes with the unsaturated subordinate verb.

(456)

$$
\begin{array}{ccccc}
\textit{ang kotse} & \textit{ay} & \textit{sinabi} & \textit{ni Pedro} & \textit{na ninakaw ni Josie} \\
\hline
\mathsf{np}_n & (\mathsf{s}\backslash_{\star}\mathsf{np}_n)/_{\star}(\mathsf{s}/_{\diamond}\mathsf{np}_n) & \mathsf{s}_i\{/\ddot{\mathsf{s}}_{c\,nom}, /_{\times}\mathsf{np}_g\} & \mathsf{np}_g & \ddot{\mathsf{s}}_c/\mathsf{np}_n \\
 & & \multicolumn{2}{c}{\xrightarrow{\hspace{6em}}>} & \\
 & & \multicolumn{2}{c}{\mathsf{s}_i/\ddot{\mathsf{s}}_{c\,nom}} & \\
 & & \multicolumn{3}{c}{\xrightarrow{\hspace{10em}}>\mathbf{B}} \\
 & & \multicolumn{3}{c}{\mathsf{s}_i/\mathsf{np}_n} \\
 & \multicolumn{4}{c}{\xrightarrow{\hspace{14em}}>} \\
 & \multicolumn{4}{c}{\mathsf{s}_i\backslash_{\star}\mathsf{np}_n} \\
\multicolumn{4}{c}{\xrightarrow{\hspace{14em}}<} & \\
\multicolumn{4}{c}{\mathsf{s}_i} &
\end{array}
$$

In fact, the same constraint which holds for long distance *wh*-extraction holds for long distance *ay*-inversion. The matrix verb must be in a voice which distinguishes the sentential complement. Thus, as long as *ay*-inversion takes place within the embedded clause, the matrix verb may vary its voice marking (457), but whenever long distance *ay*-inversion takes place the verb is restricted to a single voice, as demonstrated by the impossibility of (458a,b).

(457) a. Nagsabi si Pedro na ang kotse ay ninakaw ni Josie.
say-AV Nom-Pedro Comp Nom-car Ay steal-OV Gen-Josie
Pedro said that the car, Josie stole.

b. Sinabihan ako ni Pedro na ang kotse ay ninakaw ni Josie.
say-DV Nom-I Gen-Pedro Comp Nom-car Ay steal-OV Gen-Josie
I was told by Pedro that the car, Josie stole.

(458) a. *Ang kotse ay nagsabi si Pedro na ninakaw ni Josie.
Nom-car Ay say-AV Nom-Pedro Comp steal-OV Gen-Josie
(for: *The car, Pedro said that Josie stole.*)

b. *Ang kotse ay sinabihan ako ni Pedro na ninakaw ni Josie.
NOM-car AY say-DV NOM-I GEN-Pedro COMP steal-OV GEN-Josie
(for: *The car, I was told by Pedro that Josie stole.*)

These inversions are blocked for precisely the same reason as the ungrammatical long distance extractions shown in (436).

In addition to the fact that the MULTI-MODAL CCG machinery is far more constrained than the transderivational constraints of Nakamura (1998), it is here that the present analysis provides the greatest challenge to Nakamura's, which appears to make the wrong prediction about the examples in (459). For arguments of voiced verbs, the key to being extracted on Nakamura's analysis is to move to SPEC of IP, where the distinguished argument receives its case. Though Nakamura does not discuss *ay*-inversion, an analysis compatible with his account would involve overt NP-movement of the distinguished noun phrase to the SPEC of a functional projection higher than IP and lower than CP. This would have to be so for two reasons: *ay*-inversion can occur in subordinate clauses and the verb is assumed to move to INFL in Tagalog.

Under such an account, the asymmetries in *wh*-extraction and *ay*-inversion from clauses with a voiced verb would arise for the same reasons. However, it overgenerates. Both of the arguments of a non-voiced verb (e.g., (437)) should be invertible since there is no better competing derivation which would block the inversion in either case. This, in fact, is exactly why Nakamura's analysis allows both *wh*-extractions in such cases. However, neither argument may be inverted as the examples in (451) demonstrate.

A further problem arises when we consider the possibility of extracting an argument which has already been inverted. An *ay*-inverted noun phrase (which is always nominative) cannot be *wh*-extracted, whereas non-inverted nominative arguments are always extractable. (459a) is the unsuccessful attempt to extract the *ay*-inverted noun phrase from (449b), and (459b) likewise fails on the extraction from (455b).

(459) a. *Sino ang ay bumili ng libro.
who NOM AY buy-AV GEN-book
(for: *Who bought the book?*)

b. *Ano ang sinabi ni Pedro na ay ninakaw ni Josie.
what NOM say-OV GEN-Pedro COMP AY steal-OV GEN-Josie
(for: *What did Pedro say that Josie stole?*)

On a purely syntactic level, Nakamura's analysis appears to predict that *ay*-inverted arguments should be extractable, contrary to (459). This is because the inverted argument has moved to a functional projection even higher than SPEC of IP, and thus should be quite extractable according to Nakamura's Minimal Link Condition. The present analysis predicts that this asymmetry in the extractability of inverted and non-inverted arguments should arise as a basic syntactic effect of the directional slashes on categories. This is due to the fact that *ay* is a function which changes the direction in which the verb looks for its nominative argument, leading to the following directional mismatch between the result category of *ay* and the argument category of *ang*.

(460) *Sino [ang]$_{\mathsf{np}_n/_\star(\mathsf{s}/_\diamond\mathsf{np})}$ [ay bumili ng libro]$_{\ddot{\mathsf{s}}_c\backslash_\star\mathsf{np}_n}$

Note also that these categories cannot combine through backward crossed composition because neither category licenses crossed composition. In the restrictional CCG setting, it would be necessary to place a restriction on the backward crossed composition rule to block strings like *ang ay bumili ng libro lalaki* from being derived with the category $\ddot{\mathsf{s}}_c$.

The challenge to an analysis such as Nakamura's is to retain the similarity between the constraints on *ay*-inversion and *wh*-extraction in voiced contexts and long distance constructions without the overgenerations mentioned here. Nakamura could perhaps claim that the examples in (459) can be blocked because of clashes of pragmatic functions, but this avenue is unavailable for the examples in (451).

7.2 Toba Batak Asymmetries

Toba Batak is an interesting language to consider with respect to the discussion of asymmetries and how they arise in Tagalog and English.[6] Tagalog and Toba Batak are alike in many ways—most importantly, both are verb-initial and both have a rich voice system which allows verbs to distinguish one of their arguments. However, Toba Batak and English are alike in that they have relatively fixed word-order. As we shall see, Toba Batak and English exhibit asymmetries because of limitations on crossed compositions rules, whereas Tagalog's arise because of limitations on harmonic composition.

[6]Though I discuss only Toba Batak here, what is said for Toba Batak here appears to hold for Malagasy as well.

While Toba Batak has a voice system like Tagalog, it distinguishes the distinguished argument via word order rather than case marking.[7]

(461) a. Mangida si John si Bill
AV-see PNM-John PNM-Bill
Bill saw John.

b. Diida si Bill si John
OV-see PNM-Bill PNM-John
Bill saw John.

We see in (461) that it is the clause-final noun phrase that is distinguished—changing the order of the noun phrases would result in a reversed interpretation for both sentences. We give the following categories for the two voices of *ida* 'see' which will capture these facts as in derivations (463) and (464).

(462) a. $\mathit{mangida} \vdash (\mathsf{s}_i/_\diamond\mathsf{np}^{\mathbf{Act}})/_\diamond\mathsf{np}^{\mathbf{Pat}}$

b. $\mathit{diida} \vdash (\mathsf{s}_i/_\diamond\mathsf{np}^{\mathbf{Pat}})/_\diamond\mathsf{np}^{\mathbf{Act}}$

(463)
$$\dfrac{\dfrac{\dfrac{\mathit{mangida}}{(\mathsf{s}_i/_\diamond\mathsf{np}^{\mathbf{Act}})/_\diamond\mathsf{np}^{\mathbf{Pat}}} \quad \dfrac{\mathit{si\ John}}{\mathsf{np}^{\mathbf{Pat}}}}{\mathsf{s}_i/_\diamond\mathsf{np}^{\mathbf{Act}}}> \quad \dfrac{\mathit{si\ Bill}}{\mathsf{np}^{\mathbf{Act}}}}{\mathsf{s}_i}>$$

(464)
$$\dfrac{\dfrac{\dfrac{\mathit{diida}}{(\mathsf{s}_i/_\diamond\mathsf{np}^{\mathbf{Pat}})/_\diamond\mathsf{np}^{\mathbf{Act}}} \quad \dfrac{\mathit{si\ Bill}}{\mathsf{np}^{\mathbf{Act}}}}{\mathsf{s}_i/_\diamond\mathsf{np}^{\mathbf{Pat}}}> \quad \dfrac{\mathit{si\ John}}{\mathsf{np}^{\mathbf{Pat}}}}{\mathsf{s}_i}>$$

As before, the labeling of categories with dependency roles in no way performs grammatical work and is meant only to aid the readability of the derivations and stress that the categories are indeed capturing the correct dependencies in the undisplayed logical forms.

The $\diamond$ modality placed on the arguments of the categories for *mangida* and *diida* makes it impossible for the arguments to permute via type-raising and backward crossed composition, as the following blocked derivation demonstrates.

[7] All of my Toba Batak examples are adapted from Clark (1991).

(465)

$$
\begin{array}{ccc}
\textit{*mangida} & \textit{si Bill} & \textit{si John} \\
\hline
(\mathsf{s}_i/_\diamond\mathsf{np}^{\mathbf{Act}})/_\diamond\mathsf{np}^{\mathbf{Pat}} & \mathsf{np}^{\mathbf{Act}} & \mathsf{np}^{\mathbf{Pat}} \\
 & \overline{\ddot{\mathsf{s}}_i\backslash_\diamond(\ddot{\mathsf{s}}_i/_\diamond\mathsf{np}^{\mathbf{Act}})}^{<\mathbf{T}} & \\
\multicolumn{2}{c}{***\ \overline{\ddot{\mathsf{s}}_i/_\diamond\mathsf{np}^{\mathbf{Pat}}}^{<\mathbf{B}_\times}\ ***} & \\
\multicolumn{3}{c}{\overline{\ddot{\mathsf{s}}_i}^{>}}
\end{array}
$$

Thus, we cannot interpret *Mangida si Bill si John* as *Bill saw John.* This proves crucial for explaining the subject/object asymmetry of Toba Batak.

7.2.1 *Wh*-extraction

Just as Tagalog's asymmetries centered around the bifurcation between the distinguished argument and other arguments, the same is true in Batak, as the following data for *wh*-extraction illustrates.

(466)
a. Ise mangida si John?
who AV-see PNM-John
Who saw John?

b. *Ise diida si John?
who OV-see PNM-John
(for: *Who saw John?*)

c. *Aha mangida si Bill?
what AV-see PNM-Bill
(for: *What did Bill see?*)

d. Aha diida si Bill?
what OV-see PNM-Bill
What did Bill see?

The use of the $\diamond$ modality forced on the categories for the verbs by rigid word order requirements predicts this asymmetry. *Wh*-words have the category $\ddot{\mathsf{s}}_c/_\diamond(\mathsf{s}_i/_\diamond\mathsf{np})$, which is nearly the same as that for object-extracting *wh*-words in English. This immediately gives us the derivations for (466a,d).

(467)

$$
\begin{array}{ccc}
\textit{ise} & \textit{mangida} & \textit{si John} \\
\hline
\ddot{\mathsf{s}}_c/_\diamond(\mathsf{s}_i/_\diamond\mathsf{np}) & (\mathsf{s}_i/_\diamond\mathsf{np}^{\mathbf{Act}})/_\diamond\mathsf{np}^{\mathbf{Pat}} & \mathsf{np}^{\mathbf{Pat}} \\
 & \multicolumn{2}{c}{\overline{\mathsf{s}_i/_\diamond\mathsf{np}^{\mathbf{Act}}}^{>}} \\
\multicolumn{3}{c}{\overline{\ddot{\mathsf{s}}_c}^{>}}
\end{array}
$$

(468)

$$\begin{array}{ccc}
\textit{aha} & \textit{diida} & \textit{si Bill} \\
\hline
\ddot{s}_c/_{\diamond}(s_i/_{\diamond}np) & (s_i/_{\diamond}np^{\mathbf{Pat}})/_{\diamond}np^{\mathbf{Act}} & np^{\mathbf{Act}} \\
 & \multicolumn{2}{c}{\overline{}>} \\
 & \multicolumn{2}{c}{s_i/_{\diamond}np^{\mathbf{Pat}}} \\
\multicolumn{3}{c}{\overline{}>} \\
\multicolumn{3}{c}{\ddot{s}_c}
\end{array}$$

The most deeply nested syntactic argument of either verb can thus be easily extracted. Extracting the **Patient** argument of *mangida* as in (466c) would require deriving the category $s_i/_{\diamond}np^{\mathbf{Pat}}$ for the string *mangida si Bill* by using backward type-raising and backward crossed composition, but this has already been shown to be impossible in (465). A similar case holds for the **Actor** argument of *diida*. The categories already assumed for Toba Batak verbs in conjunction with the standard *wh*-extracting category thus predict the asymmetry of (466).

7.2.2 Adverb Placement

This compelling result receives further support from adverb placement. Adverbs such as *nantoari* 'yesterday' cannot appear between the verb and its first argument.

(469) a. *Mangida nantoari si John si Bill
AV-see yesterday PNM-John PNM-Bill
(for: *Bill saw John yesterday.*)

b. Mangida si John nantoari si Bill
AV-see PNM-John yesterday PNM-Bill
Bill saw John yesterday.

c. *Diida nantoari si Bill si John
OV-see yesterday PNM-Bill PNM-John
(for: *Bill saw John yesterday.*)

d. Diida si Bill nantoari si John
OV-see PNM-Bill yesterday PNM-John
Bill saw John yesterday.

The adverb placement data follows immediately if we give *nantoari* 'yesterday' the unsurprising category $(s/_{\diamond}np)\backslash(s/_{\diamond}np)$. The failure of (469a) and success of (469b) are given in (470) and (471), respectively.

(470)

$$\begin{array}{cccc}
\textit{*mangida} & \textit{nantoari} & \textit{si John} & \textit{si Bill} \\
\hline
(s_i/_{\diamond}np^{\mathbf{Act}})/_{\diamond}np^{\mathbf{Pat}} & (s/_{\diamond}np)\backslash(s/_{\diamond}np) & np^{\mathbf{Pat}} & np^{\mathbf{Act}} \\
\multicolumn{2}{c}{***\overline{}<\mathbf{B}_{\times}\ ***} & &
\end{array}$$

(471)

$$
\begin{array}{cccc}
\textit{mangida} & \textit{si John} & \textit{nantoari} & \textit{si Bill} \\
\overline{(\mathsf{s}_i/_{\diamond}\mathsf{np}^{\mathbf{Act}})/_{\diamond}\mathsf{np}^{\mathbf{Pat}}} & \overline{\mathsf{np}^{\mathbf{Pat}}} & \overline{(\mathsf{s}/_{\diamond}\mathsf{np})\backslash(\mathsf{s}/_{\diamond}\mathsf{np})} & \overline{\mathsf{np}^{\mathbf{Act}}} \\
\hline
\multicolumn{2}{c}{\mathsf{s}_i/_{\diamond}\mathsf{np}^{\mathbf{Act}} \quad >} & & \\
\hline
\multicolumn{3}{c}{\mathsf{s}_i/_{\diamond}\mathsf{np}^{\mathbf{Act}} \quad <} & \\
\hline
\multicolumn{4}{c}{\mathsf{s}_i \quad >}
\end{array}
$$

In (470), the adverb and the verb cannot combine due to the the fact that *mangida*'s category does not license backward crossed composition to push any constituents between it and its patient argument. However, once the verb has taken its first argument as in (471), it produces a result which the adverbial category may take via backward application.

Tagalog, on the other hand, allows adverbs to immediately follow the verb since the slash types assumed for Tagalog verbs license backward crossed composition.

(472) Tagalog:

Sumulat	kahapon	ng liham	kay Maria	si Juan
AV-write	yesterday	GEN-letter	DAT-Maria	NOM-Juan

Juan wrote a letter to Maria yesterday.

Assuming the following categories for *sumulat* and *kahapon* allows the derivation of (7.2.2) to go through.

(473) *sumulat* $\vdash \mathsf{s}_i\{/\mathsf{np}_n^{\mathbf{Act}}, /_{\times}\mathsf{np}_d^{\mathbf{Addr}}, /_{\times}\mathsf{np}_g^{\mathbf{Pat}}\}$

(474) *kahapon* $\vdash \mathsf{s}\backslash\mathsf{s}$

7.2.3 Extraction from Ditransitive Verbs

Further confirmation of this analysis for Toba Batak comes from ditransitive verbs. Clark (1985) shows that the **Addressee** and the distinguished argument may scramble with respect to each other, leading to the following two possibilites for an active voice ditransitive verb:

(475) a. Mangalean missel i tu soridadu jeneral i.

Mangalean	missel i	tu soridadu	jeneral i.
AV-give	missile-the	to soldier	general-the

The general gave the missile to the soldier.

b. Mangalean missel i jeneral i tu soridadu.

Mangalean	missel i	jeneral i	tu soridadu.
AV-give	missile-the	general-the	to soldier

The general gave the missile to the soldier.

The **Addressee** is preceded by *tu*, which I assume marks a noun with dative case. These sentences are simple to model in MULTI-MODAL CCG with the following category:

(476) *mangalean* $\vdash (\mathsf{s}_i/_{\diamond}\{\mathsf{np}^{\mathbf{Act}}, \mathsf{np}_d^{\mathbf{Addr}}\})/_{\diamond}\mathsf{np}^{\mathbf{Pat}}$

The two variations in (475) are obviously derivable using this single category.

An interesting consequence of this category assignment is that it correctly predicts that both the **Addressee** and the distinguished argument may be extracted.[8] The following data shows these possibilities.

(477)
a. Ise mangalean missel i tu soridadu?
who AV-give missile-the to soldier
Who gave the missile to the soldier?

b. Tu ise mangalean missel i jeneral i?
to whom AV-give missile-the general the
To whom did the general give the missile?

These questions have the following derivations, which show how the assumptions forced on the category of *mangalean* by (475) make both extractions of (477) possible.

(478)
$$\begin{array}{cccc}
\textit{ise} & \textit{mangalean} & \textit{missel i} & \textit{tu soridadu}\\
\hline
\ddot{\mathsf{s}}_c/_{\diamond}(\mathsf{s}_i/_{\diamond}\mathsf{np}) & (\mathsf{s}_i/_{\diamond}\{\mathsf{np}^{\mathbf{Act}}, \mathsf{np}_d^{\mathbf{Addr}}\})/_{\diamond}\mathsf{np}^{\mathbf{Pat}} & \mathsf{np}^{\mathbf{Pat}} & \mathsf{np}_d^{\mathbf{Addr}}\\
\cline{2-3}
 & \multicolumn{2}{c}{\mathsf{s}_i/_{\diamond}\{\mathsf{np}^{\mathbf{Act}}, \mathsf{np}_d^{\mathbf{Addr}}\}\ >} & \\
\cline{2-4}
 & \multicolumn{3}{c}{\mathsf{s}_i/_{\diamond}\mathsf{np}^{\mathbf{Act}}\ >}\\
\hline
\multicolumn{4}{c}{\ddot{\mathsf{s}}_c\ >}
\end{array}$$

(479)
$$\begin{array}{cccc}
\textit{tu ise} & \textit{mangalean} & \textit{missel i} & \textit{jeneral i}\\
\hline
\ddot{\mathsf{s}}_c/_{\diamond}(\mathsf{s}_i/_{\diamond}\mathsf{np}_d) & (\mathsf{s}_i/_{\diamond}\{\mathsf{np}^{\mathbf{Act}}, \mathsf{np}_d^{\mathbf{Addr}}\})/_{\diamond}\mathsf{np}^{\mathbf{Pat}} & \mathsf{np}^{\mathbf{Pat}} & \mathsf{np}^{\mathbf{Act}}\\
\cline{2-3}
 & \multicolumn{2}{c}{\mathsf{s}_i/_{\diamond}\{\mathsf{np}^{\mathbf{Act}}, \mathsf{np}_d^{\mathbf{Addr}}\}\ >} & \\
\cline{2-4}
 & \multicolumn{3}{c}{\mathsf{s}_i/_{\diamond}\mathsf{np}_d^{\mathbf{Addr}}\ >}\\
\hline
\multicolumn{4}{c}{\ddot{\mathsf{s}}_c\ >}
\end{array}$$

Of course, the **Patient** argument *missel i* cannot be extracted, for the same reasons as those discussed for transitive verbs. Similar effects can be seen for the Object Voice form *dilean* 'give', except with *dilean*, it is the **Actor** argument which cannot be extracted.

[8] I am grateful to Robin Clark for pointing out this consequence of the analysis.

7.3 Discussion

The analyses of Tagalog and Toba Batak presented in this chapter and of English in Chapter 4 demonstrate that syntactic asymmetries can be captured in the lexicon and that their unbounded realizations are mediated by the directionally and modally sensitive rules of MULTI-MODAL CCG. They furthermore show that local scrambling and asymmetries in Tagalog can be dealt with on separate tiers. In a phrase structural analysis like that of Guilfoyle *et al.* (1992), extractability is tied to the ability to be in or move to particular structural positions, a solution which forces them to assume that there are certain "basic" word orders. As Kroeger (1993) points out, their highly articulated phrase structure would require powerful scrambling mechanisms to handle even the basic local scrambling examples shown in (152). Instead of singling out a particular argument through a special structural position like SPEC of VP or SPEC of IP, the present proposal bestows that argument with greater access to the combinatory rules of MULTI-MODAL CCG, specifically the harmonic composition rules.

The assumptions about the Tagalog lexicon needed to capture local extraction asymmetries explain the constraints on long distance extraction and *ay*-inversion. Furthermore, it is shown that the very same mechanisms are responsible for both long distance extraction asymmetries and inversion asymmetries. The lexicon is very small and admits of very little lexical ambiguity while providing the most complete coverage of Tagalog asymmetry phenomena to date. The analysis thus provides an extensive model of the data while making no claims as to *why* the extracting categories of Tagalog require associatively licensed arguments. This is left open for different learning theories to be applied to the analysis, including a declarative Principles and Parameters style theory of the lexicon, a machine learning algorithm, or some combination of both. The learner would be supported by the fact that every extracting category of the Tagalog lexicon demands an argument which is a category missing a associatively licensed argument, as the list in (480) shows.

(480) a. *ang* $\vdash \mathsf{np}_n /_{\star} (\mathsf{s} /_{\diamond} \mathsf{np})$

b. *ay* $\vdash (\mathsf{s} \backslash_{\star} \mathsf{np}_n) /_{\star} (\mathsf{s} /_{\diamond} \mathsf{np}_n)$

c. *saan* $\vdash (\ddot{\mathsf{s}}_c /_{\diamond} \ddot{\mathsf{s}}_c) /_{\star} (\mathsf{s}_i /_{\diamond} \ddot{\mathsf{s}}_c)$

d. *na/ng* $\vdash (\mathsf{s}_v / \mathsf{np}_n) \{ /_{\star} (\mathsf{s} /_{\diamond} \mathsf{np}), \backslash_{\star} (\mathsf{s}_v /_{\diamond} \mathsf{np}) \}$

Though I have not discussed them here, raising predicates and floated quantifiers also take the category $\mathsf{s}/_{\diamond}\mathsf{np}$ as an argument. Thus, we see extensive generalization of this kind in the lexicon, as we would expect a language learner constructing a lexicon to do. The core asymmetries of Tagalog then follow from this generalization on second order categories in combination with the way such categories license the combinatory rules.

English and Toba Batak cannot fully license crossed composition because of word order constraints, and this in turn produces the asymmetries we saw in those languages. What is not yet clear is exactly why Tagalog places limits on *harmonic* composition. One possible explanation comes from long distance scrambling. My Tagalog informants indicate that Tagalog appears to have very limited long distance scrambling, with some speakers allowing the nominative argument to scramble out of the clause. If all arguments except for the distinguished one are trapped below the complementizer (see (436)) long distance scrambling will be very limited indeed. This supports the present analysis' use of the the non-harmonic modality $\times\!\!\diamond$ on the slashes of non-nominative arguments in Tagalog, thereby limiting their long distance scrambling potential in addition to blocking them from composing harmonically as in (436).

Tagalog is not alone in limiting the applicability of harmonic composition. Trechsel (2000) provides a CCG analysis of extraction and pied-piping in Tztotzil which requires the use of a feature [±FC] and a definition of forward harmonic composition which permits only functors which are +FC to serve as input to it. This feature provides a similar effect to the modalities of MULTI-MODAL CCG (though without the explicit logical foundations), and it is crucial for restricting the extraction of possessors of subjects of unergative verbs. Trechsel's analysis could thus be readily translated into the MULTI-MODAL CCG framework without the stipulative [±FC] feature, and it provides another cross-linguistic data point backing up the present characterization of syntactic asymmetries.

It should also be pointed out that the basic AB system upon which MULTI-MODAL CCG is built is itself responsible for certain extraction facts. (460) demonstrates a directional mismatch which causes category unification failure and leads to the unextractability of an *ay*-inverted argument. *Ay*-inversion turns a verb with a VSO category into one with an SVO category, and we discover that the inverted constituent cannot be extracted. This is just what we would expect for English relativization if English, with its SVO categories, lacked the leftward (subject) extracting category given in (114a). Tagalog, of course, has no such category (such as $\mathsf{np}_n/_{\star}(\mathsf{s}\backslash\mathsf{n})$) and thus we can never extract an *ay*-inverted constituent.

The data regarding ditransitives in Toba Batak provides extra support for the MULTI-MODAL CCG generalization of CCG. The distinguished argument and the **Addressee** could scramble with respect to each other, but other than that, word order variability is quite restricted. Putting both arguments in a set labeled with the $\diamond$ modality permits *limited* local scrambling, which in turn predicts the consequent extractability of both arguments.

This chapter demonstrates the linguistic applicability of MULTI-MODAL CCG. In addition to inheriting most of the desirable aspects of standard CCG, it permits an explanatory analysis of asymmetries from a cross-linguistic perspective while using very restricted and well-motivated formal devices. The modalities of MULTI-MODAL CCG, which hold the primary responsibility for modeling asymmetries, are based on fine-grained structural rules from Categorial Type Logic which permit us to examine closely the effects which they license. The sets are based on MULTISET-CCG and its account of scrambling, but the limitations placed on them by the MULTI-MODAL CCG rules and their modalities permit their use in languages like Tagalog which exhibit local scrambling and only limited long distance scrambling. Applying MULTISET-CCG itself to Tagalog would quickly lead to a multitude of rule restrictions to keep the grammar from overgenerating, but with MULTI-MODAL CCG, no such restrictions are necessary.

Chapter 8

Implementation of Multi-Modal CCG

In this dissertation, I have thus far pursued two distinct yet complementary threads: formal modifications to the CCG formalism and linguistic analyses of phenomena in several languages couched within the modified framework, Multi-Modal Combinatory Categorial Grammar. Because of CCG's computationally attractive properties, it is important to consider the effects that the proposed modifications have on the the ability to implement and use CCG in a computational setting. In this chapter, I consider this issue in the context of Grok, a CCG-based NLP system that I have extended to support the categories and rules of MULTI-MODAL CCG. I also discuss the implementations of the linguistic analyses proposed in this dissertation for English, Dutch, Turkish, and Tagalog, which I have used to validate the predictions and coverage of the analyses.

I begin with a brief overview of previous implementations of CCG and then discuss how I have adapted the Grok system to support MULTI-MODAL CCG grammars and some of the consequences of doing so. After that, I turn to the grammars which I have implemented based on the linguistic analyses proposed in this dissertation and provide some example interactions with Grok.

8.1 CCG Parsing

Due to their semantic transparency and precise formalization (Wood, 1993), one of the great appeals of categorial grammars has always been their amenability to computational implementation. The fact that categorial grammars are typically lexicalized also confers many benefits for using them for natural language processing (NLP) – in particular, it means that the search space for parsing is limited to the categories which have been licensed by some input word (König, 1999). CCG has a polynomial (n^6) worst-case parsing algorithm (Vijay-Shanker and Weir, 1990), and Komagata (1999) has reported average case parsing times around n^3 for parsing restricted English and Japanese medical texts using a worst-case exponential chart parser. CCG's flexible constituency, while responsible for problems of spurious ambiguity, permits Steedman (2000b) to give an algorithm for processing sentences incrementally, thus providing the basis for a psychologically motivated parser.

A number of CCG-based grammars and parsing systems have been developed over the years, each with its own emphasis on either examining particular properties of CCG or using the framework as a means to other ends. Much of the focus has centered on how lexicons are defined and created, and systems range in the amount of human labor that is needed to produce their grammars. Wittenburg (1986, 1987) explores the use of *predictive combinators* to limit the non-determinism induced by spurious ambiguities and presents a parser that takes advantage of them. Komagata (1999) and Bierner (2001) both describe systems which use hand-crafted lexicons that produce fine-grained semantic analyses of sentences. Villavicencio (1997) and Doran and Srinivas (2000) semi-automatically translate large hand-crafted lexicons from other frameworks to produce CCG lexicons. More recently, Julia Hockenmaier has produced the CCGBank (Hockenmaier *et al.*, 2001; Hockenmaier and Steedman, 2002a), a version of the Penn Treebank (Marcus *et al.*, 1993) that replaces tree structures with normal-form CCG derivations and which has been utilized to create a wide-coverage CCG lexicon and to estimate models for statistical parsing of CCG (Hockenmaier and Steedman, 2002b; Clark *et al.*, 2002). Also, Hockenmaier *et al.* (2001) and Bierner (2001) discuss techniques for combining hand-crafted lexicons with corpus-based ones in order to take advantage of the semantic detail of the former and the wide coverage of the latter.

8.2 Adapting Grok for Multi-Modal CCG

The Grok system described in Hockenmaier *et al.* (2001) and Bierner (2001) was originally designed to work with standard CCG grammars, and I have extended it to support MULTI-MODAL CCG as well. Rather than serving as the basis of a practical application, the goal of the implementation is to demonstrate the computational suitability of MULTI-MODAL CCG and to explore the ways in which its properties can be exploited to improve the declarations of grammars and take advantage of some high-level parsing strategies.

In this section, I discuss some of the more interesting aspects of the implementation provided in Grok. I begin with a brief history of the development of Grok and then discuss the ways in which I have extended it for MULTI-MODAL CCG. These extensions include the modalization of slashes, the use of $'s and multisets in category structures, the use of hard-coded combinatory rule implementations, and finally the incorporation of flexible semantic representations and support for declaring lexical inheritance hierarchies. The section ends with mention of a few other aspects of Grok that bear on the grammatical architecture put forth in this dissertation.

8.2.1 A Brief Historical Note on Grok

Grok is a Java-based natural language processing system that has gone through several phases of development. In 1998, Gann Bierner and I began writing Grok as a small parsing system for simple CCG grammars and over the following year it took on many of the architectural aspects of the XTAG system (Doran *et al.*, 2000). We wrote a graphical user interface and implemented small scale modules for syntactic, semantic, and discourse analysis that served as a useful infrastructure for experimenting with different aspects of the computational analysis of language. During this time, Grok was released as an open source system under the Lesser GNU General Public License and early in 2000 the Grok website and code was hosted on the Sourceforge site for open source software development (`http://sourceforge.net`).

During 2000, we sought to improve Grok's robustness and coverage. One aspect of this involved developing a pre-processing architecture that allowed many low-level components such as tokenizers, sentence splitters, part-of-speech taggers, and name detectors to be hooked together in a pipeline that marked up text to prepare it for parsing.[1] We then developed components to allow the Grok English lexicon to be merged with the CCG lexicon which Julia Hockenmaier had produced semi-

[1] At that time, the GATE system (Cunningham, 2000) was not open source and its architecture was too centralized, so we chose to begin a new architecture — see Bierner (2001), Chapter 5, for a discussion of this issue. Today, GATE is open source, uses a pipeline architecture, and is more general and far better documented than Grok's preprocessing components. Thus, a number of Grok's components have been migrated to the GATE architecture, and probably the rest will eventually follow.

automatically from the Penn Treebank, resulting in a grammar that could parse a wide-range of sentences and return predicate-argument structures which captured not only high-level dependencies, but also the fine-grained interpretations created from categories in the hand-crafted lexicon. The results of this work were reported in Hockenmaier *et al.* (2001) and Bierner (2001), with the former focusing on the creation of the wide-coverage lexicon and the latter on Grok itself and its use in the computational analysis of alternative phrases (e.g. ***other*** *people* ***than*** *Lee, Sue, Scott, and Lisa*, ***such as*** *Fred, Mary, and Justin* and ***besides*** *Fernanda*.

Since mid-2001, development of Grok has been almost entirely in my hands, and I started by stripping out the discourse-oriented components and paring the system down to be used only as a CCG parsing engine. I also removed the components for interfacing with the corpus-based lexicon. Thus, it is important to note that when I speak of Grok in this chapter, I am referring to the reduced system, which no longer supports much of the behavior reported in Hockenmaier *et al.* (2001) and Bierner (2001). During the past year, I have re-implemented parts of the system and added new capabilities as discussed below.

8.2.2 Modalized Slashes

One of the most straightforward modifications to Grok's data structures to support MULTI-MODAL CCG was to add modalities to the slashes. Modalities are an attribute of the data structure for slashes and can be either an actual modality or a variable standing for one. Modality variables are needed for type-raised categories (see the modalized type-raising rules given in (196)). The hierarchy of modalities used in this dissertation (see Figure 5.1, page 100) is hard-coded into the implementation, which handles their unification behavior appropriately. It should, however, be straightforward to generalize this to permit hierarchies to be defined in a more modular fashion.

Another way in which slashes were modified was the addition of another attribute that characterizes a slash as *inert* or *active*. This distinction is used to implement the antecedent government feature ±ANT discussed in §4.1.1 (page 74), §5.2.6, and §5.2.7. In rule applications for which an "antecedent-governed" argument arises, it is handled by making its slash inert rather than by unifying its feature structure with the feature +ANT. The implementations of the combinatory rules then ensure that the slashes of the functor category (e.g. the slash of $\mathsf{X/Y}$ in the forward composition rule) is not inert. By dealing with antecedent-government behavior in this way, it is unnecessary to declare all lexical categories as -ANT, and it more directly reflects the fact that the CCG notion of antecedent-government is one which mediates structural configurations and combinatory potential.

The fact that the data structure for slash has become more articulated in two dimensions – modality and active/inert – points to the possibility to break the modalities down into ever more specific behaviors that control the combinatory potential of categories along many, potentially exclusive, dimensions. For example, we could add another modality for representing headedness, along the lines of Kruijff (2001), in a way that does not complicate the use of the modalities which control the combinatory rules. Also, it could be useful to have a distinction between *lexically connected* slashes such as that in the category $\mathsf{s/n}$ of *Brazil defeated* and *derivationally connected* slashes such as that in the category $\mathsf{s/n}$ of *I think that Brazil defeated.* Of course, any such attributes on slashes must be non-recursive data structures like modalities to ensure that the power of the system is not unexpectedly boosted.

The modalization of slash with CCG is already being utilized outside of Grok. Beavers (in progress) describes the implemention of a CCG grammar in the LKB system, and he uses two modalities to encode the headedness of categories to control the application of lexical rules. This provided a simple and effective solution to a problem in which a lexical rule for auxiliary verbs applied erroneously to adverbs since the categories for both were structurally identical ($(\mathsf{s}\backslash\mathsf{np})/(\mathsf{s}\backslash\mathsf{np})$). By defining the lexical rule to apply only to heads (e.g. $(\mathsf{s}\backslash\mathsf{np})/_{\rightarrow}(\mathsf{s}\backslash\mathsf{np})$), adjuncts ((e.g. $(\mathsf{s}\backslash\mathsf{np})/_{\leftarrow}(\mathsf{s}\backslash\mathsf{np})$) were appropriately excluded.

8.2.3 Category Data Structures

Categorial grammar implementations typically utilize a curried representation of complex categories. While this has certain advantages with respect to the simplicity of basic unification regimes, it is more limited for dealing with categories that utilize $'s (Convention 2, page 26) and multisets (Definition 2, page 35). Although it is possible to hide them from the parser by interpreting them as schematizations that generate multiple standard categories, we can also consider using them directly in the categories that are fed to the parser. If we take this option, then the data structures for complex categories become somewhat more complex and it becomes advantageous to use a non-curried representation. Categories can be construed as an atomic result category plus a stack of arguments (e.g. $\mathsf{s}, [\backslash \mathsf{n}_1, /\mathsf{n}_2, /\mathsf{n}_3]$ rather than as successive result categories terminating in an atomic result $((\mathsf{s}\backslash \mathsf{n}_1)/\mathsf{n}_2)/\mathsf{n}_3$.

The first reason to use non-curried representations is that \$'s act as stack variables which take a slice out of a complex category's argument list when unified against it. Although it is possible to implement this behavior with a curried representation, it is more complicated. For example, if we wish to unify $\mathsf{s}\$_0/\mathsf{n}_3$ and $((\mathsf{s}\backslash \mathsf{n}_1)/\mathsf{n}_2)/\mathsf{n}_3$, the $\$_0$ variable unifies with the stack $[\backslash \mathsf{n}_1, /\mathsf{n}_2]$. If we use a curried representation, we must iteratively peel off each of the arguments in the stack to get the unification result. A non-curried representation, on the other hand, already has the entire argument list as a stack such as $[\backslash \mathsf{n}_1, /\mathsf{n}_2, /\mathsf{n}_3]$, and the $\$_0$ variable can be directly unified against the sub-stack consisting of the first two elements. Also, \$ variables can unify with the empty stack, which is simpler to handle with the non-curried representation.

Multisets lend themselves naturally to stack-like structures. Thus, we can represent a multiset in a category like $(\mathsf{s}_i\{/_\diamond \mathsf{np}^{\mathbf{Act}}, /_\diamond \mathsf{np}^{\mathbf{Addr}}_{dat}\})\{/_\diamond \mathsf{np}^{\mathbf{Pat}}\}$ (the Toba Batak ditransitive category) as stacks within a stack: $\mathsf{s}_i[[/_\diamond \mathsf{np}^{\mathbf{Act}}, /_\diamond \mathsf{np}^{\mathbf{Addr}}_{dat}], /_\diamond \mathsf{np}^{\mathbf{Pat}}]$. It is then possible to easily scan through a multiset and pluck out and/or replace particular elements without massive restructuring of the overall category, especially during rule application.

Though it would be more general to place all arguments in multisets so that the argument list is a stack of stacks (e.g. $\mathsf{s}_i[[/_\diamond \mathsf{np}^{\mathbf{Act}}, /_\diamond \mathsf{np}^{\mathbf{Addr}}_{dat}], [/_\diamond \mathsf{np}^{\mathbf{Pat}}]]$), I chose not to do so to make the data structures simpler for others who do not need to work with multisets. Because both simple and stack arguments are available in the system, whenever a multiset is reduced to one element, it is converted into a simple argument.

A final reason to use non-curried representations is that they provide immediate access to the atomic result category, which proves quite useful when applying the combinatory rules to pairs of categories. Before attempting to unify all of the arguments of two categories together, we can first check that the result categories are an appropriate match.

While utilizing \$'s and multisets directly in categories can reduce categorial ambiguity considerably, it does complicate the unification procedures, especially if we wish to make unification of categories containing them entirely general. In the Grok implementation, I have thus made several simplifying assumptions about the kinds of places in which \$'s may appear based on their actual use in grammars. These simplifications are not particularly interesting in and of themselves, but it bears pointing out for anyone interested in working with or extending the implementation. Note that a strategy which creates multiple standard categories from categories containing multisets is quite similar to the way in which Foster (1990) projects ordered categories from unordered ones.

8.2.4 Hard-coded Combinatory Rules

Perhaps the most important feature of MULTI-MODAL CCG for implementation is that it has an invariant rule set that is keyed to particular modalities, which leads to several desirable properties.

First, it becomes possible to use hard-coded rules that are active and unmodified for every grammar. This was not possible before since non-trivial restrictions could be placed on rules for each grammar, and some linguistic analyses even require that the structure of the categories themselves be modified. As an example of a particularly complex restriction that requires modification to the rule categories themselves, Steedman (2000b) defines the rule of forward harmonic composition for Dutch given in (282), repeated here as (481).

(481) *Dutch Forward Harmonic Composition* ($>\mathbf{B}$)

$$\mathsf{X/Y} \quad \mathsf{Y/(Y\backslash Z)} \quad \Rightarrow_{\mathbf{B}} \quad \mathsf{X/(Y\backslash Z)}$$

where $\mathsf{Y} = \mathsf{s}\backslash\$$

To support this kind of variance in Grok, rules were previously declared for each grammar as sets of categories containing variables which were potentially restricted to particular types. These categories were then unified against input categories during parsing, and variables needed to carry their restrictions along with them.

With modally refined rules that carry no restrictions, the functionality of each rule can instead be coded directly into procedures which attempt to combine categories. This has several ramifications. One is that it is no longer necessary to fully unify the input categories against variable-containing rule categories — instead, only sub-parts of categories need to be unified. For example, when trying to compose $(\mathsf{s}\backslash\mathsf{n})/\mathsf{s}_{\pm SPEC}$ with $(\mathsf{s}_{-SPEC}/\mathsf{n})/\mathsf{n}$, rather than unifying these input categories with $\mathsf{X/Y}$ and $(\mathsf{Y/Z})/\$$, respectively, it is only necessary to unify the argument category $\mathsf{s}_{\pm SPEC}$ of the primary functor and the result category s_{-SPEC} of the secondary functor (the categories that match up with the "Y" of the rule). Having performed this unification, the outcome of the rule (i.e. $(\mathsf{X/Z})/\$$) is constructed by merging the prefix of the primary functor (e.g. $\mathsf{s\backslash n}$) with the suffix of the secondary one (e.g. $\mathsf{/n/n}$). Finally, the outcome category (e.g. $\mathsf{((s\backslash n)/n)/n}$) is then filled with the new values of any variables that were resolved during the unification of the "Y" portions of the input categories.

This strategy has the further effect of automatically generalizing the combinatory rules – without the need for stack variables. To generalize rules using customized rules for each grammar, it is necessary to create several rules from a single schema. For example, the schema for generalized forward composition given in (482) would produce the rules in (483) if composition is bounded at three.

(482) $\mathsf{X/Y \quad (Y/Z)/\$_1 \quad \Rightarrow_{\mathbf{B}^n} \quad (X/Z)/\$_1}$ $(>\mathbf{B}^n)$

(483) a. $\mathsf{X/Y \quad Y/Z_1 \quad \Rightarrow_{\mathbf{B}^1} \quad X/Z_1}$ $(>\mathbf{B}^1)$

b. $\mathsf{X/Y \quad (Y/Z_1)/Z_2 \quad \Rightarrow_{\mathbf{B}^2} \quad (X/Z_1)/Z_2}$ $(>\mathbf{B}^2)$

c. $\mathsf{X/Y \quad ((Y/Z_1)/Z_2)/Z_3 \quad \Rightarrow_{\mathbf{B}^3} \quad ((X/Z_1)/Z_2)/Z_3}$ $(>\mathbf{B}^3)$

With the hard-coded rules, we can thus focus on encoding the core behavior of a rule, and it is thus unnecessary to use *multiple* rule instances to handle peripheral aspects of that functionality, such as the size of the stack of the secondary functor. It also removes the need for a clean-up rule to get rid of empty sets (see the discussion below the multiset-based application rules (359) on page 143).

Another way in which the hard-coded rules reduce the need to define multiple rule instances is that it is not necessary to provide implementations of separate rules to handle the behavior of antecedent-government. Even though I have listed them separately in the text, the group of four rules given in (223) on page 109 for backward crossed composition and the group of four rules given in (235) on page 112 for forward crossed composition can each be handled by a single rule implementation. Since the functionality is the same except for the possibility that certain modalities trigger antecedent-government, the effects of the rules which invoke antecedent-government are handled as a check for the relevant modalities *after* the rule has otherwise succeeded. If any trigger modalities are present in the relevant positions, the slashes in the remainder stack are made inert as described above in §8.2.2.

Another advantage of hard-coded rules that are modally sensitive is that it is often possible to fail on rule applications much more quickly than is otherwise possible. The first thing that the rule implementation do is to check that the slashes of the input categories are compatible. With rules that naively unify rule categories with input categories, it is common to do an entire unification of the $\mathsf{X/Y}$ portion with the primary functor (including variable restrictions) before having the chance to see if the secondary functor even has the right slash to make the rule applicable.

Multisets can be difficult to handle efficiently with naive rules. The hard-coded rules, on the other hand, can invoke procedures for scanning multisets for the appropriate arguments depending on the other inputs, thereby considerably reducing the indeterminacy introduced by multisets. Even a very simple example can demonstrate this – consider using forward application to combine $\mathsf{s}\{/\mathsf{n}, /\mathsf{s}\}$ and s. If we unify the functor category with that of the rule itself, $\mathsf{X}(\alpha \uplus \{/_{\!.}\mathsf{Y}\})$, we must try two different unification paths – one where $\mathsf{Y}{=}\mathsf{s}$ and the other where $\mathsf{Y}{=}\mathsf{n}$. However, we can cut this down by first inspecting the secondary category and then looking for matches in the functor's set. The same strategy is utilized in the implementation of composition. This is simple to handle with the declarative procedures that encode the rules' behaviors, and the ability to write such procedures is dependent on the invariant rule base provided by MULTI-MODAL CCG.

The implementations of the rules thus handle different category combinations as different cases, and they attempt to utilize the particular aspects of each case to try to fail unifications more quickly or to cut out some of the possible avenues of unification. This does involve a fair amount of complexity in the procedures which implement the behavior of the rules; however, the commonalities of the rules generated for each combinator can be exploited so that a single procedure can be written for each combinator. For example, the specific instantiations for the forward and backward varieties of harmonic and crossed composition are created by supplying the appropriate slashes and locations of the primary and secondary functors as parameters to the general procedure for composition. This still results in individual rule instantiations when parsing, but it considerably reduces the programming complexity.

8.2.5 Flexible Semantic Structures

Most work in categorial grammar utilizes simply-typed λ-calculus expressions to represent the semantic interpretations produced during derivations. While these expressions are useful in demonstrating that a given derivation produces an interpretation with the correct properties, they are deficient in several ways. Recent work has highlighted its inadequacies for both linguistic (Kruijff, 2001) and computational concerns (Copestake *et al.*, 1999, 2001) of representing natural language semantics. Minimal Recursion Semantics (MRS) (Copestake *et al.*, 1999, 2001) is a framework for computational semantics that is designed to simplify the work of algorithms which produce or use semantic representations. MRS provides the means to represent interpretations with a flat, underspecified semantics using terms of the predicate calculus and generalized quantifiers. Copestake et al. argue that these flat representations facilitate a number of computational tasks, including machine translation and generation, without sacrificing linguistic expressivity. Also, flatness permits semantic equivalences to be checked more easily than in structures with deeper embedding, and underspecification simplifies the work of the parser since it does not have to compute every possible reading for scope-bearing elements.

Kruijff (2001) couples a CTL-based grammatical framework to hybrid logic (Blackburn, 2000) to formalize a dependency-based perspective on meaning called Hybrid Logic Dependency Semantics (HLDS). As a brief example of the kind of structures involved with HLDS, the following is a simplified representation of the sentence *the referee gave Ronaldinho a red card*:

(484) $$\begin{aligned} @_{h_1}(\mathbf{give} &\wedge \langle \textsc{Actor}\rangle(d_0 \wedge \mathbf{referee}) \\ &\wedge \langle \textsc{Patient}\rangle(d_5 \wedge \mathbf{card} \wedge \langle \textsc{GenRel}\rangle(d_7 \wedge \mathbf{red})) \\ &\wedge \langle \textsc{Addressee}\rangle(d_9 \wedge \mathbf{Ronaldinho})) \end{aligned}$$

The hybrid logic nominals (e.g. h_1 and d_0) act as discourse referents for the entities or events whose meaning they are tied to. The modal relations (e.g. $\langle \textsc{Actor}\rangle$) explicitly encode the named dependency relations of FGD (Sgall *et al.*, 1986) and relate heads such as **give** to their dependents.

In addition to encoding dependency relations and predicate-valency structures, HLDS can represent other aspects of sentential meaning such as spatio-temporal structure, contextual reference, and information structure. Kruijff (2001) also shows how HLDS can be used to model discourse interpretation, and thereby cover the track from grammar to discourse with a *single* meaning formalism. Though HLDS was not originally motivated by concerns of underspecification and flexibility, Baldridge and Kruijff (2002) show how HLDS nonetheless fulfills the criteria laid out by (Copestake *et al.*, 1999) for a computational semantics framework. Like the representations of MRS, HLDS terms can be split apart into lists of elementary predications that contain potentially underspecified scope bearing elements.

Baldridge and Kruijff (2002) also demonstrate how HLDS terms can be produced with CCG, and as part of this work, I added the ability in Grok to use terms of HLDS in semantic representations. The main relevance of this for the present discussion is that the flexibility of these terms made it possible to support the use of lexical inheritance hierarchies in Grok lexicons, as described in the following sections. Also, the implementation of HLDS provides quick and simple hashing functions that are used for improving subsumption checks (Karttunen, 1989), which are used to block new subconstituents from being added to the parse chart if an equivalent constituent has already been built. This significantly reduces the proliferation of spurious ambiguities during parsing.

8.2.6 Lexical Inheritance Hierarchies

As discussed in §3.2, Villavicencio (2002) uses an inheritance hierarchy, shown in Figure 3.2, to encode the systematic relationships between verbs of different arity. In this section, I discuss a few problems with Villavicencio's assumptions and show how multiset categories provide the means to allow the variation needed for languages like Turkish, Tagalog, and Toba Batak.

When one category extends another, Villavicencio adds a new argument to the end of the subcategorization list. While this provides a very simple account of how a category subtype expands its parent's specification, it relies on the convenient fact that the categories for English verbs generally follow this pattern and on Villavicencio's use of the rule of Generalized Weak Permutation (Briscoe, 1997) to reorder the arguments in case the category specifies a canonical order that is different from some of the potential surface orders. For example, the category of English ditransitive verbs extends the transitive category by adding a syntactic argument corresponding to the **Addressee**:

(485) $\underbrace{((\mathsf{s}\backslash\mathsf{np}^{\mathbf{Act}})/\mathsf{np}^{\mathbf{Pat}})}_{\textit{inherited}}\underbrace{/\mathsf{np}^{\mathbf{Addr}}}_{\textit{added}}$

Villavicencio's lexical rule for dative shift applies to this category to create the following category:

(486) $((\mathsf{s}\backslash\mathsf{np}^{\mathbf{Act}})/\mathsf{np}^{\mathbf{Pat}})/\mathsf{pp}^{\mathbf{Addr}}$

This category clearly will allow sentences such as *the referee gave to Ronaldinho a red card*, and Generalized Weak Permuation is invoked for the more standard order *the referee gave a red card to Ronaldinho*:

(487)
$$
\begin{array}{cccc}
\textsf{The referee} & \textsf{gave} & \textsf{a red card} & \textsf{to Ronaldinho} \\
\hline
\mathsf{np^{Act}} & \mathsf{((s\backslash np^{Act})/np^{Pat})/pp^{Addr}} & \mathsf{np^{Pat}} & \mathsf{pp^{Addr}} \\
 & \overline{\mathsf{((s\backslash np^{Act})/pp^{Addr})/np^{Pat}}}{>}\mathsf{GWP} & & \\
 & \multicolumn{2}{c}{\overline{\mathsf{(s\backslash np^{Act})/pp^{Addr}}}{>}} & \\
 & \multicolumn{3}{c}{\overline{\mathsf{s\backslash np^{Act}}}{>}} \\
\multicolumn{4}{c}{\overline{\mathsf{s}}{<}}
\end{array}
$$

Even though both orders are grammatical in English, the use of Generalized Weak Permutation will overgenerate for the original ditransitive category by providing a derivation for *The referee gave Ronaldinho a red card* in which *Ronaldinho* is interpreted as the **Patient** (i.e. the thing given) instead of as the **Addressee**. Though this strategy is sufficient for Villavicencio's purpose of studying language acquisition with English data, the kind of overgeneration that it engenders is unacceptable from the perspective of the present work, which aims to provide extremely discriminating analyses by strictly regulating the associative and permutative aspects of the grammar.

If we consider a VSO language, the use of some sort of category restructuring rule like Generalized Weak Permutation is required if the grammar is to be defined in terms of a simple definition of how category subtypes extend their parents. The (simplified) categories for intransitive and transitive verbs in a VSO language are those shown in (488).

(488) a. Intransitive: $\mathsf{s/np^{Act}}$

b. Transitive: $\mathsf{(s/np^{Pat})/np^{Act}}$

Simply adding the new argument of the transitive category to the end of the argument list of the intransitive results in the category $\mathsf{(s/np^{Act})/np^{Pat}}$, which encodes VOS order. Generalized Weak Permutation can restructure this category so that it becomes (488b) in order to obtain VSO order, but even though this would potentially work in a permutation friendly verb-initial language like Tagalog, it will be quite unsuitable for Toba Batak, which has quite rigid word order with respect to its **Actor** and **Patient** arguments (see §7.2).

Fortunately, Villavicencio's approach can be extended to work with multiset categories. By parameterizing how one category extends another, we can accommodate VSO languages comfortably within Villavicencio's suggested inheritance hierarchy. Thus, in order to define the category of the transitive, the intransitive can be extended by adding the new argument to either the beginning or the end of the argument list for any given grammar – for a VSO language we add to the beginning, and for VOS we add to the end.

It is also interesting to consider how subtype categories are created for argument permuting languages like Turkish and Tagalog. The multiset categories of MULTI-MODAL CCG provide another parametric option to that described above: new arguments can be added to not only the beginning or the end, but they also can be *inserted* into the beginning or ending multiset of the parent category. For example, the Turkish intransitive is $\mathsf{s}_i\{\backslash_{\cdot}\mathsf{np}_{nom}\}$, and the transitive category extends that by placing the accusative argument in the same multiset as the nominative one to create $\mathsf{s}_i\{\backslash_{\cdot}\mathsf{np}_{nom}, \backslash_{\cdot}\mathsf{np}_{acc}\}$. Toba Batak provides a more interesting example with respect to extending the transitive category to create the ditransitive one. If we consider Active Voice verbs, we see in (489) that the order of the **Actor** and **Patient** arguments matters in Toba Batak – changing the order of the arguments changes the interpretation. However, the examples in (490) show that the **Actor** and **Addressee** arguments can be permuted without changing the interpretation.

(489) a. Mangida si John si Bill
AV-see PNM-John PNM-Bill
Bill saw John.

b. Mangida si Bill si John
AV-see PNM-Bill PNM-John
John saw Bill.

(490) a. Mangalean missel i tu soridadu jeneral i.
AV-give missile-the to soldier general-the
The general gave the missile to the soldier.

b. Mangalean missel i jeneral i tu soridadu.
AV-give missile-the general-the to soldier
The general gave the missile to the soldier.

As discussed in §7.2, the transitive Active Voice category for Toba Batak is (491a), and the ditransitive is (491b). The categories are written in full, without the notational conventions that I generally use, to emphasize the multiset categories.

(491) a. $(\mathsf{s}_i\{/_{\diamond}\mathsf{np}^{\mathbf{Act}}\})\{/_{\diamond}\mathsf{np}^{\mathbf{Pat}}\}$

b. $(\mathsf{s}_i\{/_{\diamond}\mathsf{np}^{\mathbf{Act}}, /_{\diamond}\mathsf{np}^{\mathbf{Addr}}_{dat}\})\{/_{\diamond}\mathsf{np}^{\mathbf{Pat}}\}$

It is thus apparent that providing an inheritance based account of the regularities in the categories of Turkish, Tagalog, and Toba Batak requires allowing at least four options for extending parent categories – (a) adding at the beginning of the argument list, (b) inserting into the multiset at the beginning of the list, (c) inserting into the multiset at the end of the list, or (d) adding at the end of the list, as shown below:

(492) $\underset{\substack{\uparrow\\ a}}{\mathsf{root}}\{\mathsf{arg}^1_1 \ldots \underset{\substack{\uparrow\\ b}}{\mathsf{arg}^1_i}\}\{\mathsf{arg}^2_j \ldots \mathsf{arg}^2_k\} \ldots \{\mathsf{arg}^n_l \ldots \underset{\substack{\uparrow\uparrow\\ c\,d}}{\mathsf{arg}^n_m\}}$

The ability to extend parent categories in this way is supported in Grok. However, even with such capability, separate hierarchies are still required for each voice of a language such as Toba Batak since there is a structural difference between the categories. For example, the simple transitive category of the Active Voice form of the root *ida* 'see' is $(\mathsf{s}_i/_\diamond\mathsf{np}^{\mathbf{Act}})/_\diamond\mathsf{np}^{\mathbf{Pat}}$ while that of the Objective Voice is $(\mathsf{s}_i/_\diamond\mathsf{np}^{\mathbf{Pat}})/_\diamond\mathsf{np}^{\mathbf{Act}}$, so there is no simple way of underspecifying a generic transitive category which is constrained by the type declarations for the different voices. We might instead assume an approach in which an abstract hierarchy of Toba Batak verbs is built such that each category subtype simply adds its new argument to one multiset, and then lexical rules generate the categories to be used for processing, potentially restructuring the categories for each voice that a verb may take.

8.2.7 Other Aspects of Grok

There are many aspects of Grok that support categorial grammar parsing in general which have not been discussed here, such as exactly how the lexicon is defined and how morphological variants are keyed into categories in the lexicon. The strategies followed in Grok for grammar organization are generally modeled after those used in the XTAG system (Doran *et al.*, 2000). A major limitation is that Grok's overall set-up does not support inheritance and typing in general — instead it provides rigid ways of linking groups of categories with lexical stems and then tying those stems to their morphological variants. While this simple organization can be used to create lexicons of considerable complexity, it is not as principled, flexible, extendable, and robust infrastructure as that provided by systems such as the LKB (Copestake, 2002), which use the Type Definition Language (Krieger and Schäfer, 1994) to construct hierarchies of typed default feature structures. Some of the advantages of such an infrastructure were pointed out in §3.2 (page 66). Without going into specifics, it suffices to say that having this kind of support in Grok would have led to much more compact lexicons and stronger consistency and error checking.

An efficiency limitation of Grok is that its unification architecture does not perform any structure sharing and requires the creation of many copies of data structures during unification. It does have the advantage of supporting very local declarations of unification behavior and this makes it possible to order the unifications in more optimal manners than with a more general unification scheme.

Despite these limitations, it is still very useful to have a tool like Grok which specializes in categorial grammar rather than being a general system for working with feature structures. Thus, an ideal situation from the categorial perspective would be to have a Java system for performing unification and working with and constructing inheritance hierarchies that Grok can interface with. This way, all the advantages of using typed default feature structures defined over multiple inheritance hierarchies could be married with Grok's specialization in categorial grammar and in particular its ability to use efficient implementations of the combinatory rules.

Other than using subsumption checks and exploiting the universal rule set as described in this chapter, I have used no other means to improve the efficiency of the basic CKY chart parser used in Grok. Grok has proven effective for the purposes of this dissertation, but the lexicons defined herein are small and do not suffer from much categorial ambiguity. To scale the system up for wide-coverage parsing, it should be possible to use Grok in conjuncion with corpus-induced lexicons as was previously done by Hockenmaier *et al.* (2001), and then to extend Grok to take advantage of methods for the statistical parsing of CCG such as those given in Hockenmaier and Steedman (2002b) and Clark *et al.* (2002). These techniques will help in improving both the selection of categories to use in parsing a given sentence and the selection of the correct reading out of the potentially many parses a given sentence might recieve.

8.3 Implemented Grammars

I have written grammars for the English, Dutch, Turkish, and Tagalog analyses presented in the preceding chapters. The focus for these grammars has been on replicating the core aspects put forth in the linguistic argumentation and verifying that the various aspects of MULTI-MODAL CCG can be implemented effectively. The grammars use only the features which were discussed in the dissertation and which were crucial for capturing the phenomena discussed for each of the languages; thus, they do not enforce aspects such as number and person agreement.

Though I have utilized an inheritance hierarchy to define atomic categories (see Figure 3.1), Grok does not actually support inheritance in general. The categories in the lexicons are thus broadly delineated into sentential and nominal categories, and the features are set explicitly for each category rather than being able to reference particular types with the appropriate feature settings already instantiated.

With MULTI-MODAL CCG's invariant rule component, the variation between the lexicons can occur entirely in the lexicon. In the implemented grammars, this is the case for all but the type-raising rules. One reason for this discrepancy is that the analysis of Dutch requires that all noun phrases are type-raised in the lexicon, so these rules are omitted from the Dutch grammar. Also, I have allowed type-raising to be parameterized as to whether it incorporates a \$ into the result of the rule, as in the difference between $\mathsf{s}/(\mathsf{s}\backslash\mathsf{np})$ and $\mathsf{s}\$_1/(\mathsf{s}\$_1\backslash\mathsf{np})$. In many cases, such as forward type-raising in English, the \$ schematization is completely unnecessary and would simply slow down parsing since it extends the number of wayso in which argument clusters can be formed. Apart from this, the set of rules declared for each grammar is the same and includes all of the rules — though only the English grammar actually makes use of the substitution rules to handle the example sentences.

Each grammar comes with a testbed of strings which it either parses or fails to parse, as appropriate. These testbeds were crucial for ensuring that the grammars continued to function as desired as new constructions were added, and to make sure that changes made to Grok itself worked appropriately. The strings in the testbeds correspond to those given in the text of the dissertation, though some additional strings which do not appear in the dissertation are also included. Also, a few examples from the dissertation are omitted due to the fact that they would be redundant.

The grammars all produce simple logical forms as terms of HLDS. Little attention has been paid to scope-bearing elements in these representations as the main focus for present purposes was on demonstrating that the correct dependencies are determined during parsing. The non-English grammars use English predicates to improve the understandability of the parse output.

In the sections below, I discuss details of the four grammars which were implemented. Clearly, the focus for all of these grammars is on providing syntactic coverage of particular constructions as discussed in the dissertation rather than wide-coverage for any of the languages; they are therefore all quite small and have little or no ambiguity for the categories of verbs.

8.3.1 The English Grammar

The English lexicon is the most extensive of the four and is the only one that makes use of all of the rule types. There are 61 strings in the testbed, ranging from 3 to 15 words in size and averaging 7.42 words per string.

There is one ungrammatical string in the testbed which is accepted by the grammar — that of the example of extraction out of the subject **player that a book about astonished me* (247). The implementation of the English lexicon does not incorporate the suggestion by Steedman (2000b) that all noun phrases are already type-raised in the lexicon, and therefore does not block composition by verbs into categories such as the $\mathsf{np}/_{\triangleleft}\mathsf{np}$ of *a book about.*

8.3.2 The Dutch Grammar

There are 31 strings in the testbed for Dutch, ranging from 4 to 10 words in size and averaging 7.13 words per string.

In addition to the standard features used throughout the dissertation, the Dutch lexicon uses the ±SUB feature discussed in §5.3.4 (page 128). Its use is restricted solely to distinguishing subordinate and main clause verbs.

A notable feature of the Dutch grammar is that all noun phrases are type-raised in the lexicon and there are also categories for topicalized arguments, all of which use the $ schematization. For example, the categories for proper nouns like *Cecilia* are $\mathsf{s}\$/_i(\mathsf{s}\$\backslash_i\mathsf{np})$, $\mathsf{s}\$\backslash_i(\mathsf{s}\$/_i\mathsf{np})$, and $\ddot{\mathsf{s}}_c\$/_\diamond(\mathsf{s}_i\$|_.\mathsf{np})$. When the parser encounters a large cluster of noun phrases (as, for example, in (254) on page 116), it thus has many ways of using the combinatory rules to combine them, and this significantly slows down parsing.

Another interesting aspect of implementing the Dutch lexicon relates to the category for *denk* 'think', which is necessary for capturing the extraction of embedded subjects and objects, such as (267), repeated here as (493).

(493) a.

de	arts	die	ik	denk	dat	het werk	heeft	gedaan
the	doctor	who	I	think	that	the work	has	done

**the doctor who I think that did the work*

b.

het werk	dat	ik	denk	dat	ze	heeft	gedaan
the work	that	I	think	that	she	has	done

the work that I think that she did

The initial hypothesis was to assign *denk* the category $(\mathsf{s}_i\backslash\mathsf{np})/\ddot{\mathsf{s}}_c$, but parsing (493a) showed that this category allowed an overgeneration in which *denk* forward cross composed with *het werk heeft gedaan* to create the category $(\mathsf{s}_i\backslash\mathsf{np})\backslash\mathsf{np}$ and then consume *ik* as the subject of *heeft*. This problem was resolved by using the slash $/_{\triangleleft}$ for the sentential argument of *denk*, which thus had the category $(\mathsf{s}_i\backslash\mathsf{np})/_{\triangleleft}\ddot{\mathsf{s}}_c$. Then, the composition with *het werk heeft gedaan* forced the passed-up argument to be marked for antecedent government: $(\mathsf{s}_i\backslash\mathsf{np})\backslash\mathsf{np}_{+ANT}$. What is even more interesting about the category $(\mathsf{s}_i\backslash\mathsf{np})/_{\triangleleft}\ddot{\mathsf{s}}_c$ is that it is virtually the same as that of the category $(\mathsf{s}_i\backslash\mathsf{np})/_{\triangleleft}\mathsf{s}_{fin}$ of words like *think* and *knew* in English (see (238), page 112). As shown in §5.2.7, the same slash appeared in English to allow embedded subjects to be extracted while ensuring that they were marked for antecedent government. This discovery of the commonality in the categories for the two languages might have been overlooked if the grammars had not been implemented.

8.3.3 The Tagalog Grammar

There are 77 strings in the testbed for Tagalog, ranging from 3 to 11 words in size and averaging 6.97 words per string.

Perhaps the most significant aspect of the Tagalog grammar is that apart from a simple error in the value of the MARKING feature originally given for the category of *ay*, it transfered from the paper to the implementation straightforwardly and immediately captured all the data it was designed to account for.

8.3.4 The Turkish Grammar

There are 22 strings in the testbed for Turkish, ranging from two to six words in size and averaging 4.14 words per string.

Apart from validating the coverage of the Turkish analysis, an interesting aspect of the Turkish grammar is that it takes very little time to parse the doubly scrambled sentence given in (376). In addition to the fact that multisets do not grow as they do in MULTISET-CCG this is certainly due in part to the unique case-marking on all the noun phrases, so it would be interesting to extend the grammar to deal with scrambled sentences with more ambiguity.

The Turkish grammar implementation also caught an overgeneration due to the interaction of contraposed categories with matrix verbs. The contraposed category $\ddot{\mathsf{s}}_c\backslash_{\diamond}(\mathsf{s}_i\backslash\mathsf{np}_{+ANT})$ given on 147 was initially assumed to bear the more powerful modality $\cdot$ as in $\ddot{\mathsf{s}}_c\backslash_{\cdot}(\mathsf{s}_i\backslash_{\cdot}\mathsf{np}_{+ANT})$; however, this led to a derivation for the ungrammatical sentence (379), repeated below.

(494) *Ahmet [benim t_i okudugumu] Fatma'ya *kitabı*$_i$ soyledi.
*Ahmet [I-GEN t_i read-GER-ACC] Fatma-DAT *book*-ACC say
(for: Ahmet told Fatma that I read the BOOK.)

Using the more restricted modality $\diamond$ excluded such derivations and actually accords better with the intended use of the category. Note that the point made in the text below (379) (page 148) regarding the fact that MULTI-MODAL CCG blocks the overgeneration because of its restricted power is still valid — there are no operations in MULTI-MODAL CCG that can successfully provide a derivation given the categories assumed in derivation (380). Given these same categories, MULTI-MODAL CCG will combine them successfully unless significant restrictions are placed on the type-raising and composition rules, as shown by Hoffman (1995). Thus, even though assuming a category like $\ddot{\mathsf{s}}_c\backslash_{\cdot}(\mathsf{s}_i\backslash_{\cdot}\mathsf{np}_{+ANT})$ can provide a derivation, this is a matter of lexical assignment and the grammar simply cannot overgenerate from an interaction of the more basic categories with the combinatory rules.

8.4 Summary

This chapter highlighted a number of aspects of the support for MULTI-MODAL CCG provided by Grok, focusing on the computational benefits that the properties of MULTI-MODAL CCG bring. Ways of packing the functionality of several rules into a single procedure were shown to be possible because of the modal sensitivity and invariancy of the MULTI-MODAL CCG rules. Grok itself demonstrates the computational feasibility of MULTI-MODAL CCG and the grammars provided for English, Dutch, Turkish, and Tagalog verify the descriptive correctness of the linguistic analyses given in this dissertation. The process of implementing the grammars also demonstrated the *linguistic* benefit of testing the analyses in this way.

Grok is a practical framework for hand-built applications of CCG grammars which offers simple interfaces for interacting with and testing grammars and can furthermore serve as a Java API for other applications which use the output of the parser. Though it has presently been disconnected from the corpus-based lexicon (previously done in Hockenmaier *et al.* (2001)), it has the potential to be linked to a newer, cleaner version of that lexicon (Hockenmaier and Steedman, 2002a) and be extended for statistical parsing methods. Of particular interest for the present enterprise will of course be the interaction of modalities with induced lexicons and their potential use as features in statistical models for parsing.

Appendix: Downloading and Running Grok

I have made a special MULTI-MODAL CCG release of Grok which provides the functionality and the grammars discussed in this chapter. It can be downloaded from either the Grok webpage or my personal webpage:

- `http://grok.sourceforge.net`
- `http://www.iccs.informatics.ed.ac.uk/~jmb/dissertation/`

Instructions for setting up Grok are given in the file `GETTING_STARTED` provided with the download .

There are two command line scripts for using Grok: `tgrok` and `groktest`. The script `tgrok` invokes a simple text interface for parsing a given grammar and is called by typing `tgrok <grammar_file>` at the command line. For example, to run the English grammar, change to the directory that contains it and type the following at the command line:

```
> tgrok english.gram
```

After which you will be presented with the following:

```
Enter strings to parse.
Type ':h' for help on display options and ':q' to quit.

grok>
```

At the `grok>` prompt, you can enter sentences to parse, as in the following example of parsing the sentence *the referee gave Ronaldinho a red card*:

```
grok> the referee gave Ronaldinho a red card
1 parse found.

Parse result:
s
```

To see the derivation, type `:derivs` at the prompt, hit enter, and then hit enter again without typing anything, which sends the previous string to the parser again.

```
grok> :derivs
grok>
Parsing: the referee gave Ronaldinho a red card
1 parse found.

Parse result:
s
------------------------------
(lex) the := n/^n
(lex) referee := n
(>) the referee := n
(>T) the referee := s/i(s\in)
(lex) gave := s\n/n/<n
(>B) the referee gave := s/n/<n
(lex) Ronaldinho := n
(<T) Ronaldinho := s$1\i(s$1/in)
(<) the referee gave Ronaldinho := s/n
(lex) a := n/^n
(>B) the referee gave Ronaldinho a := s/^n
(lex) red := n/^n
(>B) the referee gave Ronaldinho a red := s/^n
(lex) card := n
(>) the referee gave Ronaldinho a red card := s
```

Note that Grok uses ASCII symbols to representing the modalities as follows:

$$\texttt{*}\rightarrow\star \quad \texttt{\^{}}\rightarrow\diamond \quad \texttt{<x}\rightarrow\lessdot\!\times \quad \texttt{x>}\rightarrow\times\!\gtrdot \quad \texttt{<}\rightarrow\triangleleft \quad \texttt{>}\rightarrow\triangleright \quad \texttt{.}\rightarrow\cdot$$

The directionality portions of the modalities are suppressed when they match that of their slash (see Convention 9 on page 101). Also note that complex categories are displayed without parentheses whenever possible, using the convention that slashes associate to the left. Thus, a category such as $((\mathsf{s}\backslash\mathsf{n})\backslash(\mathsf{s}\backslash\mathsf{n}))/\mathsf{n}$ will appear in Grok as $\mathsf{s}\backslash\mathsf{n}\backslash(\mathsf{s}\backslash\mathsf{n})/\mathsf{n}$.

To inspect the features of the result category and the logical form produced by the parse, type `:reset` to hide the derivations, then `:show semantics`, and finally hit enter to parse the sentence again:

```
grok> :reset
grok> :show semantics
grok> :show feats
```

```
grok>
Parsing: the referee gave Ronaldinho a red card
1 parse found.

Parse result:
s{marking=unmarked, vform=fin}
 :@h3((give ^ <Actor>(e2 ^ referee)
       ^ <Patient>(e9 ^ (card ^ <GenRel>red))
       ^ <Addressee>(e4 ^ Ronaldinho)))
```

The lexicons are typically stored in the file `lexicon.xml`, and the entries are declared as XML elements. These files are quite difficult to read, so a better way of inspecting the categories in the lexicon is to enter the words that are associated with them at the grok command line. To see all of the entries, type `:show all` and type the word whose categories you wish to see, as shown in the following interaction to get the categories for *that*:

```
grok> :show all
grok> that
3 parses found.

Parse 1: n{\*n/*(s/.n)}
Parse 2: s/^s
Parse 3: n{\*n/*(s\.n)}
```

The other script, `groktest` is used for batch processing a set of test strings, usually stored in a file called `testbed.xml`. Each string is accompanied by an attribute that declares how many parses the grammar should produce for it. During a test, this is compared to the actual number of parses produced and the result is reported in the output of `groktest`. While this does not provide a rigid check that the output categories are the desired ones, it usually suffices to highlight when the grammar has been negatively disturbed by deletions or additions.

To use this functionality with the English grammar, type `groktest english.gram` at the command line, which produces the following kind of output:

```
> groktest english.gram
Result  #Parses Sentence
------------------------------------------------
good    1       Brazil defeated Germany
good    0       *defeated Brazil Germany
good    2       player that the referee gave a yellow card
good    1       the referee gave a yellow card the aggressive player
```

```
good    1       team that John knew that Brazil would defeat
good    0       *team that John knew that would defeat China
good    0       *John Brazil knew that would defeat China
good    1       team that John knew would defeat China
good    1       the referee gave the player today a well-deserved red card
good    1       player that I read a book about yesterday
bad     1       *player that a book about astonished me
```

Sentences which should recieve no parses are marked with a star, and the result is reported as good or bad. Note that the last sentence is a documented overgeneration of the grammar and that a relativized clause such as *player that the referee gave a yellow card* receives two parses, one corresponding to the extraction of the **Patient** and the other to that of the **Addressee**.

Chapter 9

Conclusion

In this dissertation, I have focused on a number of aspects of the CCG framework that until now have not been addressed or have proven problematic. Though I have dubbed the framework put forth in this dissertation Multi-Modal Combinatory Categorial Grammar, it should be noted that this is not intended to put it in opposition to CCG as it has been defined and utilized in previous work. Rather, this should be viewed as an evolution of the overall CCG approach, based on formal devices made available in CTL and MULTISET-CCG.

Figure 9.1 provides a summary comparing how CCG, CTL, and MULTI-MODAL CCG differ in (a) the theoretical underpinning of the formalisms, (b) the universal formal bases they provide for creating grammars, and (c) the components which are involved in creating individual grammars. As can be seen from this overview, MULTI-MODAL CCG merges the combinatory basis of CCG with the principled resource-management regime of CTL. The restricted generative capacity of the combinatory basis leads to cross-linguistic predictions about the space of natural language grammars, and resource-management through modalities allows tighter control within that restricted basis (**Thesis 1**). Whereas CTL provides a very general formal theory for the modal approach to resource-management, MULTI-MODAL CCG makes commitments about the actual modes and rules which are utilized in all grammars. As this dissertation has shown, these commitments lead to cross-linguistic predictions and correlations that are stronger than those available under CCG. The key aspect of MULTI-MODAL CCG that makes this result possible is its universal rule component, which leaves all variation to the lexicon (**Thesis 4**).

I have proposed a small set of modalities and defined how they are used in conjunction with the CCG rules, showing how these *particular* modalities and rule for-

	Underlying Theory	Universal Formal Base	Components for Each Grammar
CCG	a. Combinators b. Principles	a. Feature-Based Categories b. Combinatory Rule Schemata	a. Lexicon b. Language Specific Combinatory Rules
CTL	a. Resource-Sensitive Logic b. Multiple Modes of Composition	a. Categories as Formulas b. Base Logic c. Structural Reasoning	a. Lexicon b. Language Specific Structural Rules c. Language Specific Modes
MM-CCG	a. Combinators b. Principles c. Multiple Modes of Composition	a. Modalized, Feature-Based Categories b. Modally Sensitive Combinatory Rules c. Universal Set of Modes	a. Lexicon

Figure 9.1: Overview of the architectures of CCG, CTL, and MULTI-MODAL CCG.

mulations positively impact the analysis of linguistic phenomena in English, Dutch, Turkish, Tagalog, and Toba Batak. The modalities I have proposed might be subject to revision in the future, but the linguistic analyses provided in this dissertation have demonstrated that we must use *some* set of modalities to control the applicability of combinatory rules. One possible expansion of the use of modalities is to associate arrays of modalities with every slash as a way of cleanly separating modes controlling associativity and permutativity from those of other dimensions, such as headedness, the active versus inert distinction, or the lexically connected versus derivationally connected distinction discussed in §8.2.2. If we wish to encode all of these distinctions but are forced to do so with a single modality on each slash, the relationships between the various modalities will become considerably more complex than an approach which factors their behaviors into separate dimensions.

By implementing a multi-modal perspective in CCG, I hope to bring the logical and rule-based categorial traditions closer together. Indeed, the analyses provided in this dissertation should transfer fairly straightforwardly to the CTL setting, using the CTL rules that I have defined. This may require either using basic feature structures in CTL or modifying the categories so that features are declared as unary modalities. Also, categories defined with multisets would need to be expanded into all of their rigid instantiations. Nonetheless, the core properties of the analyses arise from the categories and the modes of grammatical composition they license, and therefore the two systems will demonstrate the same behaviors for the same reasons on this more fundamental level. Based on the relationship between MULTI-MODAL CCG and the CTL rules I have defined to ground those rules, one can furthermore envision a *general* technique for translating CTL grammars into rule-based instantiations, essentially by running the proof system to create a set of theorems which define the rule base. The proof for the division rule given in (353) on page 137 using the CTL basis for MULTI-MODAL CCG is an example of how rules can be derived from the proof theory. Such translations would lack the full logical power and generality of CTL, but it could prove a useful avenue for more efficient parsing with CTL grammars.

Another major formal contribution of this work has been to improve CCG's ability to deal with flexible word orders by incorporating the multiset definition of categories into CCG. I have done so in a way that does not increase the generative power of the system, thus providing a kind of de-extension of Hoffman's MULTISET-CCG. The ability of MULTI-MODAL CCG to handle limited, but not unbounded, levels of long distance scrambling provides further support to the arguments of Joshi *et al.*

(2000) and Kulick (2000) that limits on scrambling can and should be considered as a property of linguistic competence, in contrast with the viewpoint put forth by Rambow (1994) and Hoffman (1995), who claim that such limitations lie in the domain of performance (**Thesis 3**).

It was also shown that the flexibility provided by categories with multisets can be used in languages like English and Toba Batak without engendering too much freedom in word order. Using very restricted modalities on the arguments in the multisets greatly limits their long distance combinatory potential while providing the necessary local permutativity. MULTI-MODAL CCG thus provides a single system that can be utilized for configurational languages such as English and flexible ones like Turkish, catering to the individual needs of each with a small set of universal modalities.

It remains to be worked out how the MULTI-MODAL CCG account of word order can be integrated with an account of information structure. It should be possible to use a parallel information structure derivation such as that proposed by Hoffman (1995) without much modification, but an account which provides the information structure as part of the MULTI-MODAL CCG system proper would be preferable. This would be in line with the evolution of Steedman's account of intonation that has progressed from dual derivations in Steedman (1991) to the use of intonational features as in Steedman (2000a). Ultimately, we would like an account of information structure which explains the actual use of different word orders, such as that given by Kruijff (2001).

The resource-sensitivity of MULTI-MODAL CCG makes it possible to make the rules inviolable and leave all variation to the lexicon (**Thesis 4**). This reduction in the parametricity of the rule-base is a principled one. Under the restrictional setting previously assumed in CCG, it was possible to restrict parts of the rules to very specific categories and/or block certain categories from serving as input to the rule. This led to a situation in which the same restriction often needed to be stated for more than one rule — with no explanation as to why this should be the case. By pushing the control into the lexicon through the use of categories that can combine only through particular modes of grammatical composition, we see that the categories can be appropriately blocked from all the relevant combinatory rules. And not only have the linguistic predictions become stronger — it is also considerably easier for the working categorial grammarian to ensure that a grammar is suitably controlled by using resource-sensitive categories than it is by using globally effective rule restrictions.

The flip-side of the invariant rule component is that it becomes possible to exploit rules which would otherwise need to be absolutely banned from some grammars, such as forward crossed composition in the grammar of English. Despite the universal availability of the rules, the system's resource-sensitivity allows any given grammar to select only a subset of the rules through the use of modes that are restricted to certain rule groups. For example, a MULTI-MODAL CCG grammar that uses only the more restrictive modality $\star$ will exhibit the same behavior as the AB calculus. Even though it seems unlikely that any natural language grammar would be quite so strict, it does allow different phenomena to be handled with different levels of formal power in a principled manner. Furthermore, if it ever is determined necessary to augment the rule-base with more powerful rules, we can do so in controlled manner that will preserve previous analyses. For example, we *can* use more powerful rules such as those proposed by Hoffman in a controlled manner by defining a super-permuting modality that licenses these rules, as discussed in §5.5 and §6.5. In this manner, MULTI-MODAL CCG can grow in power without losing the discrimination provided by the present system.

The MULTI-MODAL CCG formulation itself is completely compatible with a parametric view of the rule base, therefore leaving ample theoretical room to switch to a less extreme lexicalist position than that stated in **Thesis 4**. However, while I recognize that it may be necessary to allow languages to differ with respect to *which* rules they utilize, the rules themselves must remain unchanged from their given form and cannot be restricted or tailored for a given language in any way. This accords with the approach standardly assumed in CTL, in which a particular grammar can utilize different structural rules but cannot place arbitrary restrictions on them.

Because a significant amount of variation has been pushed from the rule component into the lexicon, the need for a theory of the CCG lexicon becomes more pressing than ever. The inheritance-based approach of Villavicencio (2002) appears to be an excellent starting point for encoding lexical redundancy in a systematic and well-motivated fashion. I have shown in §8.2.6 that the structure of MULTI-MODAL CCG categories can be utilized to extend this approach to parametrically diverse languages, though some challenges remain for Toba Batak. I have also defined a preliminary hierarchy for atomic categories that contains some of the major distinctions which are needed in the linguistic analyses offered in this dissertation. Even though it is incomplete and merits a much more detailed specification such as that which is typically made available in HPSG, it provides a more principled approach to atomic categories than has thus far been advanced and the distinctions were crucial for the analysis of every languages investigated in this dissertation.

This dissertation has also covered extensive linguistic ground over phenomena occurring in English, Dutch, Turkish, Tagalog, and Toba Batak. The formal developments were shown to provide significant improvements for existing CCG analyses of English and Dutch. Whereas the rule sets previously declared for the two languages had significant differences in terms of language-specific restrictions, omission of some rules in one and not the other (e.g. forward crossed composition in English), and changes to the categories of the rules themselves (e.g. forward harmonic composition in Dutch), the MULTI-MODAL CCG analyses utilized precisely the same rules for both grammars.

MULTI-MODAL CCG was also shown to adequately handle local and long distance scrambling in Turkish without overgenerating in the way that the MULTISET-CCG analysis of Hoffman (1995) did. In particular, MULTI-MODAL CCG provided inherent limits due to its restricted generative capacity and limits enforced straightforwardly by slashes with restrictive modalities. Hoffman instead required complex rule restrictions to avoid overgeneration.

Due to the formal devices made available by MULTI-MODAL CCG, the categorial analysis of syntactic extraction asymmetries in Tagalog given in Chapter 7 greatly enhances the coverage and simplicity of the one provided in Baldridge (1998). Where restrictions and *ad hoc* mechanisms were previously needed to control Tagalog's long distance asymmetries, the MULTI-MODAL CCG analysis relies on cross-linguistically motivated ways of limiting associativity. The result is the most extensive coverage of Tagalog asymmetries to date, while using far less generative power to achieve that task than other accounts, especially that of Nakamura (1998).

Previous linguistic analyses of Tagalog have struggled to reconcile the local scrambling behavior of Tagalog with its extraction asymmetries because they have relied on specific phrase-structural positions to single out a particular argument for extraction. In the MULTI-MODAL CCG analysis provided in this dissertation, on the other hand, local permutativity and extractability are not intertwined — local permutativity is permitted because the arguments of Tagalog's verbal categories are contained in multisets, and the asymmetries arise because of differences in the modes of grammatical composition specified for those arguments (**Thesis 2**).

Another unique aspect of the Tagalog analysis is that it crucially depends on assuming that the syntactic types of word classes can differ across languages. For example, in English and many other languages nouns have the category n, while in Tagalog nouns have the category s_v/np. This assumption is supported by the distributional evidence of basic Tagalog sentences, and it forces new assumptions about the categories of other word classes, such as case markers, that make the correct predictions about a wide range of constructions. It would be of great interest to examine other languages which lack a copular verb to see if similar patterns emerge.

Finally, a basic categorial analysis was given for Toba Batak which demonstrates the interesting interaction of verbal voice, word order, adverb placement, and extraction asymmetries in the language. The analysis of Toba Batak rounds out the overall cross-linguistic characterization of syntactic extraction asymmetries provided in this dissertation. Extraction asymmetries arise essentially because the functor categories of the grammar of any given language combine with their arguments through particular modes of grammatical composition that mediate the possibilities for associative and permutative operations to be utilized in the grammar (**Thesis 5**). This affects what the possible consituents and word orders are, and this in turn makes different arguments accessible or inaccessible for extraction. In English and Toba Batak, it was shown that important categories, such as those for complementizers and verbs, license only associative operations and hence limit the possible constituents which can be created. In Tagalog we find the reverse — asymmetries arise because generally only one argument of a verb can license associative operations, whilst all of them allow permutative operations. I am unaware of any extraction asymmetries in Turkish, but it would hardly be surprising if there are none — given the language's propensity to allow not only local scrambling, but also long distance scrambling, the categories of the Turkish grammar must license both associative and permutative operations. Though Hoffman (1995) reports that there are some islands to extraction in Turkish, she attributes them to semantic incoherence rather than failure of the syntax.

The fact that languages like Tagalog and Toba Batak have been dealt with is particularly significant from the categorial perspective. Categorial grammars have suffered from the perception by some that, although they may provide interesting accounts of phenomena in European languages such as English and Dutch, they are not generally suitable for wider linguistic application to more parametrically-diverse languages in other language families. This dissertation thus adds to the growing body of work that demonstrates that such a perception is ill-founded.

Finally, the invariant nature of MULTI-MODAL CCG rules both simplifies the task of grammar development and makes it possible to write more efficient implementations of combinatory rules than is possible for a parametric, restrictable rule component. The category structures used by MULTI-MODAL CCG are also important for defining lexical inheritance hierarchies for diverse language types. The effective use of these properties has been demonstrated in Grok, a practical framework for developing and using MULTI-MODAL CCG grammars (**Thesis 6**). Grok was crucial for checking the validity of the analyses and the process of implementing the grammars even led to some interesting linguistic observations, as discussed in Chapter 8, §8.3.

A meta-goal of this dissertation has been to demonstrate that it is necessary to recognize the individual character and elegance of many different approaches to natural language grammar and learn from them. I hope to have shown the advantages of incorporating techniques, devices, distinctions, and perspectives from a variety of approaches, which are so often concerned with orthogonal issues. Their solutions can at times be utilized complementarily, and we should be on constant look-out for cross-fertilization of this nature.

The many specific formal, linguistic, and computational points made throughout this dissertation come together to provide strong justification for the central thesis set out in the introduction — that an explanatory theory of natural language grammar can be based on a categorial grammar formalism which allows cross-linguistic variation only in the lexicon and has computationally attractive properties. CCG's notion of universal grammar just got more universal, and we now await a fuller and more cross-linguistically articulated theory of the lexicon.

Bibliography

Abeillé, Anne and Rambow, Owen, 2000. "Tree Adjoining Grammar: an overview." In *Tree Adjoining Grammar: formalisms, linguistic analysis, and processing*, Stanford: CSLI Publications. 1–68.

Ades, Anthony and Steedman, Mark, 1982. "On the order of words." *Linguistics & Philosophy* 7:639–642.

Ajdukiewicz, Kazimierz, 1935. "Die syntaktische Konnexität." In Storrs McCall (ed.), *Polish Logic 1920-1939*, Oxford University Press, 207-231. translated from *Studia Philosophica*, 1, 1-27.

Baldridge, Jason, 1998. *Local Scrambling and Syntactic Asymmetries.* Master's thesis, University of Pennsylvania.

Baldridge, Jason and Kruijff, Geert-Jan, 2002. "Coupling CCG and Hybrid Logic Dependency Semantics." In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics.* Philadelphia, Pennsylvania, 319–326.

Bar-Hillel, Yehoshua, 1953. "A Quasi-arithmetical Notation for Syntactic Description." *Language* 29:47–58.

Bar-Hillel, Yehoshua, Gaifman, C., and Shamir, E., 1964. "On categorial and phrase structure grammars." In Yehoshua Bar-Hillel (ed.), *Language and information*, Reading MA: Addison-Wesley. 99–115. 1964.

Beavers, John, in progress. "Type-Driven Combinatory Categorial Grammar." unpublished ms., Stanford University.

Becker, Tilman, 2000. "Patterns in Metarules for TAG." In Anne Abeillé and Owen Rambow (eds.), *Tree Adjoining Grammars: Formalisms, Linguistic Analysis and Processing*, Stanford, CA: CSLI Publications. 331–342.

Bernardi, Raffaella, 2002. *Reasoning with Polarity in Categorial Type Logic.* Ph.D. thesis, University of Utrecht.

Bierner, Gann, 2001. *Alternative Phrases: Theoretical Analysis and Practical Applications.* Ph.D. thesis, Division of Informatics, University of Edinburgh.

Blackburn, Patrick, 2000. "Representation, Reasoning, and Relational Structures: a Hybrid Logic Manifesto." *Logic Journal of the IGPL* 8(3):339–365.

Bozsahin, Cem, 1998. "Deriving the Predicate-Argument Structure for a Free Word Order Language." In *Proceedings of COLING-ACL '98, Montreal.* Cambridge MA: MIT Press, 167–173.

Bozsahin, Cem, 2002. "Lexical Origins of Word Order and Word Order Flexibility." In preparation for a chapter in Theoretical Issues in Word Order, S. Ozsoy (ed.), available from http://www.ceng.metu.edu.tr/~bozsahin/lowowof12.ps.gz.

Bresnan, Joan, 1977. "Variables in Transformations." In Peter Culicover, Thomas Wasow, and Adrian Akmajian (eds.), *Formal Syntax*, New York: Academic Press. 157–196.

Briscoe, Edward, 1997. "Co-evolution of Language and the Language Acquisition Device." In *Proceedings of 35th Annual Meeting of the Association for Computational Linguistics, Madrid, July.* 418–427.

Bröker, Norbert, 1998. "Separating Surface Order and Syntactic Relations in a Dependecy Grammar." In *Proceedings of COLING-ACL '98, Montreal.* Cambridge MA: MIT Press, 174–180.

Carpenter, Bob, 1995. "The Turing-Completeness of Multimodal Categorial Grammar." ms, http://www.illc.uva.nl/j50/contribs/carpenter/index.html.

Carpenter, Bob, 1999. "German Word Order and "Linearization" in Type-Logical Grammar." Lucent Technologies.

Cena, Resty, 1979. "Tagalog Counterexamples to the Accessibility Hierarchy." *Studies in Philippine Linguistics* 31:119–24.

Chomsky, Noam, 1965. *Aspects of the Theory of Syntax.* Cambridge MA: MIT Press.

Chomsky, Noam, 1981. *Lectures on Government and Binding.* Dordrecht: Foris.

Chomsky, Noam, 1994. "Bare Phrase Structure." In *MIT Occasional Papers in Linguistics.* MIT Working Papers in Linguistics, Cambridge, MA.

Chomsky, Noam, 1995. *The Minimalist Program.* Cambridge, Mass.: MIT Press.

Chomsky, Noam and Lasnik, Howard, 1993. "The Theory of Principles and Parameters." In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, and Theo Vennemann (eds.), *Syntax: An International Handbook of Contemporary Research*, Berlin: Walter de Gruyter.

Clark, Robin, 1985. "The Syntactic Nature of Logical Form: Evidence from Toba Batak." *Linguistic Inquiry* 16:663–669.

Clark, Robin, 1991. "Towards a Modular Theory of Coreference." In James Huang and Robert May (eds.), *Logical Structure and Linguistic Structure: Cross-Linguistic Perspectives*, Dordrecht: Kluwer. 49–78.

Clark, Stephen, Hockenmaier, Julia, and Steedman, Mark, 2002. "Building Deep Dependency Structures using a Wide-Coverage CCG Parser." In *Proc. of the 40th Annual Meeting of the Association of Computational Linguistics*. Philadelphia, PA, 327–334.

Copestake, Ann, 2002. *Implementing Typed Feature Structure Grammars*. Stanford: CSLI Publications.

Copestake, Ann, Flickinger, Dan, Sag, Ivan, and Pollard, Carl, 1999. "Minimal Recursion Semantics: An Introduction." unpublished ms, available from http://www-csli.stanford.edu/~aac/newmrs.ps.

Copestake, Ann, Lascarides, Alex, and Flickinger, Dan, 2001. "An Algebra for Semantic Construction in Constraint-based Grammars." In *Proceedings of the 39th Annual Meeting of the Association of Computational Linguistics*. Toulouse, France, 132–139.

Cornell, Thomas L., 1997. "A Type-Logical Perspective on Minimalist Derivations." In Geert-Jan M. Kruijff, Glyn V. Morrill, and Richard T. Oehrle (eds.), *Formal Grammar 1997, Proceedings of the Conference held in Aix en Provence, France, August 9-10*.

Cunningham, Hamish, 2000. *Software Architecture for Language Engineering*. Ph.D. thesis, University of Sheffield.

Curry, Haskell B. and Feys, Robert, 1958. *Combinatory Logic: Vol I*. North Holland, Amsterdam.

Davis, Anthony R., 1996. *Linking and the Hierarchical Lexicon*. Ph.D. thesis, Department of Linguistics, Stanford University, Stanford CA.

Doran, Christy, Hockey, Beth Ann, Sarkar, Anoop, Srinivas, Bangalore, and Xia, Fei, 2000. "Evolution of the XTAG System." In Anne Abeillé and Owen Rambow (eds.), *Tree Adjoining Grammars: Formalisms, Linguistic Analysis and Processing*, Stanford, CA: CSLI Publications. 371–404.

Doran, Christy and Srinivas, Bangalore, 2000. "Developing a Wide-Coverage CCG System." In Anne Abeillé and Owen Rambow (eds.), *Tree Adjoining Grammars: Formalisms, Linguistic Analysis and Processing*, Stanford, CA: CSLI Publications. 405–426.

Dowty, David, 1988. "Type-raising, Functional Composition, and Non-constituent Coordination." In Richard T. Oehrle, Emmon Bach, and Deirdre Wheeler (eds.), *Categorial Grammars and Natural Language Structures*, Dordrecht: Reidel. 153–198. Proceedings of the Conference on Categorial Grammar,Tucson, AR, June 1985.

Evers, Arnold, 1975. *The Transformational Cycle in Dutch and German*. Ph.D. thesis, University of Utrecht. pub. Indiana Linguistics Club.

Foster, John, 1990. *A Theory of Word-Order in Categorial Grammar, with Special Reference to Spanish.* Ph.D. thesis, University of York.

Frank, Robert, 1992. *Syntactic Locality and Tree Adjoining Grammar: Grammatical, Acquisition, and Processing Perspectives.* Ph.D. thesis, University of Pennsylvania, Philadelphia PA.

Frank, Robert, 2002. *Phrase Structure Composition and Syntactic Dependencies.* Cambridge, MA: MIT Press.

Gazdar, Gerald, 1988. "Applicability of Indexed Grammars to Natural Languages." In Uwe Reyle and Christian Rohrer (eds.), *Natural Language Parsing and Linguistic Theories*, Dordrecht: Reidel. 69–94.

Gazdar, Gerald, Klein, Ewan, Pullum, Geoffrey K., and Sag, Ivan A., 1985. *Generalised Phrase Structure Grammar.* Oxford: Blackwell.

Geach, Peter, 1972. "A Program for Syntax." In Donald Davidson and Gilbert Harman (eds.), *Semantics of Natural Language*, Dordrecht: Reidel. 483–497.

Gil, David, 1993. "Tagalog Semantics." In *Proceedings of the 19th Berkeley Linguistics Society.* pp.390–403.

Ginzburg, Jonathan and Sag, Ivan, 2000. *Interrogative Investigations: The Form, Meaning, and Use of English Interrogatives.* CSLI Publications.

Girard, Jean-Yves, 1987. "Linear Logic." *Theoretical Computer Science* 50:1–102.

Guilfoyle, Eithne, Hung, Henrietta, and Travis, Lisa, 1992. "Spec of IP and Spec of VP: two subjects in Austronesian Languages." *Natural Language and Linguistic Theory* 10:375–414.

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

Hepple, Mark, 1994. "A General Framework for Hybrid Substructural Categorial Logics." Technical Report 94-14, IRCS, University of Pennsylvania, Philadelphia PA.

Hepple, Mark, 1997. "A Dependency-Based Approach to Bounded and Unbounded Movement." In *Proceedings of the Fifth Meeting on Mathematics of Language (MOL-5).* DFKI-D-97-02.

Hepple, Mark, 1999. "An Earley-style Predictive Chart Parsing Method for Lambek Grammars." In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*. Maryland, 465–475.

Hockenmaier, Julia, Bierner, Gann, and Baldridge, Jason, 2001. "Extending the Coverage of a CCG System." In *Journal of Language and Computation.*

Hockenmaier, Julia and Steedman, Mark, 2002a. "Acquiring Compact Lexicalized Grammars from a Cleaner Treebank." In *Proceedings of the Third International Conference on Language Resources and Evaluation*. Las Palmas, 1974–1981.

Hockenmaier, Julia and Steedman, Mark, 2002b. "Generative Models for Statistical Parsing with Combinatory Categorial Grammar." In *Proc. of the 40th Annual Meeting of the Association of Computational Linguistics*. Philadelphia, PA, 335–342.

Hoekstra, Eric, 1986. "On the Structural Subject Position in Tagalog." *Lingua* 70:41–55.

Hoffman, Beryl, 1995. *Computational Analysis of the Syntax and Interpretation of 'Free' Word-order in Turkish*. Ph.D. thesis, University of Pennsylvania. IRCS Report 95-17.

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*, Foris Dordrecht.

Iatridou, Sabine and Kroch, Anthony, 1992. "The Licensing of CP-recursion and its Relevance to the Germanic Verb-Second Phenomenon." In *Working Papers in Scandinavian Syntax*, volume 50. 1–24.

Joshi, Aravind, 1988. "Tree Adjoining Grammars." In David Dowty, Lauri Karttunen, and Arnold Zwicky (eds.), *Natural Language Parsing*, Cambridge: Cambridge University Press. 206–250.

Joshi, Aravind, Rambow, Owen, and Becker, Tilman, 2000. "Complexity of Scrambling: A New Twist to the Competence-Performance Distinction." In Anne Abeillé and Owen Rambow (eds.), *Tree Adjoining Grammars: Formalisms, Linguistic Analysis and Processing*, Stanford, CA: CSLI Publications. 167–182.

Kaplan, Ron and Zaenen, Annie, 1989. "Long-distance Dependencies, Constituent Structure, and Functional Uncertainty." In M.R. Baltin and A.S. Kroch (eds.), *Alternative Conceptions of Phrase Structure*, Chicago University Press, Chicago.

Kaplan, Ronald and Bresnan, Joan, 1982. "Lexical-Functional Grammar: A formal system for grammatical representation." In *The Mental Representation of Grammatical Relations*, Cambridge, MA: MIT Press. 173–281.

Karamanis, Nikiforos, 2001. "A Categorial Grammar for Greek." In *Proceedings of the 15th International Symposium on Theoretical and Applied Linguistics*. Aristotle University of Thessaloniki.

Karttunen, Lauri, 1989. "Radical Lexicalism." In M.R. Baltin and A.S. Kroch (eds.), *Alternative Conceptions of Phrase Structure*, Chicago University Press, Chicago.

Keenan, Ed and Comrie, Bernard, 1977. "Noun Phrase Accessibility and Universal Grammar." *Linguistic Inquiry* 8:63–99.

Keenan, Edward, 1978. "The Syntax of Subject-Final Languages." In Winfred Lehmann (ed.), *Syntactic Typology: Studies in the Phenomenology of Language*, The Harvester Press. 267–328.

Komagata, Nobo, 1999. *Information Structure in Texts:A Computational Analysis of Contextual Appropriateness in English and Japanese*. Ph.D. thesis, University of Pennsylvania.

Komagata, Nobo, 2002. "Coordination of Unlike (?) Categories: How *Not* to Distinguish Categories." http://www.cis.upenn.edu/˜komagata.

König, Esther, 1999. "LexGram - a practical categorial grammar formalism." *Journal of Language and Computation* 1(1):33–52.

Krieger, H. and Schäfer, U., 1994. "TDL: A Type Description Language for Constraint-Based Grammars." In *Proceedings of the 15th International Conference on Computational Linguistics*. 893–899.

Kroeger, Paul, 1993. *Phrase Structure and Grammatical Relations in Tagalog*. Stanford: CSLI Publications.

Kruijff, Geert-Jan M., 2001. *A Categorial Modal Architecture of Informativity: Dependency Grammar Logic & Information Structure*. Ph.D. thesis, Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic.

Kruijff, Geert-Jan M. and Baldridge, Jason M., 2000. "Relating Categorial Type Logics and CCG through Simulation." unpublished ms, available from http://www.iccs.inf.ed.ac.uk/˜jmb/simulation.ps.gz.

Kulick, Seth, 2000. *Constraining Non-local Dependencies in Tree Adjoining Grammar: Computational and Linguistic Perspectives*. Ph.D. thesis, University of Pennsylvania.

Kurtonina, Natasha and Moortgat, Michael, 1997. "Structural Control." In Patrick Blackburn and Maarten De Rijke (eds.), *Specifying Syntactic Structures*, Stanford CA: CSLI Publications & FoLLi. 75–113.

Laenzlinger, Christopher, 1998. *Adverbs, Pronouns, and Clause Structure in Romance and Germanic*. John Benjamins.

Lambek, Joachim, 1958. "The mathematics of sentence structure." *American Mathematical Monthly* 65:154–169.

Lascarides, Alex, Copestake, Ann, and Briscoe, Ted, 1996. "Order Independent Persistent Typed Default Unification." *Linguistics and Philosophy* 19(1):1–89.

Lecomte, Alain, 2001. "Categorial Minimalism." In *Logical Aspects of Computational Linguistics*. Springer, 143–158.

Maclachlan, Anna and Nakamura, Masanori, 1997. "Case-checking and specificity in Tagalog." *The Linguistic Review* 14:307–333.

Marcus, Mitchell P., Santorini, Beatrice, and Marcinkiewicz, Mary Ann, 1993. "Building a large annotated corpus of English: the Penn Treebank." *Computational linguistics* 19:313–330.

Milward, David, 1994. "Dynamic Dependency Grammar." *Linguistics & Philosophy* 17(6):561–605.

Miyagawa, Shigeru, 1997. "Against Optional Scrambling." *Linguistic Inquiry* 28:1–25.

Moens, Marc, Calder, Jo, Klein, Ewan, Reape, Michael, and Zeevat, Henk, 1989. "Expressing Generalizations in Unification-Based Grammar Formalisms." In *Proceedings of the 4th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*. Manchester, England, 66–71.

Montague, Richard, 1973. "The proper treatment of quantification in ordinary English." In Jaakko Hintikka, J. Moravcsik, and P. Suppes (eds.), *Approaches to natural language: proceedings of the 1970 Stanford workshop on grammar and semantics*, Dordrecht: Riedel. 221–242. Reprinted in Montague 1974, 247-279.

Moortgat, Michael, 1988. *Categorial Investigations: Logical and Linguistic Aspects of the Lambek Calculus*. Dordrecht, The Netherlands: Foris.

Moortgat, Michael, 1997. "Categorial Type Logics." In Johan van Benthem and Alice ter Meulen (eds.), *Handbook of Logic and Language*, Amsterdam New York etc.: Elsevier Science B.V.

Moortgat, Michael, 1999. "Constants of grammatical reasoning." In Gosse Bouma, Erhard W. Hinrichs, Geert-Jan M. Kruijff, and Richard T. Oehrle (eds.), *Constraints and Resources in Natural Language Syntax and Semantics*, Stanford CA: CSLI Publications.

Moortgat, Michael and Morrill, Glyn, 1991. "Heads and Phrases: Type Calculus for Dependency and Constituent Structure." Unpublished manuscript. Available from http://www-lsi.upc.es/~glyn/.

Moot, Richard, 2002. *Proof Nets for Linguistic Analysis*. Ph.D. thesis, University of Utrecht.

Morrill, Glyn V., 1994. *Type Logical Grammar: Categorial Logic of Signs*. Dordrecht, Boston, London: Kluwer Academic Publishers.

Nakamura, Masanori, 1994. "An Economy Account of Wh-extraction in Tagalog." In *Proceedings of WCCFL XII*. Stanford: CSLI Publications.

Nakamura, Masanori, 1998. "Reference Set, Minimal Link Condition, and Parameterization." In P. Barbosa, P.H.D. Fox, M. McGinnis, and D. Pesetsky (eds.), *Is the Best Good Enough?*, Cambridge, MA: MIT Press. 291–313.

Oehrle, Richard T., 1988. "Multi-Dimensional Compositional Functions as a Basis for Grammatical Analysis." In Emon Bach, Richard T. Oehrle, and Deirdre Wheeler (eds.), *Categorial Grammars and Natural Language Structures*, Dordrecht The Netherlands: D. Reidel. 349–389.

Oehrle, Richard T., 1994. "Term-Labelled Categorial Type Systems." *Linguistics & Philosophy* 17(6):633–678.

Oehrle, Richard T., 1995. "Some 3-Dimensional Systems of Labelled Deduction." *Bulletin of the IGPL* 3(2,3):429–448. Special Issue on Deduction and Language.

Oehrle, Richard T., to appear. "Multi-Modal Type-Logical Grammar." to appear in Borsley and Börjars *Non-transformational Syntax*, available at http://www.let.uu.nl/esslli/Courses/moortgat/moortgat-oehrle3.ps.

Oehrle, Richard T. and Zhang, Shi, 1989. "Lambek Calculus and Prepositioning of Embedded Subjects." In *Proceedings of the 25th meeting of the Chicago Linguistic Society*. Chicago.

Pentus, Mati, 1997. "Product-Free Lambek Calculus and Context-Free Grammars." *The Journal of Symbolic Logic* 62(2):648–660.

Pollard, Carl, 1984. *Generalized Phrase Structure Grammars, Head Grammars, and Natural Languages*. Ph.D. thesis, Stanford University, Stanford, CA.

Pollard, Carl and Sag, Ivan, 1987. *Information-based Syntax and Semantics*, volume 1. Chicago: CSLI/Chicago University Press.

Pollard, Carl and Sag, Ivan, 1994. *Head Driven Phrase Structure Grammar*. CSLI/Chicago University Press, Chicago.

Pollock, Jean-Yves, 1989. "Verb Movement, Universal Grammar, and the Structure of IP." *Linguistic Inquiry* 20:365–424.

Pullum, Geoffrey K. and Gazdar, Gerald, 1982. "Natural languages and context-free languages." *Linguistics and Philosophy* 4:471–504.

Rambow, Owen, 1994. *Formal and Computational Aspects of Natural Language Syntax*. Ph.D. thesis, University of Pennsylvania, Philadelphia PA.

Reape, Michael, 1994. "Domain Union and Word Order Variation in German." In Klaus Netter John Nerbonne and Carl Pollard (eds.), *German in Head-driven Phrase Structure Grammar*, CSLI/Cambridge University Press, Cambridge, 151-197.

Retoré, Christian and Stabler, Edward, to appear. "Resource Logics and Minimalist Grammars." special issue of the Journal of Language and Computation.

Richards, Norvin, 2000. "Another Look at Tagalog Subjects." In *Formal Issues in Austronesian Linguistics*, Dordrecht: Kluwer.

Rizzi, Luigi, 1982. *Issues in Italian Syntax.* Foris, Dordrecht.

Ross, John Robert, 1967. *Constraints on Variables in Syntax.* Ph.D. thesis, MIT. Published as "Infinite Syntax!", Ablex, Norton, NJ. 1986.

Sag, Ivan, Gazdar, Gerald, Wasow, Thomas, and Weisler, Steven, 1985. "Coordination and How to Distinguish Categories." *Natural Language and Linguistic Theory* 3:117–172.

Schachter, Paul, 1976. "The Subject in Philippine Languages: Topic, Actor-Topic, or None of the Above." In C. Li (ed.), *Subject and Topic*, New York: Academic Press. 491–518.

Schachter, Paul and Otanes, Fe, 1972. *Tagalog Reference Grammar.* Berkeley: University of California Press.

Sgall, Petr, Hajičová, Eva, and Panevová, Jarmila, 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects.* Dordrecht, Boston, London: D. Reidel Publishing Company.

Shieber, Stuart, 1985. "Evidence against the context-freeness of natural language." *Linguistics and Philosophy* 8:333–343.

Shieber, Stuart, 1986. *An Introduction to Unification-based Approaches to Grammar.* CSLI/Chicago University Press, Chicago.

Smullyan, Raymond, 1985. *To Mock a Mockingbird.* New York: Knopf.

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, 1988. "Combinators and Grammars." In Richard T. Oehrle, Emmon Bach, and Deirdre Wheeler (eds.), *Categorial Grammars and Natural Language Structures*, Dordrecht: Reidel. 417–442. Proceedings of the Conference on Categorial Grammar,Tucson, AR, June 1985.

Steedman, Mark, 1991. "Structure and Intonation." *Language* 67:262–296.

Steedman, Mark, 1996. *Surface Structure and Interpretation.* Cambridge Mass.: MIT Press. Linguistic Inquiry Monograph, 30.

Steedman, Mark, 2000a. "Information Structure and the Syntax-Phonology Interface." *Linguistic Inquiry* 34:649–689.

Steedman, Mark, 2000b. *The Syntactic Process.* Cambridge Mass.: The MIT Press.

Trechsel, Frank, 2000. "A CCG Account of Tzotzil Pied Piping." *Natural Language and Linguistic Theory* 18:611–663.

Uszkoreit, Hans, 1987. *Word Order and Constituent Structure in German.* CSLI.

Van Benthem, Johan F.A.K., 1988. "The semantics of variety in categorial grammar." In Wojciech Buszkowski, Johan F.A.K. Van Benthem, and W. Marciszewski (eds.), *Categorial Grammar*, Amsterdam: John Benjamins, volume 25 of *Linguistic and Literary Studies in Eastern Europe.* 37–55. Reprint of Report 83-29, Department of Mathematics, Simon Fraser University, Vancouver, 1983.

Vermaat, W., 1999. "Controlling Movement: Minimalism in a deductive perspective."

Vijay-Shanker, K. and Weir, David, 1990. "Polynomial Time Parsing of Combinatory Categorial Grammars." In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics, Pittsburgh.* 1–8.

Vijay-Shanker, K. and Weir, David, 1994. "The Equivalence of Four Extensions of Context-free Grammar." *Mathematical Systems Theory* 27:511–546.

Villavicencio, Aline, 1997. *Building a wide-coverage Combinatory Categorial Grammar.* Master's thesis, Cambridge.

Villavicencio, Aline, 2002. *The Acquisition of a Unification-Based Generalised Categorial Grammar.* Ph.D. thesis, University of Cambridge.

Weir, David, 1988. *Characterising Mildly Context-sensitive Grammar Formalisms.* Ph.D. thesis, University of Pennsylvania. Tech. Report CIS-88-74.

Wintner, Shuly and Sarkar, Anoop, 2002. "A Note on Typing Feature Structures." *Computational Linguistics* 28(3):389–397.

Wittenburg, Kent, 1986. *Natural Language Parsing with Combinatory Categorial Grammar in a Graph-Unification Based Formalism.* Ph.D. thesis, University of Texas at Austin.

Wittenburg, Kent, 1987. "Predictive Combinators: a Method for Efficient Processing of Combinatory Grammars." In *Proceedings of the 25th Annual Conference of the ACL, Stanford.*

Wood, Mary McGee, 1993. *Categorial Grammar.* Routledge.

XTAG-group, 1999. "A Lexicalized Tree Adjoining Grammar for English." Technical Report IRCS-98-18, University of Pennsylvania.

Zeevat, Henk, 1988. "Combining Categorial Grammar and Unification." In Uwe Reyle and Christian Rohrer (eds.), *Natural Language Parsing and Linguistic Theories*, Dordrecht: Reidel. 202–229.

Zeevat, Henk, Klein, Ewan, and Calder, Jo, 1987. "An Introduction to Unification Categorial Grammar." In N. Haddock et al. (ed.), *Edinburgh Working Papers in Cognitive Science, 1: Categorial Grammar, Unification Grammar, and Parsing*, University of Edinburgh. 195–222.

Index