

Title of Your Dissertation

A Suitable Subtitle
on Multiple Lines

Published by

LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6111
e-mail: lot@uu.nl
<http://www.lotschool.nl>

Cover illustration: The Tower of Babel, by Pieter Bruegel

ISBN: 000-11-22222-33-4
NUR: 000

© CC-BY-SA 3.0¹ by Konstantinos Kogkalidis

You are free to:

Share – copy and redistribute the material in any medium or format
Adapt – remix, transform, and build upon the material

under the following terms:

Attribution - You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests I endorse you or your use.
NonCommercial – You may not use the material for commercial purposes.
ShareAlike – If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

¹<https://creativecommons.org/licenses/by-nc-sa/3.0/legalcode>

Title of Your Dissertation

A Suitable Subtitle
on Multiple Lines

Nederlandstalige Titel

De Tweede Regel van Je Titel,
ook Vrij Lang

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. Henk Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

door

Konstantinos Kogkalidis

geboren 19 Juli 1991
te Thessaloniki, Greece

Promotores: Prof. Dr. M. J. Moortgat
Prof. Dr. R. Moot

The research reported here was supported by the Netherlands Organization for Scientific Research under the scope of the project “A composition calculus for vector-based semantic modelling with a localization for Dutch” (project number 360-89-070).

todo

Contents

Preface	ix
I Introduction	1
1 The Simple Theory of Types	3
1.1 Intuitionistic Logic	3
1.1.1 Proof Equivalences	5
1.2 The Curry-Howard Correspondence	7
1.2.1 Term Equivalences	10
1.2.2 In Place of a Summary	11
1.3 Intermezzo	11
2 Going Linear	14
2.1 Linear Types	15
2.1.1 Proof & Term Reductions	16
2.2 Proof Nets	17
3 Lambek Calculi	22
3.1 Dropping Commutativity	22
3.1.1 Proof & Term Reductions	24
3.2 Dropping Associativity	24
3.2.1 Proof & Term Reductions	26
3.3 The Full Landscape	26
4 Restoring Control	27
4.1 The Logic of Modalities	28
4.1.1 Proof & Term Reductions	30
4.1.2 A Digression on Modal Terms	31
4.1.3 Properties	32
4.2 Structural Reasoning	32
5 The Linguistic Perspective	35
5.1 Type-Logical Grammars	36
5.1.1 The Role of Modalities	40
5.1.2 Intricacies of the Lexicon	41

5.1.3 Subtleties of Proof Search	44
5.1.4 Syntax-Semantics Interface	45
5.2 Abstract Categorical Grammars	50
5.2.1 Basic Definitions	50
5.2.2 Artificial Languages	51
5.2.3 Human Languages	53
5.3 Other Formalisms	56
5.3.1 Combinatory Categorical Grammars	56
5.3.2 Hybrid Type-Logical Grammars	57
6 Key References & Further Reading	58
Chapter Bibliography	59
II Typing Dependency Structure	63
7 Phrase vs. Dependency Structure	65
7.1 Phrase Structure Grammars	65
7.2 Dependency Grammars	67
8 Modalities for Dependency Demarcation	69
8.1 Two Dimensional Predicates	69
8.2 Modal Dependents	70
8.2.1 Complements vs. Adjuncts	71
8.2.2 Grammatical Functions	72
8.3 Inference with Dependency-Enhanced Types	73
8.3.1 Initial Lexical Adjustments	73
8.3.2 Dependencies and Structural Reasoning	74
8.4 Interfaces	81
8.4.1 Dependency Trees	81
8.4.2 Semantics	83
9 Æthel	84
Chapter Bibliography	85

Preface

Greetings, reader. Out of coincidence, or some weird turn of events, I have written this dissertation to-be and you have stumbled upon it. Introductions would normally be in order, but since this communication channel is asynchronous and unidirectional I will be doing double duty for the both of us.

So, let's start with you. A few scenarios are plausible as to why you are browsing these pages. Most likely you are an acquaintance of mine, either social – in which case you are wondering what it is I spent 5 years in Utrecht for, or academic – which means you are probably trying to figure out whether I am worthy of the title of doctor. In the latter case, I hope you won't be disappointed (especially so if you happen to be a member of the examination committee, for both our sakes). Or, perhaps, you are lazily scrolling through the opening pages to evaluate my fitness for a potential job in an organization you are representing? If so, you should definitely go for me – unless this happens to be any of the big five¹, in which case: shoo, and shame on you, future me. Otherwise, could it even be that you are actually interested in the subject matter of this thesis? That would be a shock; pleasant mostly, but also slightly alarming: I feel a bit conscious knowing that you might be putting my words under a critical lens – I'll do my best not to fail your expectations. In the unlikely event that you do not fall in any of the above categories, excuse my lack of foresight and know that you are still very welcome, and I am happy to have you around. In the more likely event that nobody ever reads this (far), let this transmission be forever lost to the void.

But enough with you, what about me? At the time of writing, I am in my early thirties and I call myself Kokos. I had the enormous luck of crossing paths with my supervisor, Michael, five years ago, during the first weeks of my graduate studies in Utrecht. The repercussions of this encounter were (and still are) unforeseeable. Coming from an engineering background that basked in practicality, with an obsessive repulsion to anything formal, his course offered me a glimpse of a whole new world. I got to see that proofs are not irrel-

¹Preemptive apologies for not making sense to readers after the imminent tech collapse.

evant bureaucracies to avoid, but objects of interest in themselves, hidden in plain sight from the working hacker under common programming patterns. If this naive revelation came as shock, you can imagine my almost mystical awe when I was shown how proof & type theories also offer suitable tools and vocabulary for the analysis of human languages. Despite my prior ignorance, the “holy trinity” between constructive logics, programming languages and natural languages has been (with its ups and downs) at the forefronts of theoretical research for well over a century. This dissertation aims to be my tiny contribution to this line of work, conducted from the angle of a late convert, a theory-conscious hacker.

If all this sounds enticing and you plan on sticking around, at least for a bit longer, I think it would be beneficial if we set down the terms and conditions of what is to follow. It is no secret that dissertations are often boring to read, and it can be easy to lose track of context in seemingly unending walls of text. Striking a balance between being pedantic and making too many assumptions on background knowledge is no easy task: the only way to spare you unnecessary headaches requires a mutual contract. On my part, I will try to clearly communicate my intentions, both about the thesis in full, and its parts in isolation: the idea is to make this manuscript as self-contained as possible, but without nitpicking on details or taking detours unnecessary for the presentation of the few novelties I have to contribute. Of you, I ask to remain conscious of what you are reading and aware of my own biases and limitations. The absence of feedback means that I can only model you in my imagination; I will inadvertently skip things that to me seem self-evident, and rant at length about others that you take for granted. So, feel free to skip ahead when something reads trivial, and do not judge too harshly when you encounter an explanation you find insufficient.

What this thesis is about

todo

contributions

sectioning

CHAPTER I

Introduction

“In the beginning there was the Word, and the Word had a type.”

Our story begins with the (over-ambitious, in hindsight) musings of one of the world’s most well-renowned mathematicians, David Hilbert. Unhappy with the numerous paradoxes and inconsistencies of mathematics at the end of the 19th century, Hilbert would postulate the existence and advocate the formulation of a finite set of axiomatic rules, which, when put together, would give rise to the most well-behaved system known to [wo]mankind, capable of acting as a universal meta-theory for all mathematics, in the process absolving all mathematicians of their sins. The idea was of course appealing and gained traction, not the least due to Hilbert’s influence over the field (and his will to exercise it). As with all ideas that generate traction, however, it was not long before a cultural counter-movement would develop. Intuitionism, with Luitzen Egbertus Jan Brouwer as its forefather, would challenge Hilbert’s program by questioning the objective validity of (any) mathematical logic. What it would claim, instead, is that mathematics is but a subjective process of construction that abides by some rules of inference, which, internally consistent as they may be, hold no reflection of deeper truth or meaning. In practice, intuitionists would reject the law of the excluded middle (an essential tool for Hilbert’s school of formalists) and argue that for a proof to be considered valid, it has to provide concrete instructions for the construction of the object it claims to prove. The dispute went on for a couple of decades, its flame carried on by the respective students of the two rivals. Logic, intrigue, conflict, fame, no L^AT_EX errors... these truly were the years to be an active mathematician. Eventually, in a critical moment of clarity and inspiration, and tired by the

ongoing drama, Kurt Gödel, with his famous incompleteness theorem, would declare Hilbert's program unattainable, thus putting a violent end to the line of formalist heathens, and paving the way for the true revolution that was to come. This is in reference, of course, to the biggest discovery of the last century¹, made independently (using wildly different words every time) by various mathematicians and logicians spanning different timelines. Put plainly, what is now known as the Curry-Howard correspondence establishes a syntactic equivalence between deductive systems in intuitionistic brands of logic and corresponding computational systems, called λ -calculi. Put even more plainly, it suggests that valid proofs in such logics constitute in fact compilable code for functional programs, bridging in essence the seemingly disparate fields of mathematical logic and computer science. The repercussions of this discovery were enormous, and are more tangible today than ever before; type systems comprised of higher-order λ -calculi and their logics provide the theoretical foundations for modern programming languages and proof assistants (this last fact is both important and interesting, but won't concern us much presently).

In a more niche (but equally beautiful) fragment of the academic world, and in parallel to the above developments, applied logicians and formally inclined linguists have been demonstrating a stunning perserverance in their self-imposed quest of modeling natural language syntax and semantics, making do only with the vocabulary provided by formal logics. This noble endeavour traces its origins back to Aristotle, but its modern incarnation is due to Jim Lambek, who was the first to point out that the grammaticality of a natural language utterance can be equated to provability in a certain logic (or type inhabitation, if one is to borrow the terminology of constructive type theories), if the grammar (a collection of empirical linguistic rules) were to be treated as a substructural logic (a collection of formal mathematical rules). Funnily enough, the kind of logics Lambek would employ for his purposes would be exactly those at the interesection of intuitionistic and linear logic, the latter only made formally explicit in a breakthrough paper by Jean-Yves Girard almost three decades later. By that time, Richard Montague had already come up with the fantastically novel idea of seeing no distinction between formal and natural languages, single-handedly birthing and popularizing the field of formal semantics (which would chiefly involve semantic computations using λ -calculus notation). With this, he fulfilled Gottlob Frege's long-prophesized principle of compositionality, which would once and for all put the Chomskian tradition to rest², ushering linguistics into a new era. With the benefit of posterity, it would be tempting for us to act smart and exclaim that Lambek and Montague's ideas were remarkably aligned. In reality, it took another couple of decades for someone to notice. The credit is due to Johan van Benthem, who basically pointed out that Lambek's calculi make for the perfect syntactic

¹In proof theory, at least.

²In some corners of the Earth, this part of the prophecy is yet to transpire.

machinery for Montague’s program, seeing as they admit the Curry-Howard correspondence, and are therefore able to drive semantic composition virtually for free (in fact one could go as far as to say that they are the only kind of machinery that can accomplish such a feat without being riddled with ad-hoc transformations). This revelation, combined with the contemporary bloom of substructural logics, was the spark that ignited a renewed interest in Lambek’s work. The culmination point for this interest was type-logical grammars (or categorial type logics): families of closely related type theories extending the original calculi of Lambek with unary operators lent from modal logic, intended to implement a stricter but more linguistically faithful modeling of the composition of natural language form and meaning.

In this chapter, we will isolate some key concepts from this frantic timeline and expound a bit on their details. Other than reinvented notation or perhaps some fresh example, no novel contributions are to be found here; the intention is merely to establish some common grounds before we get to proceed. If confident in your knowledge of the subject matter, `goto` Chapter II, but at your own risk.

1 The Simple Theory of Types

Simple type theory is the computational formalization of intuitionistic logic. It is in essence an adornment of the rules of intuitionistic logic with the computational manipulations they dictate upon mathematical terms. Dually, it provides a decision procedure that allows one to infer the type of a given program by inspecting the operations that led up to its construction. It is a staple of almost folkloric standing for computer scientists across the globe, tracing its origins to the seminal works of Russell and Church [Russell, 1908; Church, 1940]. The adjective “simple” is not intended as either a diminutive nor a condescending remark pertaining to the difficulty of the subject matter, but rather to distinguish it from the broader class of intuitionistic type theories, which attempt to systematize the notions of quantification (universal and existential), stratification of propositional hierarchies, and more recently equivalence (neither of which we will concern ourselves with).

Our presentation will begin with intuitionistic logic. Once that is done, we will give a brief account of the Curry-Howard correspondence, which shall allow us to give a computational account of the logic, that being the simply typed λ -calculus.

1.1 Intuitionistic Logic

Intuitionistic logic is due to Arend Heyting [Heyting, 1930], who was the first to formalize Bouwer’s intuitionism. It is a restricted version of classical logic, where the laws of the excluded middle (*tertium non datur*) and the elimination of the double negation no longer hold universally. The first states that one

must choose between a proposition A and its negation $\neg A$ ($A \vee \neg A$), whereas the second that a double negation is equivalent to an identity ($\neg\neg A \equiv A$). The absence of these two laws implies that several theorems of classical logic are no longer derivable in intuitionistic logic, meaning that the logic is weaker in terms of expressivity. On the bright side, it has the pleasant effect that proofs of intuitionistic logic are constructive, i.e. they explicitly demonstrate the formation of a concrete instance of whatever proposition they claim to be proving.

Focusing on the disjunction-free fragment of the logic, we have a tiny recursive language that allows us to define the various shapes of logical *propositions* (or *formulas*).¹ Given some finite set of *propositional constants* (or *atomic formulas*) Prop_0 , and A, B, C arbitrary well-formed propositions, the language of propositions in Backus-Naur form is inductively defined as:

$$A, B, C := p \mid A \rightarrow B \mid A \times B \quad (\text{I.1})$$

where $p \in \text{Prop}_0$. Propositions are therefore closed under the two binary *logical connectives* \rightarrow and \times ; we call the first an *implication*, and the second a *conjunction*. A *complex* proposition is any proposition that is not a member of Prop_0 , and its *primary* (or *main*) connective is the last logical connective used when writing it down according to the grammar (I.1).

Besides propositions, we have *structures*. Structures are built from propositions with the aid of a single binary operation, the notation and properties of which can vary between different presentations of the logic. In our case, we will indicate valid structures with Greek uppercase letters Γ, Δ, Θ , and define structures inductively as

$$\Gamma, \Delta, \Theta := 1 \mid A \mid \Gamma, \Delta \quad (\text{I.2})$$

In other words, structures are an inductive set closed under the operator $-, -$ which satisfies associativity and is equipped with an identity element 1 (the *empty structure*), i.e. a monoid. A perhaps more down-to-earth way of looking at a structure is as a *list* or *sequence* of propositions.

Given propositions and structures, we can next define *judgements*, statements of the form $\Gamma \vdash A$. We read such a statement as a suggestion that from *assumptions* Γ (i.e. a structure of *hypotheses*) one can derive a proposition A . Formulas occurring within Γ are said to occur in *antecedent* position, whereas A is in *succedent* position.

A *rule* is a two-line statement separated by a horizontal line. Above the line, we have a (possibly empty) sequence of judgements, which we call the *premises* of the rule. Below the line, we have a single judgement, which we call the rule's *conclusion*. The rule can be thought of as a formal guarantee that if all of its premises are deliverable, then so is the conclusion. Each rule has an

¹The full logic also includes disjunctive formulas, but we will skip them from this presentation as they are of little interest to us. For brevity, we will from now on use intuitionistic logic to refer to its disjunction-free fragment.

identifying name, written directly to the right of the horizontal line.

Rules may be split in two conceptual categories. *Logical* rules, on the one hand, provide instructions for eliminating and introducing logical connectives. Figure I.1a presents the logical rules of intuitionistic logic. The first rule, the axiom of identity id , contains no premises and asserts the reflexivity of the provability operator \vdash . It states that from a proposition A one can infer, guess what, that very proposition. The remaining logical rules come in pairs, one per logical connective. The elimination of the implication (or *modus ponens*) states that, given a proof of a proposition $A \rightarrow B$ from assumptions Γ and a proof of proposition A from assumptions Δ , one can join the two to derive a proposition B . Dually, the introduction of the implication (or *deduction theorem*) states that from a proof of a proposition B given assumptions Γ, A , one can use Γ alone to derive an implicational proposition $A \rightarrow B$. In a similar manner, the elimination of the conjunction $\times E$ states that, given a proof of a proposition $A \times B$ from assumptions Γ , and a proof that the triplet Δ, A, B allows us to derive a proposition C , one could well use Γ together with Δ to derive C directly. And dually again, the introduction of the conjunction $\times I$ permits us to join two unrelated proofs, one of A from Γ and one of B from Δ into a single proof, that of their product $A \times B$, from Γ joined with Δ .

Structural rules, on the other hand, allow us to manipulate structures (who would have thought); they are presented in Figure I.1b. Structural rules have a two-fold role. First, they explicate an extra property of our structure binding operator, namely commutativity. One could also make do with an implicit exchange rule by treating structures as *multisets* rather than lists – having it explicit, however, will keep us conscious of its presence and strengthen our emotional bond to it, in turn making us really notice its absence when it will no longer be there (it also keeps the presentation tidier). Second, they give an account of the status of propositions as permanent and reusable facts. The weakening rule *weak* states that if we were able to derive a proposition B from some assumptions Γ , we will also be able to do so if the assumptions were to contain some arbitrary extra proposition A . Conversely, the contraction rule *contr* states that if we needed some assumption structure containing two instances of a proposition A to derive a proposition B , we could also make do with just one instance of it, discarding the other without remorse.

A *proof*, finally, is a heterogeneous variadic tree. At its root, it has a judgement, guaranteed to be derivable (provided we did not mess up somewhere), called its *endsequent*. Its branches are themselves proofs, fused together by a rule – the number of premises being the local tree's arity. At its leaves, it has identity axioms – the smallest kind of proof.

1.1.1 Proof Equivalences

The same judgement may be provable in more than one ways. The difference between two proofs of the same judgement can be substantial, when they indeed describe distinct derivation procedures, or trivial. Trivial variations come

$$\begin{array}{c}
\frac{}{A \vdash A} \text{id} \\
\\
\frac{\Gamma \vdash A \rightarrow B \quad \Delta \vdash B}{\Gamma, \Delta \vdash B} \rightarrow E \qquad \frac{\Gamma, A \vdash B}{\Gamma \vdash A \rightarrow B} \rightarrow I \\
\\
\frac{\Gamma \vdash A \times B \quad \Delta, A, B \vdash C}{\Gamma, \Delta \vdash C} \times E \qquad \frac{\Gamma \vdash A \quad \Delta \vdash B}{\Gamma, \Delta \vdash A \times B} \times I
\end{array}$$

(a) Logical rules.

$$\begin{array}{c}
\frac{\Gamma, \Delta \vdash A}{\Delta, \Gamma \vdash A} \text{ex} \\
\\
\frac{\Gamma \vdash B}{\Gamma, A \vdash B} \text{weak} \qquad \frac{\Gamma, A, A \vdash B}{\Gamma, A \vdash B} \text{contr}
\end{array}$$

(b) Structural rules.

Figure I.1: Intuitionistic Logic $\mathbf{IL}_{\rightarrow, \times}$.

in two kinds: syntactic equivalences (i.e. sequences of rule applications that can safely be rearranged) and redundant detours (i.e. sequences of rule applications that can altogether removed).

The first kind is not particularly noteworthy. In essence, we say that two proofs are syntactically equivalent if they differ only in the positioning of structural rule applications. This notion can be formally captured by establishing an equivalence relation between proofs on the basis of commuting conversions.

The second kind is more interesting and slightly more involved. A proof pattern in which a logical connective is introduced, only to be immediately eliminated, is called a *detour* (or β redex). Detours can be locally resolved via proof rewrites – the fix-point of performing all applicable resolutions is called *proof normalization* and yields a canonical proof form. The strong normalisation property guarantees that a canonical form exists for any proof in the logic, and in fact the choice of available rewrites to apply at each step is irrelevant, as all paths have the same end point [de Groote, 1999]. Figure I.2 presents rewrite instructions for the two detour patterns we may encounter (one per logical connective). Read bottom-up¹, the first one suggests that if one were to hypothesize a proposition A , use it within an (arbitrarily deep) proof s together with extra assumptions Γ to derive a proposition B , before finally redacting the hypothesis and composing with a proof t that derives A from assumptions Δ , it would have been smarter (and more concise!) to just plug in t directly when

¹In the small-to-big rather than literal sense! If confused: start from the proof leaves and go down.

$$\begin{array}{c}
\begin{array}{c}
\dots \quad \overline{A \vdash A} \text{ id} \\
\vdots s \\
\hline
\Gamma, A \dots \vdash B \\
\vdots \\
\hline
\Gamma, A \vdash B \\
\hline
\Gamma \vdash A \rightarrow B \quad \xrightarrow{\rightarrow I} \quad \begin{array}{c} \vdots t \\ \hline \Delta \vdash A \end{array} \\
\hline
\Gamma, \Delta \vdash B \quad \xrightarrow{\rightarrow E}
\end{array}
\quad \Rightarrow \quad
\begin{array}{c}
\begin{array}{c} \vdots t \\ \hline \Delta \vdash A \end{array} \\
\vdots s \\
\hline
\Gamma, \Delta \dots \vdash B \\
\vdots \\
\hline
\Gamma, \Delta \vdash B
\end{array}
\end{array}$$

$$\begin{array}{c}
\begin{array}{c}
\overline{A \vdash A} \text{ id} \quad \dots \quad \overline{B \vdash B} \text{ id} \\
\vdots u \\
\hline
\Theta, A, B \dots \vdash C \\
\vdots \\
\hline
\Theta, A, B \vdash C \\
\hline
\Gamma, \Delta, \Theta \vdash C \quad \xrightarrow{\times E}
\end{array}
\quad \Rightarrow \quad
\begin{array}{c}
\begin{array}{c} \vdots s \\ \hline \Gamma \vdash A \end{array} \quad \dots \quad \begin{array}{c} \vdots t \\ \hline \Delta \vdash B \end{array} \\
\vdots u \\
\hline
\Gamma, \Delta, \Theta \dots \vdash C \\
\vdots \\
\hline
\Gamma, \Delta, \Theta \vdash C
\end{array}$$

Figure I.2: Intuitionistic β redexes.

previously hypothesizing A , since then no redaction or composition would have been necessary. In a similar vein, the second suggests that if one were to derive and merge proofs s and t (of propositions A and B , respectively), only to eliminate their product against hypothetical instances of A and B that were used in proof u to derive C (together with assumptions Θ), the proof can be reduced by just plugging s and t in place of the axiom leaves of u . Note the use of horizontal dots at the axiom leaves, denoting simultaneous substitutions of *all* occurrences of redundant hypotheses, and the use of unnamed vertical dots, denoting (invertible) sequences of contr and/or ex rules.

1.2 The Curry-Howard Correspondence

The Curry-Howard correspondence asserts an equivalence between the above presentation of the logic in natural deduction, and a system of computation known as the λ -calculus. It was first formulated by Haskell Curry in the 30s before being independently rediscovered by William Alvin Howard and Nicolaas Govert de Bruijn in the 60s [Curry, 1934; de Bruijn, 1983; Howard, 1980]. The entry point for such an approach is to interpret propositions as *types* of a minimal functional programming language (a perhaps more aptly named alternative to the Curry-Howard correspondence is the *propositions-as-types interpretation*). In that sense, the set of propositional constants Prop_0 becomes the programming language's basic set of *primitive* types (think of them as built-ins). Implicational formulas $A \rightarrow B$ are read as *function* types, and conjunction formulas are read as *tuple* (or cartesian product) types. From now we will use

$$\begin{array}{c}
\frac{}{x_i : A \vdash x_i : A} \text{ id} \\
\\
\frac{\Gamma \vdash s : A \rightarrow B \quad \Delta \vdash t : B}{\Gamma, \Delta \vdash s t : B} \rightarrow E \qquad \frac{\Gamma, x_i : A \vdash s : B}{\Gamma \vdash \lambda x_i. s : A \rightarrow B} \rightarrow I \\
\\
\frac{\Gamma \vdash s : A \times B \quad \Delta, x_i : A, x_j : B \vdash t : C}{\Gamma, \Delta \vdash \text{case } s \text{ of } (x_i, x_j) \text{ in } t : C} \times E \qquad \frac{\Gamma \vdash s : A \quad \Delta \vdash t : B}{\Gamma, \Delta \vdash (s, t) : A \times B} \times I \\
\\
\frac{\Gamma, \Delta \vdash s : A}{\Delta, \Gamma \vdash s : A} \text{ ex} \\
\\
\frac{\Gamma \vdash s : B}{\Gamma, x_i : A \vdash s : B} \text{ weak} \qquad \frac{\Gamma, x_i : A, x_j : A \vdash s : B}{\Gamma, x_k : A \vdash s_{[x_i \mapsto x_k, x_j \mapsto x_k]} : B} \text{ contr}
\end{array}$$

Figure I.3: Simple type theory.

formulas, propositions and types interchangeably. Following along the correspondence allows us to selectively speak about individual, named instances of propositions – we call these *terms*. The simplest kind of term is a *variable*, corresponding to a hypothesis in the proof tree. Each logical rule is identified with a programming pattern: the axiom rule is variable *instantiation*, introduction rules are *constructors* of complex types, and elimination rules are their *destructors*. The question of whether a logical proposition is provable translates to the question of whether the corresponding type is inhabited; i.e. whether an object of such a type can be created – we will refer to the latter as a *well-formed* term.

Rather than present a grammar of terms and later ground it in the logic, we will instead simply revisit the rules we established just above, now adorning each with a term rewrite instruction – the result is a tiny yet still elegant and expressive type theory, presented in Figure I.3. Given an infinite but enumerable set Vars consisting of (unique names for) indexed variables with elements $\{x_i, x_j, x_k, x_l, \dots\}$, and denoting arbitrary but well-formed terms with s, t, u , we will use $s : A$ (or s^A) to indicate that term s is of type A . Assumptions Γ, Δ will now denote a *typing environment*:

$$x_1 : A_1, x_2 : A_2 \dots x_n : A_n \tag{I.3}$$

i.e. rather than a sequence of formulas, we have a sequence of distinct variables, each of a specific type, and a judgement $\Gamma \vdash s : B$ will now denote the derivation of a term s of type B out of such an environment.

Inspecting Figure I.3, things for the most part look good. The implication

elimination rule $\rightarrow E$ provides us with a composite term $s\ t$ that denotes the function application of *functor* s on *argument* t . Function application is left-associative: $s\ t\ u$ is the bracket-economic presentation of $(s\ t)\ u$ – we have no choice but to use brackets if want to instead denote $s\ (t\ u)$. The dual rule, $\rightarrow I$, allows us to create (so-called anonymous) functions by deriving a result s dependent on some hypothesized argument x_i which is then *abstracted* over as $\lambda x_i.s$. Any occurrence of x_i within s is then *bound* by the abstraction; variables that do not have a binding abstraction are called *free*. The conjunction introduction $\times I$ allows us to create tuple objects (s, t) through their parts s and t . Its dual, $\times E$, gives us the option to identify the two coordinates of a tuple s with variables x_i and x_j , when the latter are hypothesized assumptions for deriving some program t . If our assumptions are not in order, blocking the applicability of some rule, we can put them back where they belong with *ex*. With *contr* we can pretend to be using two different instances x_i and x_j of the same type before identifying the two as a single object x_i in term s with term t "); note here the meta-notation for *variable substitution*, $s_{[x_i \mapsto t]}$, which reads as “replace any occurrence of variable x_i . And finally, we can introduce throwaway variables into our typing environment with *weak* (arguably useful for creating things like constant functions).

There’s just a few catches to beware of. The first has to do with tracing variables in a proof; the concatenation of structures Γ, Δ is only valid if Γ and Δ contain no variables of the same name; if that were to be the case, we would be dealing with variable shadowing, a situation where the same name could ambiguously refer to two distinct objects (a horrible thing). The second has to do with do with the *ex* rule. The careful reader might notice that the rule leaves no imprint on the term level, meaning we cannot distinguish between a program where variables were *a priori* provided in the correct order, and one where they were shuffled into position later on. This is justifiable if one is to treat the rule as a syntactic bureaucracy that has no real semantic effect, i.e. if we consider the two proofs as equivalent, following along the commuting conversions mentioned earlier (supporting the idea that in this type theory, assumptions are multisets rather than sequences). A slightly more perverse problem arises out of the product elimination rule $\times E$. The rule posits that two assumptions $x_i : A$ and $x_j : B$ can be substituted by a single (derived) term of their product type $s : A \times B$. Choosing different depths within the proof tree upon which to perform this substitution will yield distinct terms (because indeed they represent distinct sequences of computation); whether there’s any merit in distinguishing between the two is, however, debatable. Finally, whereas other rules can be read as syntactic operations on terms, (this presentation of) the *contr* rule contains meta-notation that is not part of the term syntax itself. That is to say, $s_{[x_i \mapsto t]}$ is *not* a valid term – even if the result of the operation it denotes is. Generally speaking, substitution of objects for others of the same type is (modulo variable shadowing) an admissible property of the type system. Mixing syntax and meta-syntax in the same system is a dirty but useful trick people sporadically employ; this surely invites some

trouble, but conscious use of it can be worth it, since it significantly simplifies presentation.

1.2.1 Term Equivalences

There exist three kinds of equivalence relations between terms, each given an identifying Greek letter.¹

α conversion is a semantically null rewrite obtained by renaming variables according to the substitution meta-notation $s_{[x_i \mapsto x_j]}$ described above. Despite seeming innocuous at a first glance, α conversion is an evil and dangerous operation that needs to be applied with extreme caution so as to avoid variable capture, i.e. substituting a variable's name with one that is already in use. Two terms are α equivalent if we can rewrite one into the other using just α conversions, e.g.

$$\lambda x_i. x_i^A \equiv \lambda x_j. x_j^A \quad (\text{I.4})$$

Standardizing variable naming, e.g. according to the distance between variables and their respective binders, alleviates the effort required to check for α equivalence by casting it to simple syntactic equality (string matching).

β reduction The term rewrites we have so far inspected were either provided by specific rules, or were notational overhead due to the denominational ambiguity of variables. Aside from the above, our type system provides two minimal computation steps that tell us how to reduce expressions that involve the deconstruction of a just-constructed type:

$$(\lambda x_i. s) \, t \xrightarrow{\beta} s_{[x_i \mapsto t]} \quad (\text{I.5})$$

$$\text{case } (s, t) \text{ of } (x_i, x_j) \text{ in } u \xrightarrow{\beta} u_{[s \mapsto x_i, t \mapsto x_j]} \quad (\text{I.6})$$

A term on which no β reductions can be applied is said to be in β -normal form. The Church-Rosser theorem asserts first that one such form exists for all well-formed terms, and second, that this form is inevitable and inescapable – any reduction strategy followed to the end will bring us to it [Barendregt et al., 1984]. Two terms are β equivalent to one another if they both reduce to the same β -normal form.

If you are at this point getting a feeling of *deja vu*, rest assured this is not on you; we have indeed gone through this before, last time around with proofs rather than terms. If one were to replicate the above term reductions with their corresponding proofs, they would end up exactly with the proof reduction patterns of Figure I.2. I will spare you the theatrics of faking surprise

¹The denotational significance of these letters I have yet to understand – legend has it that it only starts making sense after having written your 10th compiler from scratch.

Logic	Computer Science
Propositional Constant	Base Type
Logical Connectives	Type Constructors
Implication	Function Type
Conjunction	Product Type
Axiom	Variable
Introduction Rule	Constructor Pattern
Elimination Rule	Destructor Pattern
Proof Normalization	Computation
Provability	Type Inhabitation

Table I.1: The Curry-Howard correspondence in tabular form.

at this fact, but if this not something you were exposed to previously, take a moment here to marvel at the realization that proof normalization is in reality “just” computation. This ground-shattering discovery lies at the essence of the Curry-Howard correspondence.

η conversion In contrast to β conversion, which tells us how to simplify an introduce-then-eliminate pattern, η conversion tells us how to modify an eliminate-then-introduce pattern. An η long (or normal) form of a term is one in which the arguments to type operators are made explicit (i.e. all introductions of a connective are preceded by its elimination), whereas an η contracted (or pointfree) form is one where arguments are kept hidden [Prawitz, 1965]. We refer to the simplification of an expanded form as η reduction, which is the computational dual of β reduction; the reverse process is an η expansion. Both directions are facets of η conversion – the equivalence relation enacted by this conversion is called η equivalence.

$$\lambda x_i. s \stackrel{\eta}{=} s \quad (\text{I.7})$$

$$(\text{case } s \text{ of } (x_i, x_j) \text{ in } x_i, \text{ case } s \text{ of } (x_k, x_l) \text{ in } x_l) \stackrel{\eta}{=} s \quad (\text{I.8})$$

1.2.2 In Place of a Summary

Table I.1 summarizes the subsection.

1.3 Intermezzo

We now know how to prove things (or compute with types). Before moving along with this chapter’s agenda, we will take a brief pause to provide some auxiliary definitions and notations that should prove relevant later on. This is also a chance to do a bit of warming up with some baby examples before some real world proofs start coming our way.

Table I.2: Polarity induction.

Figure I.4: Type raising.

Type Raising Type raising $A^+ \vdash (A^- \rightarrow B^+) \rightarrow B^-$ is a derivable theorem of intuitionistic logic presented in Figure I.4. It states that for A , B arbitrary propositions, from A one can derive its raised form $A \rightarrow B$. The converse, i.e. type lowering, is not true: $(A^+ \rightarrow B^-) \rightarrow B^+ \not\vdash A^-$.

$$\begin{aligned} \mathcal{O}(p \in \text{Prop}_0) &:= 0 \\ \mathcal{O}(A \rightarrow B) &:= \max(\mathcal{O}(A) + 1, \mathcal{O}(B)) \end{aligned} \tag{I.9}$$

Types whose order is above 1 are called *higher-order* types; they denote functions that accept functions as their arguments. For instance, for p and s atomic propositions of order 0, their respective identity functions $p \rightarrow p$ and $s \rightarrow s$ are of order 1, and the raised form of p into s , i.e. $(p \rightarrow s) \rightarrow s$, is of order 2.

Notational Shorthands The verbosity of term-decorated proofs can get cumbersome in the long run, and does not play well with the unforgiving horizontal margins enforced by the template imposed on writer and reader alike. It is probably inevitable that at some point proofs will need a smaller font size to fit in a page (or, worse yet, some neck-breaking rotations of the orientation plane), but in a futile attempt to postpone such emergency measures, we will occasionally make use of a shorthand notation for natural deduction proofs that avoids repetition, at the cost of maybe requiring some extra time to visually parse. In this notation, axioms will be rewritten as follows:

$$\frac{}{x_i : A \vdash x_i : A} \text{id} \quad =: \quad \frac{}{x_i : A} \text{id}$$

And assumptions will appear without type assignments (if uncertain of what some variable's type is, just trace it back to its axiom). We will always provide type declarations for derived terms (right of the turnstile). The examples of the next paragraph (and many of the ones to follow) will use this alternative notation.

Currying A product type occurring in the argument position of an implication is interderivable with a longer implication where its coordinates are sequentialized: $(A \times B) \rightarrow C \dashv\vdash A \rightarrow B \rightarrow C$. The forward direction is called currying, and the backward uncurrying; you can find a proof for each in Figure I.5. Having proven that once, we can reuse that proof for deriving implicational equivalents from conjunctions (including nested ones, provided they occur as arguments to an implication). Combined with type raising, this trick is interesting, as it permits us to indirectly argue about product types as higher-order implications, even in presentations of the theory that do not include an explicit product (and thus avoid the issues related to its elimination), e.g. we have:

$$A \times B \vdash ((A \times B) \rightarrow C) \rightarrow C \dashv\vdash (A \rightarrow B \rightarrow C) \rightarrow C \quad (\text{I.10})$$

Keep a mental note.

Proof Search Attempting to derive a judgement of the form $\Gamma \vdash A$ amounts to searching for a suitable proof of that statement, a process called *proof search*. We distinguish two directions of proof search: the *backward chaining* (or top-down) approach starts from the goal judgement and iteratively expands it into judgements with smaller assumptions – one judgement per premise generated

$$\frac{\frac{\frac{}{x_0 : (A \times B) \rightarrow C} \text{id} \quad \frac{\frac{}{x_1 : A} \text{id} \quad \frac{}{x_2 : B} \text{id}}{x_1, x_2 \vdash (x_1, x_2) : A \times B} \times I}{x_0, x_1, x_2 \vdash x_0 (x_1, x_2) : C} \rightarrow E}{x_0 \vdash \lambda x_1 x_2. x_0 (x_1, x_2) : A \rightarrow B \rightarrow C} \rightarrow I(x_2)$$

(a) Currying

$$\frac{\frac{\frac{}{x_0 : A \times B} \text{id} \quad \frac{\frac{\frac{}{x_1 : A \rightarrow B \rightarrow C} \text{id} \quad \frac{}{x_2 : A} \text{id}}{x_1, x_2 \vdash x_1 x_2 : B \rightarrow C} \rightarrow E \quad \frac{}{x_3 : B} \text{id}}{x_1, x_2, x_3 \vdash x_1 x_2 x_3 : C} \rightarrow E}{x_0, x_1 \vdash \text{case } x_0 \text{ of } (x_2, x_3) \text{ in } x_1 x_2 x_3 : C} \times E}{x_1, x_0 \vdash \text{case } x_0 \text{ of } (x_2, x_3) \text{ in } x_1 x_2 x_3 : C} \text{ex}}{\frac{}{x_1 \vdash \lambda x_0. \text{case } x_0 \text{ of } (x_2, x_3) \text{ in } x_1 x_2 x_3 : (A \times B) \rightarrow C} \rightarrow I}$$

(b) Uncurrying

Figure I.5: Interderivability of product and arrow.

by the rule of inference applied – with the intention being the eventual deconstruction of all branches into axioms of identity. The other direction is called *forward chaining* (or bottom-up), and starts from a collection of hypothesized propositions (axioms) that are glued together to form progressively more complex structures, until the goal judgement is reached. Without digressing further, it is important to realize that both directions suffer the same issue, albeit from different angles, namely hypothetical reasoning. Forward chaining requires a perfect guess of any and all propositions required in deriving A from Γ , even those that will be redacted and thus never occur in Γ . Dually, backward chaining might require introduction of substructures and subformulas that are nowhere to be found in either the antecedents or the succedent of the current judgement due to the modus ponens-like behavior of implication elimination. Long story short, proof search is hard.

2 Going Linear

We are now ready to start charting grounds in substructural territories: we will gradually impoverish our logic by removing structural rules one by one, and see where that gets us. The weakest links are the *contr* and *weak* rules. These two rules are a cultural and ideological remnant of a long-gone age infested by delusions of prosperity and abundance. In their presence, propositions are proof objects that can be freely replicated and discarded. Removing them (or controlling their applicability via other means) directs us towards a more eco-

conscious regime by turning propositions into finite resources, the production and/or consumption of which is not to be taken for granted. Removing contr yields Affine Logic, a logic in which resources can be used no more than once. Removing weak yields Relevant Logic, a logic in which resources can be used no less than once. Removing both yields Linear Logic, a logic in which resources can be used *exactly* once. The intuitionistic formulations of the above give rise to corresponding type theories [Pierce, 2004]. For the purposes of this manuscript, we will focus our presentation on linear type theory.

2.1 Linear Types

Linear logic is due to Jean-Yves Girard [Girard, 1987], and its computational interpretation due to Samson Abramsky [Abramsky, 1993]. The full logic includes two disjunctive connectives as well as a modality that allows one to incorporate non-linear propositions into the presentation, but we will happily forget about those. Note that with these missing connectives included, the logic is not impoverished but rather enhanced – full linear logic in fact subsumes intuitionistic logic; we have no use of this much expressivity here though. Insights from the previous section carry over to this one; we will no longer separate the presentation between the logic and the type theory, but instead do both in one go.

For the fragment of interest to us, the type grammar becomes:

$$A, B, C := p \mid A \multimap B \mid A \otimes B \mid A \& B \quad (\text{I.11})$$

There is not really much we have to do to manipulate these new types, other than a slight cognitive rewiring. We will note first that the meaning of the implication arrow changes from material implication to transformation process; i.e. where we previously had $A \rightarrow B$ to denote that B logically follows from A , we will now have $A \multimap B$ to denote an irreversible process that transforms a single A into a single B , consuming the former in the process (we can think of this as a perfect chemical reaction). The new, weird-looking arrow of linear implication is read as *lolli*(pop) due to its suggestive appearance.¹ Conjunction \times is now separated into two distinct operators, the multiplicative \otimes and the additive $\&$. The first denotes a linear tuple, and $A \otimes B$ is read as *both A and B*. A linear tuple offers no possibility of projection: we will need to use both coordinates going forward. The second denotes a choice, and $A \& B$ is read as *A with B*, or *choose one of A or B*. This choice is *external*, as the freedom of applying it lies with the operator rather than the proof, and is manifested by the presence of two eliminators for our new connective: a left projection $\&E_1$ and a right projection $\&E_2$; choosing one means we lose the possibility of obtaining the other. Unique to the $\&I$ rule is the fact that two proof branches used to derive each coordinate of the $A \& B$ conclusion share the same assumptions

¹If trying to typeset it yourself, DO NOT duckduckgo for “lolli in latex”. It can be found as `\multimap`. You are welcome.

$$\begin{array}{c}
\frac{}{x_i : A \vdash x_i : A} \text{id} \\
\\
\frac{\Gamma \vdash s : A \multimap B \quad \Delta \vdash t : A}{\Gamma, \Delta \vdash st : B} \multimap E \qquad \frac{\Gamma, x_i : A \vdash s : B}{\Gamma \vdash \lambda x_i. s : A \multimap B} \multimap I \\
\\
\frac{\Gamma \vdash s : A \otimes B \quad \Delta, x_i : A, x_j : B \vdash t : C}{\Gamma, \Delta \vdash \text{case } s \text{ of } (x_i, x_j) \text{ in } t : C} \otimes E \qquad \frac{\Gamma \vdash s : A \quad \Delta \vdash t : B}{\Gamma, \Delta \vdash (s, t) : A \otimes B} \otimes I \\
\\
\frac{\Gamma \vdash s : A \& B}{\Gamma \vdash \text{fst}(s) : A} \&E_1 \qquad \frac{\Gamma \vdash s : A \& B}{\Gamma \vdash \text{snd}(s) : B} \&E_2 \qquad \frac{\Gamma \vdash s : A \quad \Gamma \vdash t : B}{\Gamma \vdash \langle s, t \rangle : A \& B} \&I \\
\\
\text{(a) Logical rules.} \\
\\
\frac{\Gamma, \Delta \vdash s : A}{\Delta, \Gamma \vdash s : A} \text{ex} \\
\\
\text{(b) Structural rule.}
\end{array}$$

Figure I.6: Linear Logic $\mathbf{ILL}_{\multimap, \otimes, \&}$ and its type theory.

Γ . The subset of linear logic concerning the connectives discussed is presented in Figure I.6, together with its term rewrites (assumptions, judgements, rules and proofs look just like before). For the sake of homogeneity and explicitness, ex still makes an appearance as the sole structural rule.

Notationally, the absence of non-linear intuitionistic terms allows us to freely reuse our prior term notation without fear of ambiguity. We have three new term patterns: $\langle s, t \rangle$ to denote the choice between proof terms s and t (contrast to the linear tuple (s, t)), and $\text{fst}(_)$ and $\text{snd}(_)$ to denote the first and second projections. Similarly for the implication introduction rule $\multimap I$, no redundant variables means that x_i must now appear free once in the abstraction body s – in other words, we have no way of syntactically instantiating constant functions.

2.1.1 Proof & Term Reductions

The notions of proof and term equivalence discussed in the previous section hold also for linear logic. Proof normalization looks almost identical to before in the case of \rightarrow and \times (substituting of course for \multimap and \otimes). The key difference lies in the absence of horizontal dots and unnamed vertical dots (since the contr rule is no more, meaning that there can only be a single occurrence of each axiom replaced with a proof). The extra connective $\&$ introduces its own two redexes (one per eliminator); the reduction of the first projection is shown in Figure I.7. Its reading is straightforward: if one were to use Γ to independently derive A and B along two parallel proofs, proceed by constructing

$$\begin{array}{c}
\frac{\overline{A \vdash A} \text{ id} \quad \frac{\Gamma, A \vdash B}{\Gamma \vdash A \multimap B} \multimap I \quad \frac{\Delta \vdash A}{\Gamma, \Delta \vdash B} \multimap E}{\Gamma, \Delta \vdash B} \multimap E \quad \Rightarrow \quad \frac{\frac{\Delta \vdash A}{\Gamma, \Delta \vdash B} \quad \frac{\vdots t}{\Delta \vdash A}}{\Gamma, \Delta \vdash B}
\end{array}$$

$$\begin{array}{c}
\frac{\frac{\frac{\vdots s}{\Gamma \vdash A} \quad \frac{\vdots t}{\Delta \vdash B}}{\Gamma, \Delta \vdash A \otimes B} \otimes I \quad \frac{\overline{A \vdash A} \text{ id} \quad \overline{B \vdash B} \text{ id} \quad \frac{\vdots u}{\Theta, A, B \vdash C}}{\Gamma, \Delta, \Theta \vdash C} \otimes E}{\Gamma, \Delta, \Theta \vdash C} \otimes E \quad \Rightarrow \quad \frac{\frac{\vdots t}{\Gamma \vdash A} \quad \frac{\vdots u}{\Delta \vdash B}}{\Gamma, \Delta, \Theta \vdash C}
\end{array}$$

$$\begin{array}{c}
\frac{\frac{\frac{\vdots s}{\Gamma \vdash A} \quad \frac{\vdots t}{\Gamma \vdash B}}{\Gamma \vdash A \& B} \& I}{\Gamma \vdash A} \& E_1 \quad \Rightarrow \quad \frac{\vdots s}{\Gamma \vdash A}
\end{array}$$

Figure I.7: Linear β redexes.

a choice between the two, and then also make that choice (favoring either), there was never any need for the other in the first place. The equivalent term reduction steps this time around are:

$$(\lambda x_i. s) \, t \xrightarrow{\beta} s_{[x_i \mapsto t]} \quad (\text{I.12})$$

$$\text{case } (s, t) \text{ of } (x_i, x_j) \text{ in } u \xrightarrow{\beta} u_{[s \mapsto x_i, t \mapsto x_j]} \quad (\text{I.13})$$

$$\text{fst}(\langle s, t \rangle) \xrightarrow{\beta} s \quad (\text{I.14})$$

$$\text{snd}(\langle s, t \rangle) \xrightarrow{\beta} t \quad (\text{I.15})$$

2.2 Proof Nets

We have basked in the beauty of the natural deduction presentation adopted so far, and seen how it gives rise to a straightforward computational interpretation. We have also seen how it is at times overly bureaucratic in its explication of structural rules that are null from a computational perspective. To our luck, an additional representation makes itself available as soon as we step into linear grounds, namely *proof nets*, also due to Girard [Girard, 1987]. Proof nets are best suited for the multiplicative fragment of the logic (they are amenable to extensions with additive connectives, but things get uglier there). In our case, we have foregone the disjunctive connectives, and we have already

$$\begin{array}{c}
\frac{}{x_1 : B \multimap C} \text{id} \quad \frac{}{x_0 : A \multimap B} \text{id} \quad \frac{}{x_2 : A} \text{id} \\
\frac{}{x_0, x_2 \vdash x_0 \ x_2 : B} \multimap E \\
\frac{x_1, x_0, x_2 \vdash x_2 (x_1 \ x_0) : A \multimap C}{x_1, x_0 \vdash \lambda x_2. x_1 (x_0 \ x_2) : A \multimap C} \multimap I \\
\frac{x_1, x_0 \vdash \lambda x_2. x_1 (x_0 \ x_2) : A \multimap C}{x_0, x_1 \vdash \lambda x_2. x_1 (x_0 \ x_2) : A \multimap C} \text{ex}
\end{array}$$

Figure I.8: Linear function composition.

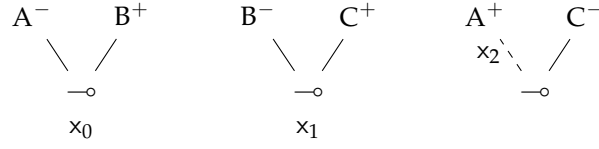
suggested that the multiplicative conjunction \otimes is (to an extent) interchangeable with the implication arrow \multimap (we actually did this for intuitionistic connectives, but there is no need to repeat ourselves – the story looks identical with their linear variants). This gives us the much needed excuse to justify limiting our presenting of proof nets to the implication-only fragment of linear logic, **ILL** $_{\multimap}$, where things are easy and intuitive.

In natural deduction, our proofs are built sequentially. We start with some hypothesized variables and combine them via rules to derive more complex terms, which then serve as premises for a next iteration of rule applications. As long as we are careful not to get stuck in some detour loop hell, we rinse and repeat, and eventually, after a finite number of steps, we end up with our conclusion, at which point we can call it a day. Proof nets offer an appealing alternative: they are parallel, in the sense that they allow multiple conclusions to be derived simultaneously. They also have no notion of temporal precedence: everything happens in a single instant, meaning that positive subformulas are good to use without having to wait for their conditions to be met.

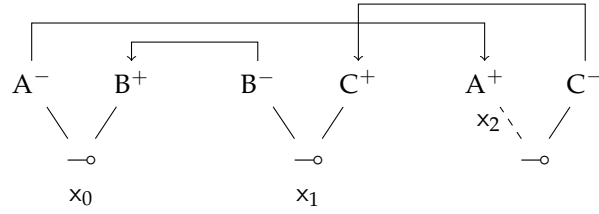
To see this in practice, a simple but concrete example will prove helpful. We will consider the natural deduction proof of linear function composition of Figure I.8 and translate it into its proof net equivalent of Figure I.9.

The first thing we need to do is write formulas as their *decomposition trees*; it sounds fancy, but in reality this is just recompiling formulas according to their underlying grammar, but now in tree form: propositional constants become leaves, and logical connectives become branch nodes. To make this a tiny bit more interesting, we will decorate atomic subformulas with a superscript denoting their polarity, according to the schema of Subsection 1.3. For formula trees with positive roots (i.e. trees occurring left of the turnstile), we will put a label beneath them to identify them with the variable that provides them. We will also distinguish tree edges originating from a negative implication and pointing to a positive tree by marking them with dashed lines – these denote positively rooted trees that are either nested within a higher-order positive implication, or in the argument position of an implication right of the turnstile.¹ Such positive formulas play the role of hypotheses that have been abstracted

¹An alternative notation employs two distinct symbols for the positive and negative versions of an implication. In the literature, these can be encountered as tensor (\otimes) and par (\wp) links.



(a) Proof frame.



(b) Proof structure.

Figure I.9: Proof net construction for the proof of Figure I.8.

over, therefore we need to also give these a name; we can put it right next to the dashed line. An arrangement of the decomposition trees of all formulas as they occur within a judgement is called a *proof frame*; the one for our running example can be seen in Figure I.9a.

A proof frame must satisfy an invariance property before the eligibility of the judgement it prescribes can even be considered. Namely, it must contain an equal number of positive and negative occurrences of each unique atomic formula that appears within it. This perfectly fits the linear logic paradigm: everything produced has to be consumed, and vice versa. In our case, we have three unique formulas, A , B and C , each one of which has a single positive and a single negative occurrence: check. The next thing to do in building a proof net is to establish a bijection between atomic formulas of opposite polarity, and draw it as an extra set of edges, pointed negative atoms to their positive matches: we call those *axiom links*. Axiom links essentially specify the elimination of function types: they identify function arguments with their concrete inputs in a geometric fashion. We do not need to put much thought for our running example: since all atoms have a single occurrence of each polarity, there is just one possible bijection to consider, presented in Figure I.9b. A proof frame with axiom links on top is called a *proof structure*, and this representation provides all of the information contained within a proof.

Alas, the relation from proofs structures is not one-to-one: there exist many more proof structures than proofs. A proof structure is a *proof net* if and only if it satisfies a correctness criterion. There have been various formulations of this criterion, each with a different time complexity, ranging from exponential to linear [Girard, 1987; Danos and Regnier, 1989; Murawski and Ong,

2000; Guerrini, 2011]. We will adopt an (informally rephrased) version of the acyclicity and connectedness criterion of Danos and Regnier [1989]. We are going to treat a proof structure as a heterogeneous graph, consisting of two node types, logical connectives and atomic formulas, and three edge types: tree-structure edges, further subcategorized according to their polarity, and axiom links. We will then attempt a traversal of that graph using an algorithm defined on the basis of the above parameterization, that further utilizes the ancillary tree labels we have assigned earlier. At each step of the traversal, the algorithm will write down a (partial) term or term instruction. We will expect the traversal to be terminating (i.e. to not get stuck in a loop) and complete without repeats (i.e. to have passed through all nodes and edges exactly once), and the term it has transcribed at its last step to be well-formed, at which point we will happily claim the proof structure to indeed be a proof net.

Traversing ILL_{\rightarrow} Nets Now, let's get our crayons out and sketch the outline of the traversal algorithm. We will have two traversal modes: negative (or upward) mode and positive (or downward) mode. In negative mode, we move upwards along negative nodes. When encountering a negative implication, we will create a λ abstraction over the variable specified by its dashed edge. When encountering a negative leaf, we will traverse across the axiom link to its positive counterpart and switch to positive mode. In positive mode, we move downwards along positive nodes. When we encounter a positive implication, we will add its negative (upward) branch to our mental stack and proceed downwards. Upon running out of positive nodes to visit, we will write down the variable label assigned to the positive tree's root, and then perform a negative traversal of each of the negative branches that live in our stack (i.e. we have encountered going down), in reverse order. We will start from the root of the formula tree occurring right of the turnstile (we know which one that is by the fact it has no variable label underneath it) in negative mode.

In our case, this is a negative implication, the dashed line of which reads x_2 , so we start by writing down $\lambda x_2.(\dots)$. We move on to C^- , which is a leaf, so we cross over to C^+ and switch to positive mode. Going down, we encounter an implication as the positive root, and write down the positive tree's name, getting $\lambda x_2.x_1(\dots)$. We proceed in negative mode to B^- , cross over to B^+ and repeat the above, getting $\lambda x_2.x_1(x_0 \dots)$ before going into negative mode again. The final axiom link points us to A^+ , which is its own root, named x_2 . At this point, our traversal has transcribed the term $\lambda x_2.x_1(x_0 x_2)$ and we have ran out of paths to explore. By now, all our nodes and edges have been visited, and our final term is both well-formed and identical to the one prescribed by the natural deduction proof of Figure I.8: a joyful outcome! If we consider ourselves bound to the permutation of assumptions dictated by the variable indices, the only thing we would need to do going backwards is guess the presence (and position) of the ex rule.

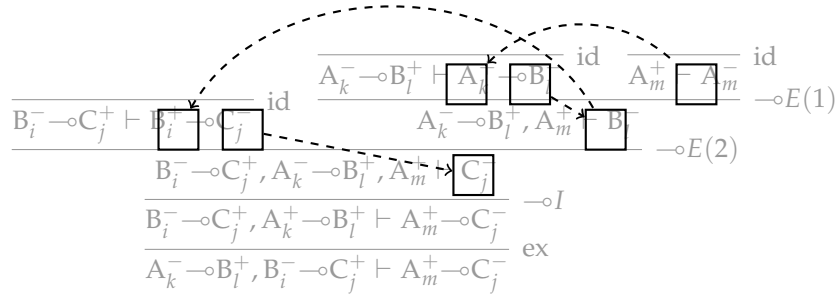


Figure I.10: Extracting axiom links from the proof of Figure I.8.

Axiom Links and Where to Find Them It would be understandable if at this point we allowed ourselves a feeling of complacency; navigating a proof net is no small feat, after all. But upon careful inspection, you might realize that you have been tricked. There never was any room for error in transitioning from the proof frame of Figure I.9a to the proof structure of Figure I.9b. A rare and lucky coincidence, or perhaps the result of carefully planned concealment? Regardless of what it might have been, we cannot reasonably anticipate there to always be a single possible set of axiom links, so we need a decision procedure that tells us how to actually extract them from a less forgiving proof.

Let's revisit the natural deduction proof of Figure I.8. This time around, we will explicate polarity information, and put an identifying index to each atomic formula occurrence at its axiom leaves; the turnstile mirrors indices faithfully, but inverts atomic polarities. Going bottom up through Figure I.10, every encounter of an implication elimination allows us (i) to identify the set of indices coming from the functor's negative part with the set of indices provided by the counterpart positive argument and (ii) to propagate the negative indices of the functor's remainder to the succedent of the next conclusion. That is, the first elimination $\rightarrow E(1)$ creates the identification $\{k \leftrightarrow m\}$, and propagates the positive leftover B_l^- to the next proof step, whereas the next elimination $\rightarrow E(2)$ identifies $\{i \leftrightarrow l\}$ and propagates C_j^- downwards. Upon reaching the proof's conclusion, we get to merge all identifications established along the proof¹, which can then be applied as a mapping (e.g. to the lexico-

¹Let's avoid any misteps here. The joining described is *not* set-theoretic union. Rather, we first take the set-theoretic union, and then iteratively reduce the set by conflating all identifications that agree in at least one element, up to a fixpoint. For instance, joining $\{i \leftrightarrow k\}$ and $\{l \leftrightarrow i, j \leftrightarrow m\}$ yields $\{i \leftrightarrow k \leftrightarrow l, j \leftrightarrow m\}$. Such a situation could arise for example when attempting to find the axiom links of non β normal proofs, like

$$\left(\lambda x_0. x_0^{A_i \rightarrow B_j} x_1^{A_k} \right) x_2^{A_l \rightarrow B_m}$$

The added burden has the enormous benefit of yielding " β normalized" links and resolving potential future headaches *a priori*.

graphically first element) yielding a link-decorate judgement, in our case:

$$A_k^- \multimap B_i^+, B_i^- \multimap C_j^+ \vdash A_k^+ \multimap C_j^-$$

Matching indices correspond exactly to the axiom links of Figure I.9b – the two representations are in fact equivalent. Now, we really do know how to freely move back and forth between the proof net and natural deduction presentation of proofs in \mathbf{ILL}_{\multimap} .

The question then is: when should we use which? The original intention of proof nets was to provide a compact, bureaucracy-free representation of proofs that abstracts away from structural rules. In that sense, their strength is also their weakness; same as the λ -terms they prescribe, they encode the semantically essential part of a proof, but hide structural subtleties that can prove hard to guess or recover. At the same time, performing search over proof nets is a horrible idea; the number of possible links we need to consider scales factorially with respect to the number of atoms in the proof frame, and checking whether a set of links is valid is in the best case linear. Due to these limitations, proof nets were envisaged as a compiled form of an existing proof, rather than a canvas to find that proof on. We will not see proof nets again for a while, but we will keep their memory warm in our hearts. Because when we do, we will challenge this perception and see how their parallel nature can actually be very convenient for proof search. Until then, we can temporarily store them in our mental backlog.

3 Lambek Calculi

3.1 Dropping Commutativity

There is only one structural rule left¹: it is time for ex to go. Dropping ex makes the structures of our logic non-commutative. The transition, however, requires some care. If we were to naively go about our business using the inherited \mathbf{ILL} connectives, we would soon stumble upon a pitfall. Recalling the shape of the $\multimap E$ rule, we come to the realization that functions carried over to this new logic are suddenly picky; they can only be applied to arguments to their right. This should raise some flags: a directionally flavoured version of the implication is not bad in itself, but the presence of just such one such version is – where shall we look for the left-biased one? The answer is simple: the conflation between the two directions was natural, up until a moment ago; having them both would not amount to much, since by ex they would be interderivable. With ex removed, the veil is lifted and we can now see this clearly: there were always two implications, except disguised by the same symbol! Let us do our newfound friend justice, and make this distinction explicit.

¹Or is there?

The logic that provides us with the tools to accomplish this is due to Jim Lambek [Lambek, 1958], and has come to be known as the Lambek calculus L . At this point, the careful reader will notice a chronological inconsistency in our presentational tour: the Lambek calculus predates Linear Logic! Despite the fact, it is in essence a *refinement* of its purely linear part – a substructural logic within a substructural logic – and our previous exposition makes us better equipped to appreciate it. With commutativity gone, the Lambek calculus brings *order* to Linear Logic – in the literal sense – assumptions must now be used exactly in the order they were instantiated. It also brings forth the notion of *adjacency*: structures joined by a rule are now immobile, and therefore obliged to remain adjacent from then on, unless broken apart by abstractions.

Formulas in the Lambek calculus are generated by the grammar:

$$A, B, C := p \mid A \backslash B \mid A / B \mid A \otimes B \quad (\text{I.16})$$

The rules of this fragment are presented in Figure I.11. Alternative presentations can include additive conjunction and/or either of the disjunctions, but the key feature of interest lies in the two implications, $/$ and \backslash . The intuitive way of reading those is as directed fractionals, the formula hidden under the cover of the slash being the divisor, and the formula lying on it the dividend. The elimination rule $/E$ (resp. $\backslash E$) can then be read as fractional simplifications, whereby right (resp. left) multiplication by the divisor cancels out the division as a whole. An analogous reading can be attributed to the introduction rules, them now being the instantiation of a division by withdrawing items from the left or right of the assumption sequence (it might be helpful to think of $/I$ as dequeuing and $\backslash I$ as popping from the assumptions in the premise). The division paradigm is of pedagogical utility only, and we will not take it any further for fear of (incorrectly) hinting at other properties of fractionals being applicable in the logic. A noteworthy change of notation appears in the elimination of the product: with $\Delta[\Gamma]$ we denote a structure Δ containing substructure Γ : $\Delta[_]$ now serves as a *context*, i.e. a structure of assumptions with a hole. The rule now claims it is acceptable to *replace* substructure A, B in Δ by Γ , if $\Gamma \vdash A \otimes B$ holds. The notions of structure and substructure depend of course on the logic used – in the current setting, Δ is a sequence, to which Γ is a subsequence. The reformulation of the rule is necessary to arbitrate elimination of nested products, since their extraction to the right or left edge of an assumption sequence is no longer possible. This also serves to better illustrate a remark made earlier: the rule can be applied at arbitrary nesting depths, each position corresponding to a supposedly different proof (consider for instance that if $\Delta[\Gamma]$, and $\Gamma[A, B]$, then it is also the case that $\Delta[A, B]$).

The Lambek calculus hails from an intuitionistic tradition, and is thus amenable to a propositions as types interpretation [Wansing, 1990]. Adorning its rules with faithful term rewrites translates into a type system that is both linear and ordered [Pierce, 2004]. Things get funky there: we now have two distinct modes of function application and λ abstraction, each pair with its

$$\begin{array}{c}
\frac{}{x_i : A \vdash x_i : A} \text{id} \\
\\
\frac{\Gamma \vdash s : B/A \quad \Delta \vdash t : A}{\Gamma, \Delta \vdash s \triangleleft t : B} /E \qquad \frac{\Gamma, x_i : A \vdash s : B}{\Gamma \vdash \lambda x_i. s : B/A} /I \\
\\
\frac{\Gamma \vdash s : A \quad \Delta \vdash t : A \setminus B}{\Gamma, \Delta \vdash s \triangleright t : B} \setminus E \qquad \frac{x_i : A, \Gamma \vdash s : B}{\Gamma \vdash \lambda x_i. s : A \setminus B} \setminus I \\
\\
\frac{\Gamma \vdash s : A \otimes B \quad \Delta \llbracket x_i : A, x_j : B \rrbracket \vdash t : C}{\Delta \llbracket \Gamma \rrbracket \vdash \text{case } s \text{ of } (x_i, x_j) \text{ in } t : C} \otimes E \qquad \frac{\Gamma \vdash s : A \quad \Delta \vdash t : B}{\Gamma, \Delta \vdash (s, t) : A \otimes B} \otimes I
\end{array}$$

Figure I.11: Lambek calculus **L**.

own reduction. We use \triangleleft and \triangleright to denote right and left application, respectively – the mnemonic is that the triangle points to the function – and λ and λ to denote the two kinds of anonymous functions.

3.1.1 Proof & Term Reductions

The proof reductions of Figure I.12 should be at this point straightforward to decode. The only addition is the symmetric version of the familiar implicational redex. For the redex of the product, the substituted A and B hypotheses are now wrapped on both sides by a context Θ , following the formulation of $\otimes E$. The corresponding term reductions are:

$$(\lambda x_i. s) \triangleleft t \xrightarrow{\beta} s_{[x_i \mapsto t]} \quad (\text{I.17})$$

$$t \triangleright (\lambda x_i. s) \xrightarrow{\beta} s_{[x_i \mapsto t]} \quad (\text{I.18})$$

$$\text{case } (s, t) \text{ of } (x_i, x_j) \text{ in } u \xrightarrow{\beta} u_{[s \mapsto x_i, t \mapsto x_j]} \quad (\text{I.19})$$

3.2 Dropping Associativity

Judging by the apparent absence of any more structural rules to remove, someone eager to be done with the whole story could at this point proclaim our substructural tour finished. We are not quite done yet, however, for one last structural equivalence still remains unchecked (one we have made extensive use of, for that matter). The culprit can be found by going back to our original definition of structures in the long and distant past of Subsection 1.1 – by

$$\begin{array}{c}
\frac{\frac{\frac{\overline{A \vdash A} \text{ id}}{\vdots s} \quad \frac{\Gamma, A \vdash B}{\Gamma \vdash B/A} /I \quad \frac{\frac{\vdots t}{\Delta \vdash A}}{\Gamma, \Delta \vdash B} /E}{\Gamma, \Delta \vdash B} \Rightarrow \frac{\frac{\vdots t}{\Delta \vdash A}}{\vdots s} \quad \frac{\Gamma, \Delta \vdash B}{}
\end{array}$$

$$\begin{array}{c}
\frac{\frac{\frac{\overline{A \vdash A} \text{ id}}{\vdots s} \quad \frac{\frac{\vdots t}{\Gamma \vdash A} \quad \frac{A, \Delta \vdash B}{\Delta \vdash A \setminus B} \setminus I}{\Gamma, \Delta \vdash B} \setminus E}{\Gamma, \Delta \vdash B} \Rightarrow \frac{\frac{\vdots t}{\Gamma \vdash A}}{\vdots s} \quad \frac{\Gamma, \Delta \vdash B}{}
\end{array}$$

$$\begin{array}{c}
\frac{\frac{\frac{\vdots s}{\Gamma \vdash A} \quad \frac{\vdots t}{\Delta \vdash B}}{\Gamma, \Delta \vdash A \otimes B} \otimes I \quad \frac{\frac{\overline{A \vdash A} \text{ id} \quad \overline{B \vdash B} \text{ id}}{\vdots u} \quad \frac{\Theta[A, B] \vdash C}{\Theta[\Gamma, \Delta] \vdash C} \otimes E}{\Theta[\Gamma, \Delta] \vdash C} \Rightarrow \frac{\frac{\vdots s}{\Gamma \vdash A} \quad \frac{\vdots t}{\Delta \vdash B}}{\vdots u} \quad \frac{\Theta[\Gamma, \Delta] \vdash C}{}
\end{array}$$

Figure I.12: Lambek β redexes.

treating them as sequences, we have mindlessly equipped them with *associativity* for free, the use of which we never made explicit. The one to notice was Lambek once more [Lambek, 1961]. In the new logic (pragmatically named the non-associative Lambek calculus **NL**) the definition of a structure changes to:

$$\Gamma, \Delta, \Theta := A \mid (\Gamma \cdot \Delta) \quad (\text{I.20})$$

i.e. the structural unit of the empty sequence is no more, and the scope of the binary structural binder is made explicit with brackets (we use the distinct symbol \cdot to tell this new structural binder apart from its associative sibling). On top of adjacency and order, the non-associative Lambek calculus further considers *constituency*; structures are now binary trees, with atomic propositions as their leaves and \cdot as branching nodes, and judgements are differentiated on the basis of the binary branching form their assumptions take. Formulas remain as they were, but the presentation of the rules changes to that of Figure I.13 in order to accommodate the new, stricter structures. Merging structures Γ and Δ via $\setminus E$, $/E$ or $\otimes I$ is translated to building up a tree with the two as branches. Decomposing a structure via an abstraction $\setminus I$ or $/I$ now requires that the formula abstracted over occurs not just at the edge of the tree's linear projection, but also at its top-most branching level. The notation $\Gamma[\Delta]$ now denotes that Δ is a *subtree* of Γ – for the product elimination $\otimes E$ to be applicable, A and B need not just be adjacent, but also commonly rooted.

The syntax of the isomorphic λ -calculus is identical to before, except this

$$\begin{array}{c}
\frac{}{x_i : A \vdash x_i : A} \text{id} \\
\\
\frac{\Gamma \vdash s : B/A \quad \Delta \vdash t : A}{(\Gamma \cdot \Delta) \vdash s \blacktriangleleft t : B} /E \qquad \frac{(\Gamma \cdot x_i : A) \vdash s : B}{\Gamma \vdash \lambda x_i. s : B/A} /I \\
\\
\frac{\Gamma \vdash s : A \quad \Delta \vdash t : A \setminus B}{(\Gamma \cdot \Delta) \vdash s \blacktriangleright t : B} \setminus E \qquad \frac{(x_i : A \cdot \Gamma) \vdash s : B}{\Gamma \vdash \setminus x_i. s : A \setminus B} \setminus I \\
\\
\frac{\Gamma \vdash s : A \otimes B \quad \Delta[(x_i : A, x_j : B)] \vdash t : C}{\Delta[\Gamma] \vdash \text{case } s \text{ of } (x_i, x_j) \text{ in } t : C} \otimes E \qquad \frac{\Gamma \vdash s : A \quad \Delta \vdash t : B}{\Gamma, \Delta \vdash (s, t) : A \otimes B} \otimes I
\end{array}$$

Figure I.13: Non-associative Lambek calculus **NL**.

time we use \blacktriangleleft and \blacktriangleright to notationally differentiate with the non-associative application (not unlike how we replaced the intuitionistic implication \rightarrow with its linear counterpart \multimap earlier). The new structural constraint on the introduction of a directed implication can be intuitively translated to a constraint on the applicability of an abstraction. Namely, the variable to abstract over needs to occur at the top-most level of a function application in the term's inductive body.¹

3.2.1 Proof & Term Reductions

Proof & term reductions are notationally identical to those of the previous subsection, modulo bracketing, and substituting white for black triangles. I trust the missing picture is easy enough to create mentally.

3.3 The Full Landscape

We have seen **NL** as a refinement of **L**, and **L**, in turn, as a refinement of **ILL**. The three can be perceived as points in a lattice of substructural logics, upon which we can move by adding or removing structural rules at a global level; this view lends **ILL** its alternative name **LP**, for the Lambek calculus with permutation (also encountered as the Lambek-van Benthem Calculus [van Benthem, 1988]). At the top of the diamond we have **ILL**, where (linearity aside), anything goes, and at the bottom we have **NL**, where neither associativity nor commutativity hold. At the center, there's **L**, where only associativity holds. Next to it, an unexpected curiosity pops up: **NLP** (for the non-associative Lam-

¹Abusing terminology, here by inductive body we mean the term itself (if it's an implicative one), the term's inner body (if it's a sequence of abstractions), the left or right coordinate (if it's a product), or the nested body on which substitution is performed (if it's a deconstructed product).

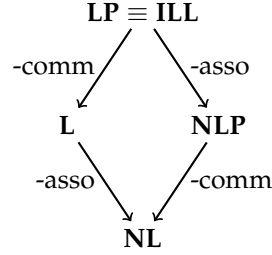


Figure I.14: (N)L(P): ILL and substructural friends.

bek with permutation), an offbeat logic where associativity holds but commutativity doesn't – its structures are *mobiles*: orderless, binary branching trees that make no distinction between left and right daughters. Unlike its relatives, **NLP** has received limited attention from theorists and practitioners alike. This will still remain the case even after (if?) this manuscript sees the light of day, but its peculiar structures will reemerge and have their moment to shine later on.

4 Restoring Control

With every step we have taken further into substructuraland, we have been paying a price in expressivity; it is now time for us to acknowledge the accumulated bill. Dropping *contr* and *weak* made us resource conscious, but theorems of **IL** that required resource duplication or erasure became undervivable. Dropping *ex* forced us to pay attention to the order of assumptions, but costed us access to theorems that required permutation to derive. Substituting the structural comma \multimap, \multimap with the non-associative $(\cdot \cdot)$ casted our sequences to trees, this time at the expense of theorems that required rebracketing. Woe is us – is there even anything left we can derive?

Perhaps this is painting an overly dramatic picture, considering that none of this is necessarily bad. From an epistemic perspective, the less structural equivalences we take for granted, the better our mental grasp of structural difference becomes. In the best case, if it just so happens that the kind of structures we want to investigate overlaps *fully* with the kind of structures our logic can explicitly reason about, the distinction between theorem and non-theorem becomes a refinement rather than a loss of expressivity. From a more pragmatic perspective, more structural constraints means easier proof search, and less theorems means faster exhaustion of possibilities. To make the scale of the combinatorics tangible, reflect for a second on this. A single judgement of n hypotheses in **NL** is but one of the Catalan number of bracketings $C(n)$ it would be syntactically undistinguishable from in **L**, each one of which in turn is but one of the factorially many permutations $n!$ it would be equivalent to

in LP.¹ In the case for checking the satisfiability of a judgement (i.e. searching for *any* valid proof), all the above would have made for potential proof candidates; in the case for attempting to enumerate the proofs of a judgement (i.e. searching for *all* valid proof), they would all have needed to be exhausted. The point to take home is that proof search becomes decidedly easier in the absence of syntactic equivalences, so perhaps a double-edged sword would have made for a better analogy than a bill.

The defeatist attitude here would be to just accept the trade-off between expressivity and complexity, weep for the theorems forever lost, take our victory and walk away. The problem lies however in the common occasion where the structure of objects under scrutiny overlaps only *partially* with a specific sub-structural flavour, modulo some exceptional but real cases that require added expressivity. In such a scenario, taking a step up in the hierarchy would cause an undesirable combinatorial explosion, whereas staying put would sacrifice our ability to argue about these exceptional cases. By contrast, the maximalist attitude makes no concessions and seeks both for the cake to be whole and the dog to be fed.² What if there was a way to keep our logic computationally tractable but with temporary and on-demand access to normally excluded reasoning tools?

4.1 The Logic of Modalities

The answer comes in the form of unary *modalities*, type-forming operators lent from modal logics, that allow navigation between logics of different structural properties. Unary modalities hold a key role in the presentation of full linear logic; there, a single operator ! (called *bang*) would allow an embedding of intuitionistic (non-linear) propositions into the linear regime, essentially acting as a licenser of contr and weak. In our case, we will make do with two modalities from temporal logic, the diamond \Diamond and the box \Box .

The two form a residuated pair, the properties of which can be formulated either (i) in the form of a type-level biconditional derivability relation:

$$\Diamond A \vdash B \text{ iff } A \vdash \Box B \quad (\text{I.21})$$

or (ii) the monotonic behavior of its parts:

$$A \vdash B \implies \Diamond A \vdash \Diamond B \quad (\text{I.22})$$

$$A \vdash B \implies \Box A \vdash \Box B \quad (\text{I.23})$$

and the adjointness of their compositions, where $\Diamond\Box(-)$ is an *interior* and $\Box\Diamond(-)$

¹Boom, goes the combinatorial explosion.

²Direct translation of a silly but fitting Greek aphorism that won by a small margin over the Italian equivalent (wine barrel full and wife drunk). In any case, cake is bad for dogs.

$$\begin{array}{c}
\frac{\Gamma \vdash s : \Box A}{\langle \Gamma \rangle \vdash \blacktriangledown s : A} \Box E \qquad \frac{\langle \Gamma \rangle \vdash s : A}{\Gamma \vdash \blacktriangle s : \Box A} \Box I \\
\\
\frac{\Gamma \llbracket \langle x_i : A \rangle \rrbracket \vdash s : B \quad \Delta \vdash t : \Diamond A}{\Gamma \llbracket \Delta \rrbracket \vdash \text{case } \forall t \text{ of } x_i \text{ in } s : B} \Diamond E \qquad \frac{\Gamma \vdash s : A}{\langle \Gamma \rangle \vdash \Delta s : \Diamond A} \Diamond I
\end{array}$$

Figure I.15: Logical rules of modal inference.

a *closure* operator:

$$\Gamma \vdash A \implies \Gamma \vdash \Box \Diamond A \quad (\text{I.24})$$

$$\Gamma \vdash \Diamond \Box A \implies \Gamma \vdash A \quad (\text{I.25})$$

The logical manipulation of these modalities is handled by corresponding elimination and introduction rules, presented in Figure I.15. The presentation is intentionally detached from a specific substructural strand – modalities are plug-and-play to any member of the **(N)L(P)** family. Their incorporation adds a new kind of structure to the ones provided by the underlying logic, altering judgements accordingly:

$$\Gamma, \Delta, \Theta := \dots \mid \langle \Gamma \rangle \quad (\text{I.26})$$

Angular brackets denote unary tree branches that behave slightly different to the rest; they act as an impenetrable barrier that permits or hinders the introduction or elimination of modal connectives in a judgement. The box elimination rule $\Box E$ grants us the option of removing a logical box from the succedent of the premise (as long as it is its main connective), but encloses the premises in angular brackets in the process. Its introduction counterpart $\Box I$ does the exact opposite: it frees a judgement's assumptions from their brackets, but puts the succedent proposition under the scope of a box. The diamond behaves just the other way around. Its introduction rule $\Diamond I$ is straightforward: it offers the possibility of putting the succedent under the scope of a diamond, in exchange wrapping the antecedents with brackets. The elimination rule $\Diamond E$ is more of a problem child, behaving akin to a unary product. Without locality restrictions, it inspects a proof of B , the assumptions of which contain a substructure $\langle A \rangle$ within context $\Gamma \llbracket _ \rrbracket$, and allows the post-hoc substitution of the hypothesis together with its brackets by a structure Δ , if from it one can derive $\Diamond A$.

Rules are adorned with term rewrite instructions in the propositions as types style, similar to how temporal logic can be operationalized in the λ -calculus [Wansing, 2002]. The mnemonic is now two-dimensional: upward triangles denote introduction and downward ones elimination, whereas black triangles are for the box, white ones for the diamond. Term constructions for the single-premise rules are uncomplicated: each type operation just leaves the corresponding term footprint. This is not the case for the $\Diamond E$ rule, which

$$\begin{array}{c}
\vdots s \\
\langle \Gamma \rangle \vdash A \\
\hline
\Gamma \vdash \Box A \quad \Box I \\
\hline
\langle \Gamma \rangle \vdash A \quad \Box E \\
\hline
\end{array}
\Rightarrow
\begin{array}{c}
\vdots s \\
\langle \Gamma \rangle \vdash A
\end{array}$$

$$\begin{array}{c}
\overline{A \vdash A} \text{ id} \quad \vdots t \\
\vdots s \quad \Delta \vdash A \\
\Gamma \llbracket \langle A \rangle \rrbracket \vdash B \quad \langle \Delta \rangle \vdash \Diamond A \\
\hline
\Gamma \llbracket \langle \Delta \rangle \rrbracket \vdash B \quad \Diamond I \\
\hline
\Gamma \llbracket \langle \Delta \rangle \rrbracket \vdash B \quad \Diamond E
\end{array}
\Rightarrow
\begin{array}{c}
\vdots t \\
\Delta \vdash A \\
\vdots s \\
\Gamma \llbracket \langle \Delta \rangle \rrbracket \vdash B
\end{array}$$

(a) Modal β redexes.

$$\begin{array}{c}
\vdots s \\
\Gamma \vdash \Box A \\
\hline
\langle \Gamma \rangle \vdash A \quad \Box E \\
\hline
\Gamma \vdash \Box A \quad \Box I \\
\hline
\end{array}
\equiv
\begin{array}{c}
\vdots s \\
\Gamma \vdash \Box A
\end{array}$$

$$\begin{array}{c}
\overline{A \vdash A} \text{ id} \quad \vdots s \\
\langle A \rangle \vdash \Diamond A \quad \Delta \vdash \Diamond A \\
\hline
\Delta \vdash \Diamond A \quad \Diamond I \\
\hline
\Delta \vdash \Diamond A \quad \Diamond E
\end{array}
\equiv
\begin{array}{c}
\vdots s \\
\Delta \vdash \Diamond A
\end{array}$$

(b) Modal η redexes.

Figure I.16: Modal redexes

requires some attention: the structural substitution of $\langle A \rangle$ for Δ necessitates a case construct that calls for a term substitution of the variable x_i for ∇t . Note that the free variables of the resulting expression (case ∇t of x_i in s) are the union of the free variables of t and those of s except for x_i , which becomes *bound* by the case construct.

4.1.1 Proof & Term Reductions

The proof patterns of Figure I.16a exhibit introduction elimination chains of modal operators, and thus constitute β redexes subject to normalization. The first one is trivial: it just says that a sequential application of $\Box I$ followed by $\Box E$ can be safely excised. The second one proposes that if a $\Diamond I$ is the last rule to have been applied on the substitution branch t of the $\Diamond E$ rule, it would make sense to simply plug proof t in place of the proposition A hypothesized in the

other branch s . On the term level, these rules correspond to computations:

$$\blacktriangledown \blacktriangle s \xrightarrow{\beta} s \quad (I.27)$$

$$\text{case } \nabla \Delta t \text{ of } x \text{ in } s \xrightarrow{\beta} s_{[x \mapsto t]} \quad (I.28)$$

The dual direction of η equivalences also holds – since these are discovered in the literature less frequently than the more pedestrian implication and product equivalences, we explicitly present them in Figure I.16b. The term equivalences they materialize are:

$$\blacktriangle \blacktriangledown s \stackrel{\eta}{=} s \quad (I.29)$$

$$\text{case } \nabla s \text{ of } x \text{ in } \Delta x \stackrel{\eta}{=} s \quad (I.30)$$

4.1.2 A Digression on Modal Terms

For the modally savvy, the term rewrites attributed to the modal rules might seem unorthodox. A more common presentation employs the simpler meta-syntax notation of term substitution. For instance, $\Diamond E$ can often be spotted in the wild as:

$$\frac{\Gamma[\langle x_i : A \rangle] \vdash s : B \quad \Delta \vdash t : \Diamond A}{\Gamma[\Delta] \vdash s_{[x_i \mapsto \nabla t]} : B} \Diamond E$$

In this disguise, the rule is again seen as realizing a retroactive substitution of x_i with ∇t , except this time around the substitution is *actually* performed, resulting in less cumbersome terms being carried around.

Opting for this alternative notation has, however, a number of negative consequences. The more superficial one is that the main term connective does not take scope at the outermost layer of rule's yield, but rather nested arbitrarily deeply within it, unlike its better behaved version. From a proof-theoretic perspective, normalization is now baked directly into the theory, as the term yield of the rule exactly coincides with its β reduced form. At the same time, all rule permutations boil down to having the exact same reduction, i.e. multiple previously distinct terms are conflated into a single representation. This establishes an implicit syntactic equivalence on proofs that claims that the exact position of the $\Diamond E$ rule is *syntactically irrelevant* (so long of course as the same variable x_i is substituted by the same term ∇t). Finally, the shorthand version hides variables; hypotheses that would be bound by the case construct are instead erased and forgotten, obfuscating the term-to-proof correspondence. All these are perhaps minor points not worth taking too seriously, but for one concerned with concrete implementation the extra merit of notational simplicity comes at the cost of equality checking become way more tedious. With this in mind (and in a rare moment of excessive formal zeal), we will exercise some self restraint and avoid indulging in the convenience of this version.

4.1.3 Properties

Situating our unary operators within the modal logic zoo is no trivial endeavour (especially if you, like me, have had limited exposure to it before). They are best characterized by the properties they satisfy, so inspecting them should shed some light on their proof-theoretic behavior (as a bonus, it will also help us get better acquainted with the kind of term rewrites their rules prescribe). Figure I.17 presents the proof transformations equivalent to the properties foretold: (a) and (b) for monotonicity, (c) and (d) for composition, and (e) and (f) for the two directions of the residuation law.

Worth a special mention are also the so-called triple laws:

$$\Diamond A \dashv\vdash \Diamond \Box \Diamond A \quad (\text{I.31})$$

$$\Box A \dashv\vdash \Box \Diamond \Box A \quad (\text{I.32})$$

which can be intuitively read as claiming that prepending an already modal type with (one or more) diamond-box pairs in alteration has no real effect, as these can unconditionally cancel out or be expanded into. Figure I.18 presents proofs of the above in both directions.

4.2 Structural Reasoning

This detour may have proven lengthy, but has hopefully helped us acquire a first taste for modalities. We now know how to introduce and eliminate them and what the effect of doing so is on the antecedent structure, and got a first glimpse of their properties, the term rewrites they prescribe and the type inequalities (in the form of unidirectional derivations) they give rise to. The question then becomes how to actually use them for the task at hand, namely disciplined traversal between substructural logics. Structural reasoning is accomplished via structural postulates, rules of inference that enact commutativity and associativity (or combinations thereof), except in a controlled fashion. These are permissible only under strict conditions on the substructures constituent to the antecedent structure – this is exactly where the new kind of structures will prove useful. There is no fixed vocabulary of structural rules, as they are intended for application-specific finetuning of a universal logical core, so we are free to design and populate it according to our own needs. Prime examples and standard items for consideration include the controlled associativity and mixed associativity-commutativity rules of Figure I.19a (and the corresponding tree transformations of Figure I.19b, if you have a disdain for brackets). The first rule ass_\Diamond allows a unary branch $\langle \Phi \rangle$ to escape its bind to its neighbour Θ , forcing it to associate to the structure Δ to its left instead. The second one mix_\Diamond allows a unary $\langle \Theta \rangle$ to swap position with its right-adjacent neighbor Φ , disassociating from its left neighbour Δ in the process. In domains where even finer control is needed, one can consider indexed families of (interacting) modalities, each with their own structural brackets and rulesets.

$$\frac{\frac{\vdots s}{x_i : A \vdash s : B} \Diamond I \quad \frac{}{x_j : \Diamond A \vdash x_j : \Diamond A} \text{id}}{x_j : \Diamond A \vdash \text{case } \nabla x_j \text{ of } x_i \text{ in } s : \Diamond B} \Diamond E$$

(a) Monotonicity of the diamond.

$$\frac{\frac{\vdots s}{x_i : A \vdash s : B} \multimap I \quad \frac{}{x_j : \Box A \vdash x_j : \Box A} \text{id}}{\frac{\langle x_j : \Box A \rangle \vdash (\lambda x_i. s) \blacktriangledown x_j : B}{x_j : \Box A \vdash \blacktriangle(\lambda x_i. s) \blacktriangledown x_j : \Box B} \Box I} \multimap E$$

(b) Monotonicity of the box.

$$\frac{\frac{\vdots s}{\Gamma \vdash s : A} \Diamond I \quad \frac{}{\langle \Gamma \rangle \vdash \Delta s : \Diamond A} \Box I}{\Gamma \vdash \blacktriangle \Delta s : \Box \Diamond A} \Box I \quad \frac{\frac{}{x_i : \Box A \vdash x_i : \Box A} \text{id} \quad \frac{}{\langle x_i : \Box A \rangle \vdash \blacktriangledown x_i : A} \Box E \quad \frac{\vdots s}{\Gamma \vdash s : \Diamond \Box A} \Diamond E}{\Gamma \vdash \text{case } \nabla s \text{ of } x_i \text{ in } \blacktriangledown x_i : A} \Diamond E$$

(c) The closure $\Diamond\Box(\cdot)$.(d) And the interior $\Box\Diamond(\cdot)$.

$$\frac{\frac{\vdots s}{x_i : A \vdash s : \Box B} \Box E \quad \frac{}{\langle x_i : A \rangle \vdash \blacktriangledown s : B} \Box E}{x_j : \Diamond A \vdash \text{case } \nabla x_j \text{ of } x_i \text{ in } s : B} \Diamond E$$

(e) Residuation law: from $A \vdash \Box B$ to $\Diamond A \vdash B$.

$$\frac{\frac{\vdots s}{x_i : \Diamond A \vdash s : B} \multimap I \quad \frac{}{x_j : A \vdash x_j : A} \text{id}}{\frac{\langle x_j : A \rangle \vdash \Delta x_j : \Diamond A}{\langle x_j : A \rangle \vdash \text{case } \nabla \Delta x_j \text{ of } x_i \text{ in } \lambda x_i. s : B} \Diamond E} \multimap E$$

(f) Ditto, the other way around.

Figure I.17: Derivations for the various aspects of residuation.

$$\frac{\frac{\overline{x_i : \Box A \vdash x_i : \Box A} \text{ id}}{\langle x_i : \Box A \rangle \vdash \nabla x_i : A} \Box E \quad \frac{\frac{\overline{x_j : \Box \Diamond \Box A \vdash x_j : \Box \Diamond \Box A} \text{ id}}{\langle x_j : \Box \Diamond \Box A \rangle \vdash \nabla x_j : \Diamond \Box A} \Box E}{\frac{\langle x_j : \Box \Diamond \Box A \rangle \vdash \text{case } \nabla x_j \text{ of } x_i \text{ in } \nabla x_i : A}{x_j : \Box \Diamond \Box A \vdash \blacktriangle(\text{case } \nabla x_j \text{ of } x_i \text{ in } \nabla x_i) : \Box A} \Box I} \Diamond E$$

(a) Contraction of $\Box \Diamond \Box (-)$ to $\Box (-)$.

$$\frac{\frac{\overline{x_i : \Box A \vdash x_i : \Box A} \text{ id}}{\langle x_i : \Box A \rangle \vdash \Delta x_i : \Diamond \Box A} \Diamond I}{x_i : \Box A \vdash \blacktriangle \Delta x_i : \Box \Diamond \Box A} \Box I$$

(b) Expansion of $\Box (-)$ to $\Box \Diamond \Box (-)$.

$$\frac{\frac{\overline{x_i : \Box \Diamond A \vdash x_i : \Box \Diamond A} \text{ id}}{\langle x_i : \Box \Diamond A \rangle \vdash \nabla x_i : \Diamond A} \Box E \quad \frac{\overline{x_j : \Diamond \Box \Diamond A \vdash x_j : \Diamond \Box \Diamond A} \text{ id}}{\langle x_j : \Diamond \Box \Diamond A \rangle \vdash \nabla x_j : \Diamond A} \Diamond E}{x_j : \Diamond \Box \Diamond A \vdash \text{case } \nabla x_j \text{ of } x_i \text{ in } \nabla x_i : \Diamond A} \Diamond E$$

(c) Contraction of $\Diamond \Box \Diamond (-)$ to $\Diamond (-)$.

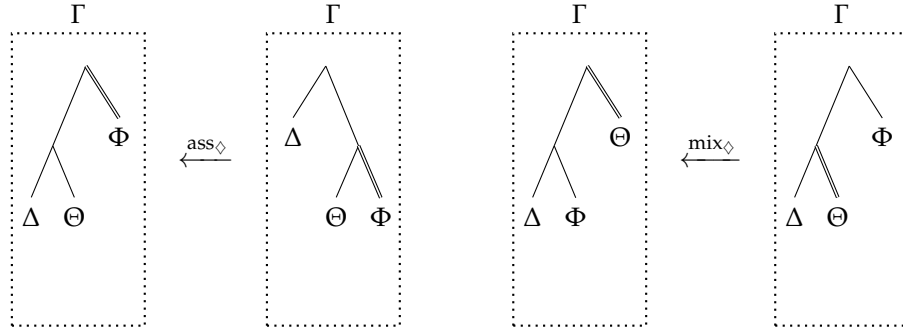
$$\frac{\frac{\overline{x_i : A \vdash \blacktriangle \Delta x_i : \Box \Diamond A} \text{ (I.17c)}}{\langle x_i : A \rangle \vdash \Delta \blacktriangle \Delta x_i : \Diamond \Box \Diamond A} \Diamond I \quad \frac{\overline{x_j : \Diamond A \vdash x_j : \Diamond A} \text{ id}}{\langle x_j : \Diamond A \rangle \vdash \text{case } \nabla x_j \text{ of } x_i \text{ in } \Delta \blacktriangle \Delta x_i : \Diamond \Box \Diamond A} \Diamond E}{x_j : \Diamond A \vdash \text{case } \nabla x_j \text{ of } x_i \text{ in } \Delta \blacktriangle \Delta x_i : \Diamond \Box \Diamond A} \Diamond E$$

(d) Expansion of $\Diamond (-)$ to $\Diamond \Box \Diamond (-)$.

Figure I.18: The triple laws for the two modalities in both directions.

$$\frac{\Gamma[\Delta, (\Theta, \langle \Phi \rangle)] \vdash A}{\Gamma[(\Delta, \Theta), \langle \Phi \rangle] \vdash A} \text{ass}_{\Diamond} \quad \frac{\Gamma[(\Delta, \langle \Theta \rangle), \Phi] \vdash A}{\Gamma[(\Delta, \Phi), \langle \Theta \rangle] \vdash A} \text{mix}_{\Diamond}$$

(a) In rule format.



(b) Corresponding tree transformations. Double edges denote bracketed substructures.

Figure I.19: Controlled associativity/mixed commutativity.

5 The Linguistic Perspective

Despite their presentation having intentionally been left vague and abstract, the ideas explored so far have been a keystone element of computer science, from its inception until recent modernity. Beyond that, they form the common theoretical underpinnings for the formal treatment of natural languages and their various aspects, where they manifest as so-called *Categorical Grammars*. Categorical grammars is a heavily overloaded term that refers to a wide and diverse family of related formalisms, each with its own ambitions, goals, strengths and weaknesses. The most encompassing way of defining a categorical grammar is thus best accomplished through a high-level intersection of their common points. A categorical grammar is usually tied to a logic, commonly a choice from the ones reviewed so far (or at least loosely inspired by one). The choice of logic is part personal preference, but is usually motivated by the degree of alignment between the options under consideration and the characteristics of the target language – a factor that also comes into play is also the trade-off between expressivity and complexity. On the basis of the chosen logic, a categorical grammar has a *lexicon*; a mapping from primitive linguistic entries (i.e. words) to formulas of that logic. Their dependence on a lexicon grants categorical grammars their *strongly lexicalized* title – as the slogan goes, words carry their combinatorics on their sleeves. With these two components in hand, compiling composite structures for complex linguistic entries (i.e. parsing) becomes a process of formal deduction dictated by the interplay between the types of the participating atomic elements, and the rules of in-

Logic	Computer Science	Linguistics
Propositional Constant	Base Type	Syntactic Category
Inference Rule	Term Rewrite	Phrase Formation
Axiom	Variable	Word (or Empty Category)
Provability	Type Inhabitation	Grammaticality
Deduction	Program Synthesis	Parsing

Table I.3: The Curry-Howard correspondence applied in linguistics.

ference the logic is equipped with. Categorical grammars are a staple of the linguistic tradition and a focal point for practitioners, logicians and linguists alike. In this section we will examine some of their main strands, with a special emphasis on two spiritual progenitors of the unique flavour that is to be developed and presented later in this thesis.

5.1 Type-Logical Grammars

The earliest take at a categorical grammar are the AB grammars attributed to Kazimierz Ajdukiewicz [Ajdukiewicz, 1935] and Yehoshua Bar-Hillel [Bar-Hillel, 1953], but it was Jim Lambek that raised the existing notation and operations into the glory of a fully-fledged type theory. In their original purpose as envisaged by Lambek, his calculi would find use as *grammar logics*, i.e. universal systems of *grammatical* computation – a perspective adopted and advanced into what has presently come to be known as type-logical grammars [Morrill, 1994; Moortgat, 1997, 2014]. In a natural language setting, the linear base of the Lambek calculi is naturally equated to the resource sensitivity of grammar: words play a single grammatical role in the phrases they help form – there’s no ignoring or reusing items at will. There, the original Lambek calculus **L** would be the logic of *strings*; it can faithfully portray the generation of natural language utterances, where arbitrary reordering is a destructive process that ruins coherence. Its stricter version **NL** would instead be the logic of *constituency trees*; on top of word order, it further specifies constituency structure, allowing a distinction between different syntactic analyses of the same surface form. Type-logical grammars extend the Curry-Howard correspondence with a new axis, that of natural language; the transference of points of interest across that axis is presented in Table I.3; our motto shall from now on be “*Parsing as Deduction*”.

To see this in action, let’s consider first an instantiation of a Lambek calculus **NL** with the set of primitive types Prop_0 populated with signs characterizing the grammatical role of a piece of text that can independently stand on its own (i.e. phrasal categories or, more crudely, parts of speech). In a toy fragment and for illustrative purposes, this could look like:

$$\text{Prop}_0 := \{N, NP, S, PP\}$$

eye	::	N	
oceans, suns, deeps , dolphins			
sea-nymphs, whirlpools	::	NP	
the	::	NP/N	
opiate, strange, unrememberable, their	::	NP/NP	
poured	::	ITV	:= NP\S
behold	::	TV	:= (NP\S)/NP
there	::	ADV\	:= (NP\S)\(NP\S)
never	::	ADV/	:= (NP\S)/(NP\S)
litten	::	(NP\NP)/PP	
may	::	AUX	:= (NP\S)/(NP\S)

Table I.4: Toy lovecraftian lexicon of pure Lambek types.

for a grammar able to reason about nouns N , noun phrases and bare nouns NP , sentential clauses S , prepositional phrases PP and functions thereof in English. One might wonder: what happened to the remaining kinds of phrasal categories like verbs, adjectives and adverbs? These would indicate grammatical functions, and in fact should be represented as such. An intransitive phrase, for instance, is a grammatical function that would consume a left-adjacent noun phrase to produce a sentence, therefore it would materialize as $NP\S$. It follows that a transitive phrase or copula would then be of type $(NP\S)/NP$, a function that requires a right-adjacent noun phrase to produce an intransitive phrase, whereas a bitransitive, requiring two, would be $((NP\S)/NP)/NP$, etc. In the same vein, determiner phrases consume right-adjacent nouns and lift them to noun phrases NP/N , whereas prenominal adjectives are noun phrase (or noun) endomorphisms modifying them but keeping their type intact, NP/NP (and the other way around for postnominal use). Adverbs would also be endomorphisms, except this time higher-order – $(NP/NP)/(NP/NP)$ for adjectival and $(NP\S)\(NP\S)$ for verbal modification, respectively.

Linguistic reasoning is not done *ex nihilo* – formulas like the above are supplied by and grounded in the lexicon. This does not exclude the option of utilizing hypotheticals instantiated by the axiom rule *id* – hypothetical reasoning lives, in fact, at the core of the type-logical inferential process, as we will soon see. It means, rather, that our building blocks will for the most part be *lexical constants*, proof objects that behave just like variables, except they are neither wantonly typed nor amenable to abstraction. To convey the difference between the two, we will instantiate the latter with a seemingly new rule of inference, *lex*, which simply performs lexical lookup, i.e. pulls a word's type from the lexicon.

The internet guide *how to write a dissertation* I am consulting insists it is important to set clear goals and stick to them. It seems like sound advice, so

$$\frac{\frac{}{\text{strange} \vdash \text{NP}/\text{NP}} \text{lex} \quad \frac{}{\text{dolphins} \vdash \text{NP}} \text{lex}}{\text{strange} \cdot \text{dolphins} \vdash \text{NP}} /E$$

(a) Derivation for “*strange dolphins*”.

$$\frac{\frac{}{\text{the} \vdash \text{NP}/\text{N}} \text{lex} \quad \frac{}{\text{eye} \vdash \text{N}} \text{lex}}{\text{the} \cdot \text{eye} \vdash \text{NP}} /E$$

(b) Derivation for “*the eye*”.

$$\frac{\frac{}{\text{litten} : (\text{NP} \backslash \text{NP})/\text{PP}} \text{lex} \quad \frac{\frac{}{\text{by} : \text{PP}/\text{NP}} \text{lex} \quad \frac{}{\text{suns} : \text{NP}} \text{lex}}{\text{by} \cdot \text{suns} \vdash \text{PP}} /E}{\text{litten} \cdot (\text{by} \cdot \text{suns}) \vdash \text{NP} \backslash \text{NP}} /E$$

(c) Derivation for “*litten by suns*”.

$$\frac{\frac{}{\text{sea-nymphs} : \text{NP}} \text{lex} \quad \frac{\frac{}{\text{of} : (\text{NP} \backslash \text{NP})/\text{NP}} \text{lex} \quad \frac{\frac{}{\text{unrememberable} : \text{NP}/\text{NP}} \text{lex} \quad \frac{}{\text{deeps} : \text{NP}} \text{lex}}{\text{unrememberable} \cdot \text{deeps} \vdash \text{NP}} /E}{\text{of} \cdot (\text{unrememberable} \cdot \text{deeps}) \vdash \text{NP} \backslash \text{NP}} /E}{\text{sea-nymphs} \cdot (\text{of} \cdot (\text{unrememberable} \cdot \text{deeps})) \vdash \text{NP}} \backslash E$$

(d) Derivation for “*sea-nymphs of unrememberable deeps*”.

$$\frac{\frac{}{\text{opiate} : \text{NP}/\text{NP}} \text{lex} \quad \frac{}{\text{oceans} : \text{NP}} \text{lex}}{\text{opiate} \cdot \text{oceans} \vdash \text{NP}} /E \quad \frac{\frac{}{\text{poured} : \text{NP} \backslash \text{S}} \text{lex} \quad \frac{}{\text{there} : \text{ADV} \backslash} \text{lex}}{\text{poured} \cdot \text{there} \vdash \text{NP} \backslash \text{S}} \backslash E$$

$$\frac{}{(\text{opiate} \cdot \text{oceans}) \cdot (\text{poured} \cdot \text{there}) \vdash \text{S}} \backslash E$$

(e) Derivation for “*Opiate oceans poured there*”.

Figure I.20: Deriving simple multiplicative phrases in **NL**.

we are going to do just that, and attempt to demonstrate the analysis of a non-contrived example in the type-logical framework. The following looks like a fitting match:

“Opiate oceans poured there, litten by suns that the eye may never behold, and having in their whirlpools strange dolphins and sea-nymphs of unrememberable deeps.”

H.P. Lovecraft, *Azathoth* (1938). In *Leaves* (2).

Let’s pave the way towards this ambitious goal with the miniature mock-up lexicon of Table I.4, and see just how far it can get us.

Figure I.20 presents derivations for parts of the goal phrase, and our very first linguistic examples (!) – their purely applicative nature should make them straightforward to decipher. The two proofs of I.20e and I.20c can readily be combined to yield a derivation for the phrase “*Opiate oceans litten by suns poured there*”. Close, but not quite there... The participial “*litten*”, which acts here as a postnominal modifier, has the special property of being able to position itself either immediately after the noun phrase “*opiate oceans*” it modifies, or deferred until after the matrix head “*poured*” has made an appearance (with any adverbials attached to it). Attempting to produce a derivation for the original version seems like a dead-end enterprise, though. We are not to blame for this incompetence: the problem lies with the grammar – we could never hope to capture this behavior with our current machinery. Despite their elegance and formal appeal, grammars relying purely on Lambek calculi suffer from an aversion to anomalies like discontinuities and long-distance dependencies, which natural languages tend to exhibit at an unfortunately striking degree.

One could of course attempt to cop out of the problem by just introducing ad-hoc raised forms for movable parts, one per distinct position they can be found at. The repercussions of such a move would soon, however, prove catastrophic. On the one hand, the once reliably concise lexicon would become overpopulated by endless variations on the same theme: each expansion point of a lexical type would percolate into all other lexical items it interacts with (either as consumers or producers thereof), the effect cascading at progressively larger lexical neighborhoods, until (if ever) an eventual equilibrium is reached. On the other hand, raised types obfuscate the functional relations and constituency structures we have worked so hard to reveal and incorporate, virtually beating the very purpose of the logic. Relaxing the structural constraints of the logic to globally allow movement and/or rebracketing is no good either. Spurious ambiguity would be the least of our concerns as we would be faced with *overgeneration*, i.e. the unwelcome ability to derive proofs that have no correspondence to correct linguistic structures whatsoever, leading us back to square zero. If you have not skipped any parts yet, your reward should now manifest as an unwavering faith for a solution, and a premonition of what is to come: modalities to the rescue!

5.1.1 The Role of Modalities

Ever since their original integration with the vanilla multiplicative toolkit, (the early pioneer being none other than my #1 supervisor!) modalities have played an indispensable role in the history and development of type-logical grammars [Hendriks, 1995; Moortgat, 1996; Kurtonina and Moortgat; Moortgat, 1997; Vermaat, 1999]. They find use as either licensors or licensees of structural rewrites, now in the form of movement and rebracketing of words and phrases. Figure I.22 progresses our agenda by accounting for the presence of a (hypothetical) movable postnominal modifier via the rules of Figure I.19. To make the hypothesis movable, we need to instantiate it as a box – for the pure function contained therein to be applicable, the box needs to be removed, enclosing the hypothesis in angular brackets, which in turn license its structural extraction to the rightmost edge of the assumptions via the mix_\diamond rule. At that point, we need to eliminate the bracketed variable with a term of the corresponding type, plus a diamond. For this to work, we need to make the tiniest of modifications to our lexicon so as to get access to the sought-after diamond:

$$\text{litten} \quad :: \quad \diamond \square (\text{NP} \backslash \text{NP}) / \text{PP} \quad (\text{I.33})$$

Intuitively, the new type requests a prepositional phrase complement to the right, after the consumption of which it produces a movable postnominal modifier that can penetrate constituent phrase boundaries to the left. Equipped with it, we can derive both the local versions hinted at earlier, and their discontinuous variations; see Figure I.21 for a proof of concept.¹

This methodology is in fact adopted from Moortgat [1999], where it finds similar use in dealing with the grammatical ambivalence of relativizers like “that” or “which”. Bound relative clauses headed by complementizers like the above contain a subordinate sentence with a *gap*, which can vary in its position. Let’s make things unnecessarily convoluted for the sake of clichéd self-referentialism by considering the relative clause “*which can vary in its position*” of the previous sentence. There, the subordinate clause “*_ can varj in its position*” contains a gap in the subject position, which the head “*a gap*” occupies implicitly. This is not the case in the last relative clause “*which the head “gap” occupies implicitly*”, whose subordinate clause “*the head “gap” occupies _ implicitly*” contains a non-peripheral (nested) gap in direct object position. What a mess! The subject-relative case can easily be dealt with in a pure Lambek grammar, as the gap hypothesis occurs adjacent to the verb phrase, but the same cannot be said for the object-relative case, whose structurally free gap seems to pose a challenge. The solution comes in the form of two distinct type assignments for the relativizer, one per grammatical role fulfilled:

$$\text{that} :: \text{REL}_s := (\text{NP} \backslash \text{NP}) / (\text{NP} \backslash \text{S}) \quad (\text{I.34})$$

$$\text{that} :: \text{REL}_o := (\text{NP} \backslash \text{NP}) / (\text{S} / \diamond \square \text{NP}) \quad (\text{I.35})$$

¹Get it? It’s an actual *proof*.

$$\begin{array}{c}
\vdots \quad \frac{\overline{x_i : \square(NP \backslash NP)} \quad \text{id}}{\langle x_i \rangle \vdash NP \backslash NP} \square E \\
\frac{\text{opiate} \cdot \text{oceans} \vdash NP \quad \langle x_i \rangle \vdash NP \backslash NP}{(\text{opiate} \cdot \text{oceans}) \cdot \langle x_i \rangle \vdash NP} \backslash E \\
\frac{\text{poured} \cdot \text{there} \vdash NP \backslash S \quad \vdots}{((\text{opiate} \cdot \text{oceans}) \cdot \langle x_i \rangle) \cdot (\text{poured} \cdot \text{there}) \vdash S} \backslash E \\
\frac{((\text{opiate} \cdot \text{oceans}) \cdot \langle x_i \rangle) \cdot (\text{poured} \cdot \text{there}) \vdash S}{((\text{opiate} \cdot \text{oceans}) \cdot (\text{poured} \cdot \text{there})) \cdot \langle x_i \rangle \vdash S} \text{mix}_\diamond
\end{array}$$

(a) Extracting a hypothetical postnominal modifier...

$$\begin{array}{c}
\text{(I.21a)} \quad \frac{\overline{\text{litten} : \diamond \square(NP \backslash NP) / PP} \quad \text{lex} \quad \frac{\overline{\text{by} : PP / NP} \quad \text{lex} \quad \overline{\text{suns} : NP} \quad \text{lex}}{\text{by} \cdot \text{suns} \vdash PP} / E}{\text{litten} \cdot (\text{by} \cdot \text{suns}) \vdash \diamond \square(NP \backslash NP)} / E \\
\frac{((\dots) \cdot (\dots)) \cdot \langle x_i \rangle \vdash S \quad \text{litten} \cdot (\text{by} \cdot \text{suns}) \vdash \diamond \square(NP \backslash NP)}{((\text{opiate} \cdot \text{oceans}) \cdot (\text{poured} \cdot \text{there})) \cdot (\text{litten} \cdot (\text{by} \cdot \text{suns})) \vdash S} \diamond E
\end{array}$$

(b) ...before substituting the hypothesis for its material instance.

Figure I.21: Deriving long-distance postnominal modification with the aid of type assignment (I.33).

The second version launches a mobile NP hypothesis via the same diamond-box pattern showcased earlier. The proof of Figure I.22 employs this typing in combination with the ass_\diamond rule to derive the object-relative clause “*that the eye may never behold*”, which applied to “*suns*” and combined with the proof of Figure I.21 yields the correct form of the postnominal modifier “*opiate oceans poured there, litten by suns that the eye may never behold*”, bringing us one step closer to success.

5.1.2 Intricacies of the Lexicon

The analysis just performed illustrated the necessity of (at least) two distinct types for the same string “*that*”, hinting at the fact that the lexicon is *not a function* from words to types, but rather a *relation* between them. One, more opinionated than I, might argue that each type is mapped to a distinct lexical item (one per relativization type), and that the identification between their strings is a mere coincidence, an idiosyncrasy of the language, or anyway irrelevant; even if a string is multi-typed, each type is a witness to a unique latent word hiding behind it. Of different effect but similar flavour would be the line of defense that appeals to null syntax, a covert process that can conditionally nominalize infinitives, determine plural nouns, relativize gerunds or do any sort of thing, really; a word is never multi-typed, but ad-hoc type conversions can take place out of the blue. Even under premises as radical as the above, occasions of type undeterminism are all but rare. Consider for instance the verb “*to have*”, whose argument structure for the possessive meaning alone) is specified (according to its FrameNet entry [Baker et al., 1998]) as

$$\begin{array}{c}
\text{(I.20b)} \quad \frac{\text{the} \cdot \text{eye} \vdash \text{NP} \quad \frac{\text{may} : \text{AUX} \quad \text{lex} \quad \frac{\text{never} : \text{ADV}_/ \quad \text{lex} \quad \frac{\text{behold} : \text{TV} \quad \text{lex} \quad \frac{\overline{x_i : \square \text{NP}} \quad \text{id}}{\langle x_i \rangle \vdash \text{NP}} \quad \square E}{\text{behold} \cdot \langle x_i \rangle \vdash \text{NP} \backslash \text{S}} \quad /E}{\text{never} \cdot (\text{behold} \cdot \langle x_i \rangle) \vdash \text{NP} \backslash \text{S}} \quad /E}{\text{may} \cdot (\text{never} \cdot (\text{behold} \cdot \langle x_i \rangle)) \vdash \text{NP} \backslash \text{S}} \quad /E}{\text{the} \cdot \text{eye} \cdot (\text{may} \cdot (\text{never} \cdot (\text{behold} \cdot \langle x_i \rangle))) \vdash \text{S}} \quad \backslash E \\
\frac{\text{the} \cdot \text{eye} \cdot (\text{may} \cdot (\text{never} \cdot (\text{behold} \cdot \langle x_i \rangle))) \vdash \text{S} \quad \text{ass}_\diamond}{\text{the} \cdot \text{eye} \cdot (\text{may} \cdot ((\text{never} \cdot \text{behold}) \cdot \langle x_i \rangle)) \vdash \text{S}} \quad \text{ass}_\diamond \\
\frac{\text{the} \cdot \text{eye} \cdot (\text{may} \cdot ((\text{never} \cdot \text{behold}) \cdot \langle x_i \rangle)) \vdash \text{S} \quad \text{ass}_\diamond}{\text{the} \cdot \text{eye} \cdot ((\text{may} \cdot (\text{never} \cdot \text{behold})) \cdot \langle x_i \rangle) \vdash \text{S}} \quad \text{ass}_\diamond \\
\frac{\text{the} \cdot \text{eye} \cdot ((\text{may} \cdot (\text{never} \cdot \text{behold})) \cdot \langle x_i \rangle) \vdash \text{S} \quad \text{ass}_\diamond}{((\text{the} \cdot \text{eye}) \cdot (\text{may} \cdot (\text{never} \cdot \text{behold}))) \cdot \langle x_i \rangle \vdash \text{S}} \quad \text{ass}_\diamond \\
\frac{((\text{the} \cdot \text{eye}) \cdot (\text{may} \cdot (\text{never} \cdot \text{behold}))) \cdot \langle x_i \rangle \vdash \text{S} \quad \overline{x_j : \diamond \square \text{NP}} \quad \text{id}}{((\text{the} \cdot \text{eye}) \cdot (\text{may} \cdot (\text{never} \cdot \text{behold}))) \cdot x_j \vdash \text{S}} \quad \diamond E \\
\frac{\text{that} : \text{REL}_o \quad \text{lex} \quad \frac{((\text{the} \cdot \text{eye}) \cdot (\text{may} \cdot (\text{never} \cdot \text{behold}))) \cdot x_j \vdash \text{S}}{(\text{the} \cdot \text{eye}) \cdot (\text{may} \cdot (\text{never} \cdot \text{behold})) \vdash \text{S} / \diamond \square \text{NP}} \quad /I}{\gamma := \text{that} \cdot ((\text{the} \cdot \text{eye}) \cdot (\text{may} \cdot (\text{never} \cdot \text{behold}))) \vdash \text{NP} \backslash \text{NP}} \quad /E
\end{array}$$

(a) Deriving an object-relative clause...

$$\begin{array}{c}
\text{(I.22a)} \quad \frac{\text{litten} : \diamond \square (\text{NP} \backslash \text{NP}) / \text{PP} \quad \text{lex} \quad \frac{\text{by} : \text{PP} / \text{NP} \quad \text{lex} \quad \frac{\text{suns} : \text{NP} \quad \text{lex} \quad \frac{\gamma \vdash \text{NP} \backslash \text{NP}}{\text{suns} \cdot \gamma \vdash \text{NP}} \quad \backslash E}{\text{by} \cdot (\text{suns} \cdot \gamma) \vdash \text{PP}} \quad /E}{\text{litten} \cdot (\text{by} \cdot (\text{suns} \cdot (\text{that} \cdot ((\text{the} \cdot \text{eye}) \cdot (\text{may} \cdot (\text{never} \cdot \text{behold})))))) \vdash \diamond \square (\text{NP} \backslash \text{NP})} \quad /E
\end{array}$$

(b) ...and using it to derive the full long-distance postnominal modifier.

Figure I.22: An object-relative clause in action, prompted by type assignment (I.35).

having mandatory owner and possession semantic arguments (corresponding to syntactic subject and direct object), but also any combination of depictive, duration, explanation, manner and temporal optional complements, in various orders – each variation necessarily expressed with a distinct type. In our case, we need the type:

$$\text{having} :: (\Diamond \Box (\text{NP} \setminus \text{NP}) / \text{NP}) / \text{PP} \quad (\text{I.36})$$

for a gerund that requisits first a prepositional complement phrase and then an object noun phrase (i.e. *having somewhere something*) to act as a movable post-nominal modifier (an argument permutation that FrameNet does not even contain an example of!).

The reality of optional arguments and non-trivial argument order variations alone should suffice to convince us of the issue at hand: *lexical type ambiguity* is a real phenomenon, and one that is here to stay. Having acknowledged that, the question shifts to how we deal with it. From a theoretical perspective, we can incorporate the question of type choice into our proof-machinery via the additive conjunction $\&$ of **ILL**, which is essentially recovering the functional nature of our lexicon, with type assignments reformulated as nested choices:

$$A_1 \& (A_2 \& (A_3 \dots (A_{n-1} \& A_n))) \quad (\text{I.37})$$

and the subscript enumerating each of the possible instantiations in the context of a single sentence. Under this regime, the lexical assignment rule *lex* would need to be followed by a sequence of projections to isolate the desired type, contributing little other than excessive verbosity.¹ Given the limited use we would have for all this “proof waste”, we will stick with the current formulation of the *lex* rule – if it helps us sleep better at night, we can imagine it as a shorthand notation for the correct sequence of projections requested by the current analysis, the construction of which we have delegated to a silent and omnipotent oracle. Be at rest knowing that this oracle will be temporary and for presentation purposes only; we will address its demystification later on.

The ambiguity problem is exacerbated and pushed to the limit by function words enacting context-dependent chameleon roles. Coordinators are the main culprit; they can bind together pairs of the same (almost) arbitrary type to produce an instance of the conjoined pair, a complex phrase of the same type. We will write:

$$(\chi \setminus \chi) / \chi \quad (\text{I.38})$$

to denote the coordinator type pattern parameterized over the *type variable* χ ,

¹A more ambitious usecase could allow the *simultaneous* derivation of multiple unique analyses, and the incorporation of derivational ambiguity arising out of lexical choice as a first class citizen of the proof theory – a proof object that resides *within* it rather than a notion in the meta-theory *above* it. The repercussions of this would be magnificent for semantic applications, but no concrete results that I am aware of were ever produced in that direction.

which can be instantiated as any type of our type grammar.¹

Armed with this last trick, we are now in possession of all the knowledge necessary to finally tackle the full derivation. First, we must instantiate the polymorphic coordinator once by substituting χ for NP to derive the noun phrase conjunction “*strange dolphins and sea-nymphs of unrememberable deeps*”, as portrayed in Figure I.23a. This, together with our freshly typed “*having*”, allows the derivation of the mobile postnominal modifier “*having in their whirlpools strange dolphins and sea-nymphs of unrememberable deeps*”, as in Figure I.23b. At this point, we must employ another instance of the polymorphic coordinator, this time substituting χ for $\Diamond\Box(\text{NP}\backslash\text{NP})$ – this opens the door to the derivation of the structurally free complex postnominal modifier “*litten by suns that the eye may never behold and having in their whirlpools strange dolphins and sea-nymphs of unrememberable deeps*”, which can apply to the nested “*opiate oceans*” in the same fashion as the proof of Figure I.21. At long last, we are rewarded with a type-checking and syntactically faithful analysis of the full sentence (and a check mark on *how to write a dissertation*). Collaging these last bits together is left as an exercise to the motivated reader, for fear of repetition sterilizing the quotation of its beauty.

5.1.3 Subtleties of Proof Search

The last sentence was merely a test to weed out the uncommitted. Of those that passed it and attempted to really proceed with the derivation, the observant ones should have found themselves at multiple crossroads regarding the order of applying the numerous modifiers in the sentence – a matter carefully concealed in the derivations presented so far. The choice of NL over L implies that scope assigned to competing modifiers should reflect in a corresponding judgement that differs to the rest in the bracketing structure of its antecedents (and of course the proof justifying it). The following endsequents are all valid alternatives provable with the lexical types of Figure I.23a:

- i. $\text{strange} \cdot (\text{dolphins} \cdot (\text{and} \cdot (\text{sea-nymphs} \cdot (\text{of} \cdot (\text{unrememberable} \cdot \text{deeps}))))))$
- ii. $\text{strange} \cdot ((\text{dolphins} \cdot (\text{and} \cdot \text{sea-nymphs})) \cdot (\text{of} \cdot (\text{unrememberable} \cdot \text{deeps})))$
- iii. $(\text{strange} \cdot (\text{dolphins} \cdot (\text{and} \cdot \text{sea-nymphs}))) \cdot (\text{of} \cdot (\text{unrememberable} \cdot \text{deeps}))$
- iv. $((\text{strange} \cdot \text{dolphins}) \cdot (\text{and} \cdot \text{sea-nymphs})) \cdot (\text{of} \cdot (\text{unrememberable} \cdot \text{deeps}))$
- v. $(\text{strange} \cdot \text{dolphins}) \cdot (\text{and} \cdot (\text{sea-nymphs} \cdot (\text{of} \cdot (\text{unrememberable} \cdot \text{deeps}))))$

¹This is in fact an exemplar of *parametric polymorphism*, which is properly formalized in second-order intuitionistic logic and its type-equivalent System F [Girard, 1972; Reynolds, 1974].

$$\frac{\Gamma, \alpha : \text{TYPE} \vdash M : \sigma}{\Gamma \vdash \lambda \alpha. M : \Pi \alpha. \sigma} \text{ III} \quad \frac{\Gamma \vdash M : \Pi \alpha. \sigma \quad \Delta \vdash B : \text{TYPE}}{\Gamma, \Delta \vdash M B : \sigma_{[\alpha \mapsto B]}} \text{ PIE}$$

The rules above showcase the introduction and elimination of types quantified over types $\Pi \alpha. \sigma$, and the term analogue of abstracting over types $\lambda \alpha. M$. In this notation, a coordinator would be a quantification of type $\Pi \chi. (\chi \backslash \chi) / \chi$, that when reduced against arbitrary type A would yield $(A \backslash A) / A$. Other than this unique occurrence of polymorphism, second order term and type constructions are an overkill to our purposes here, relegating this comment to footnote status.

$$\begin{array}{c}
\text{(I.20a)} \quad \frac{}{\text{strange} \cdot \text{dolphins} \vdash \text{NP}} \quad \frac{\text{and} : (\text{NP} \backslash \text{NP}) / \text{NP} \quad \text{lex} \quad \frac{\text{sea-nymphs} \cdot (\text{of} \cdot (\text{unrememberable} \cdot \text{deeps})) \vdash \text{NP}}{\text{and} \cdot (\text{sea-nymphs} \cdot (\text{of} \cdot (\text{unrememberable} \cdot \text{deeps}))) \vdash \text{NP} \backslash \text{NP}} \quad \text{(I.20d)} \quad /E}{\delta := (\text{strange} \cdot \text{dolphins}) \cdot (\text{and} \cdot (\text{sea-nymphs} \cdot (\text{of} \cdot (\text{unrememberable} \cdot \text{deeps})))) \vdash \text{NP}} \quad \backslash E
\end{array}$$

(a) Deriving noun-phrase coordination...

$$\begin{array}{c}
\frac{\text{having} : (\Diamond \Box (\text{NP} \backslash \text{NP}) / \text{NP}) / \text{PP} \quad \text{lex} \quad \frac{\text{in} : \text{PP} / \text{NP} \quad \text{lex} \quad \frac{\text{their} : \text{NP} / \text{NP} \quad \text{lex} \quad \frac{\text{whirlpools} : \text{NP} \quad \text{lex}}{\text{their} \cdot \text{whirlpools} \vdash \text{NP}} \quad /E}{\text{in} \cdot (\text{their} \cdot \text{whirlpools}) \vdash \text{PP}} \quad /E}{\text{having} \cdot (\text{in} \cdot (\text{their} \cdot \text{whirlpools})) \vdash \Diamond \Box (\text{NP} \backslash \text{NP}) / \text{NP}} \quad /E \quad \text{(I.23a)} \quad \frac{}{\delta \vdash \text{NP}} \quad /E}{(\text{having} \cdot (\text{in} \cdot (\text{their} \cdot \text{whirlpools}))) \cdot \delta \vdash \Diamond \Box (\text{NP} \backslash \text{NP})} \quad /E
\end{array}$$

(b) ...and using it to construct yet another postnominal modifier.

Figure I.23: Filling in the missing bits using the polymorphic type (I.38).

This is an admittedly stretched case of *derivational ambiguity*, a situation where from the same lexical assignments one can obtain multiple syntactic analyses, which may correspond to equinumerous subtly or drastically diverging semantic interpretations (more on that in a bit). Derivational ambiguity is not necessarily bad, provided the divergence in the proofs constructed is linguistically meaningful.¹ What is, however, worth noting is the structural discrepancy between what we see (a flat sequence) and what we want to parse into (a binary branching tree). Even though constituency structure is de facto acknowledged by linguistic theory, it is a latent mental construct revealed through (or assigned by) the parsing process, rather than an observable feature of text that we can assume as a given. The connotation of this is that even though backwards proof search in **NL** may find use in *verifying* the plausibility of a type-assigned, pre-bracketed phrase, forward search is necessary in *eliciting* a type and a bracketing structure from a phrase.

5.1.4 Syntax-Semantics Interface

The game played so far, challenging as it may be, might prove dull to someone indifferent to syntax or its type-theoretic formulation; we will attempt to fix that by expanding our target crowd to semanticists and Montagovian grammarians, who are said to recite daily before beditme:

I fail to see any interest in syntax except as a preliminary to semantics. [Montague, 1970]

¹Just think of all the different things you could do with pijamas, elephants, telescopes, etc.

Montague’s Insights A full exposition to Montague grammar is beyond the scope of this thesis, but a brief introduction to some of its foundations will go a long way in helping us perceive its relevance to the type-logical approach. Richard Montague was disillusioned with the tackling of natural language semantics at the time, which he found formally inadequate and lacking the elegance of contemporary approaches to mathematical syntax. He sought to fill this gap by arguing that formal and natural languages are morally indistinguishable – different instantiations of the same theory – and advocating their treatment in just the same way. Influenced by his own background on modal logic and the highly influential work of Saul Kripke on possible world semantics [Kripke, 1963], the machinery he thought was best fit for the task at hand was a model theoretic semantics axiomatized on the basis of set theory and higher-order logic; his work is marked with heavy use of λ notation, the adoption of which by today’s working linguist is largely attributed to him.

Revolutionary as it may have been at the time, this semantic machinery and its antiquated details are largely irrelevant to this work. What is of prime interest, though, is Montague’s treatment of the passage between syntax and semantics. In his view, if syntax is an algebra describing the process of synthesizing a grammatically passable sentence, semantics is another algebra providing a logical recipe for evaluating that sentence’s truth-validity. The two systems are viewed as distinct, but not independent: they are connected by a unidirectional transformation that preserves and transports (certain aspects of) the structure of the former into the latter, in other words a *homomorphism*. The slogan “syntax is an algebra, semantics is an algebra and meaning is a homomorphism between them” summarizes this notion [Janssen, 2014]. The gracefulness of this statement is easy to miss. It proclaims that the semantic expression assigned to complex linguistic entries mimics (or is at least informed by) the structural form of their syntactic analyses. This perspective actuates the ideal of compositionality, a concept passed down by Gottlob Frege and summarized as stating that the meaning of a complex expression is computable on the basis of its primitive expressions and the rules that dictate their combination [Partee et al., 1984].

The Type-Logical View Let’s appropriate this view and translate it to the type-logical setup, as done by van Benthem [1988]. Here, syntax is a type theory: a logic whose rules are equated to term rewrite instructions, and proofs to programs. Semantics can also be a type theory; one with its own types and terms, potentially more expressive and certainly unriddled by (some of) the structural constraints of grammar. The meaning interpretation would then be a homomorphism that translates syntactic proofs and programs to corresponding semantic ones – a translation from one constructive logic to another. Its design would need to follow the rule-to-rule approach, according to which every syntactic construction will have its homomorphic image in the target system [Bach, 1976]. This viewpoint is quite open-ended and admits a whole

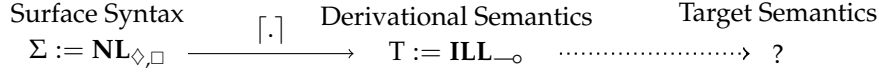


Figure I.24: The syntax-semantics interface in the type-logical setting.

lot of creative liberty with respect to the the nature of the target system and the details of the translations. The only constraint imposed is the only one that matters: the high-level principle of compositionality needs to hold, i.e. the function/argument structure specified by syntax need to be carried through to the semantics.

Interestingly, the approach permits a division of labour between syntax, semantics and everything in between: the end-to-end translation can be decomposed into a sequence of homomorphisms, each intermediate step explicating an additional layer of added expressivity (or forfeited structure) and singling out a subset of the desiderata towards the end-target. A natural first stop would be that of \mathbf{ILL}_{\multimap} as a *derivational semantics* logic: it captures the function/argument structures prescribed by the syntactic proof and respects its no-reuse principle, but without the semantically void headaches of order and bracketing structures, or that of the rules manipulating them.

To make things concrete, let's consider this in the context of the source logic Σ being identified with the instantiation of $\mathbf{NL}_{\Diamond, \Box}$ of the previous section, and the intermediate logic T being its \mathbf{ILL}_{\multimap} mirror image. Using the superscript $X := \Sigma \mid T$ to distinguish between the two logics, we will denote with Prop_0^X the set of atomic types of X , and \mathcal{U}^X its type universe, i.e. the inductive closure of types under type operators. Similarly, we will denote with Cons^X its set of constants, Vars^X its set of variable names, and Terms^X its well-formed terms, i.e. the inductive closure of terms under term operators.

The homomorphism $[\cdot]$ operates on proofs, i.e. typed terms, and thus does double duty: it transforms both terms and types of Σ to corresponding terms and types of T . It is handy, then, to define it on the basis of two components $\langle \eta, \theta \rangle$, where $\eta : \mathcal{U}^\Sigma \rightarrow \mathcal{U}^T$ and $\theta : \text{Terms}^\Sigma \rightarrow \text{Terms}^T$, such that $[s : A] = \theta(s) : \eta(A)$, where the typing relation at the right-hand side of the equation must hold (i.e. the two maps mutually respect derivability). On the type level, η must specify a pointwise mapping η_0 from the propositional constants Prop_0^Σ of the source logic to types \mathcal{U}^T of the intermediate logic. In our case, we will consider this a bijection from Prop_0^Σ to Prop_0^T , such that $\eta_0(p) \mapsto p$ (i.e. instantiating Prop_0^T as a literal copy of Prop_0^Σ). Then, to extend η_0 to η we need to specify its action on complex types, where it essentially forgets the unary modalities and removes the directionality of the implications, as shown in Table I.5. In the exact same vein, θ pointwise sends constants and variables to their copycat images, and is then inductively defined on complex terms, where it casts directional applications and abstractions to undirectional ones, drops modal decorations and performs the simplified substitution prescribed

by the $\Diamond E$ rule, as shown in Table I.6. As an example, applying $[\cdot]$ to the proof of Figure I.22 should yield the derivational term:

$$\text{litten (by (that } (\lambda x_i. (\text{may (never (behold } x_i)) \text{) (the eye))) (suns))}^{\text{NP} \multimap \text{NP}} \quad (\text{I.39})$$

\mathcal{U}^Σ	\mathcal{U}^Γ
$p \in \text{Prop}_0^\Sigma \mapsto \eta_0(p) := p \in \text{Prop}_0^\Gamma$	
$A \setminus B, B / A \mapsto \eta(A) \multimap \eta(B)$	
$\Diamond A, \Box A \mapsto \eta(A)$	

Table I.5: Translating $\mathbf{NL}_{\Diamond, \Box}$ types to $\mathbf{ILL}_{\multimap, \circ}$.

Terms $^\Sigma$	Terms $^\Gamma$
$c \in \text{Cons}^\Sigma \mapsto \theta_0(c) := c \in \text{Cons}^\Gamma$	
$x_i \in \text{Vars}^\Sigma \mapsto \theta_0(x_i) := x_i \in \text{Vars}^\Gamma$	
$s \blacktriangleleft t, t \blacktriangleright s \mapsto \theta(s) \theta(t)$	
$\lambda x_i. s, \lambda x_i. s \mapsto \lambda \theta(x_i). \theta(s)$	
$\Delta s, \blacktriangle s, \blacktriangledown s \mapsto \theta(s)$	
$\text{case } \nabla t \text{ of } x_i \text{ in } s \mapsto \theta(s)_{[\theta(x_i) \mapsto \theta(t)]}$	

Table I.6: Translating $\mathbf{NL}_{\Diamond, \Box}$ terms to $\mathbf{ILL}_{\multimap, \circ}$.

With the Curry-Howard isomorphism as our guiding star, there's no peril in navigating between syntactic and semantic theories. Syntactic proofs are equated to syntactic terms, on which our homomorphism can be applied to yield derivational semantics terms, in turn equatable to derivational semantics proofs. This might seem like a lot of work to simply “forget” syntax, but it showcases how one can step up the computational hierarchy of substructural logics in order to attain access to more expressive semantics. Note also that such a path is merely a suggestion and not an imperative; a more ambitious line of thought could maintain that word order variations (and the structural rules licensing them) can carry semantic cues which, albeit subtle, need to be upheld in the compositional meaning translation (see for instance the contemporary work of Correia [2022] for some exotic interpretations of the control modalities).

The Role of the Lexicon The sentiments of the previous paragraph could be met with some skepticism. A critical eye might argue that semantic interactions not already manifested in the syntax may never be born of this process, and thus wonder whether this added expressivity can serve any real purpose

or offer any tangible benefits. To dispel such doubts, we need to keep in mind that derivational terms refrain from specifying *lexical meaning*, i.e. they treat lexical items as black boxes, from a semantic perspective. Opening these black boxes would reveal flat entries (i.e. term constants) in the case of words providing meaning *ingredients*, as opposed to structurally rich entries (i.e. complex terms with internal structure) in the case of words providing meaning *recipes*.¹ Structurally rich lexical entries can utilize *any* term constructor made available by the semantic logic; crucially, this includes constructors that escape the narrow borders of the homomorphic codomain (i.e. do not have a syntactic origin). Of course, such terms are still bound by the promise to obey the type dictated by the homomorphic translation of their original syntactic type, and must also be derivable theorems of the semantic logic they live in. Increasing expressivity therefore may indeed not in itself add to the function/argument structures inherited by syntax, but provides the tools necessary for complex lexical semantic actions to take effect as needed.

A case in point is the coordinator “and” conjoining the two modifiers of the previous section: “*litten by ... and having in ...*”. Each individual conjunct fulfills a descriptive filter that intersects the properties of its argument with the properties attributed by its internal meaning. That is, of all objects of type $*$ (where $*$ an arbitrary type, denoting the interpretation target of NP), the first modifier withdraws all but those lit by unseeable suns, whereas the second one withdraws all but those with weird entities in their whirlpools. For the full conjunction to have the intended meaning, i.e. evoke the image of exclusively this subset of oceans characterized by both the above properties, the coordinator would need to enact the role of a portable implementation of function composition² as in Figure I.8, so as to allow the iteration of the intersective modifiers:

$$\lambda x_0 x_1 x_2. x_0 (x_1 x_2) :: (* \multimap *) \multimap (* \multimap *) \multimap * \multimap * \quad (\text{I.40})$$

Even though no non-standard term constructors are to be found in this recipe, it is nonetheless *not* a theorem of the source logic, as function composition is not derivable in **NL**. In a set-theoretic semantics domain unbound by linearity constraints, another (perhaps more reasonable) translation might make use of an added operator $\wedge : * \rightarrow * \rightarrow *$ for set-theoretic intersection ($*$ now an arbitrary set), to deliver the recipe:

$$\lambda x_0 x_1 x_2. (x_0 x_2) \wedge (x_1 x_2) :: (* \rightarrow *) \rightarrow (* \rightarrow *) \rightarrow * \rightarrow * \quad (\text{I.41})$$

¹This distinction is usually paralleled with the linguistic distinction between *content* and *function* words, but committing to this being the case is an unnecessary restriction. Depending on the end-target semantics logic and the granularity of the semantic lexicon, content words might still be assigned complex term structure – a common trick, for instance, in delivering dependent type semantics; see the book of Chatzikyriakidis and Luo [2020] for an overview of recent developments.

²Before anyone gets angry: I am neither pitching some provocative theory of conjunction semantics here, nor secretly advocating for the dot-combinator – just trying to make a point.

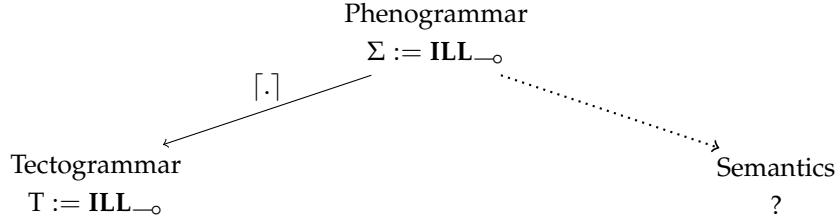


Figure I.25: The syntax-semantics interface in the abstract categorial setting.

5.2 Abstract Categorial Grammars

So far, we have been predisposed to treating syntax as the hidden process that forms grammatically correct sentences. It is insightful to contrast this treatment with the view of Curry [1961], who thought of syntax as a two-layered hierarchy of grammaticality criteria. The deep layer, called *tectogrammar*, would be concerned solely with the well-typedness of grammatical function domains and the validity of their interpretations. The shallow layer, called *phenogrammar*, would be where tectogrammatical proofs are transformed and cast to surface forms that abide by the linear order and constituency restrictions imposed by the language. Type-logical grammars pose no challenge to the legitimacy of this distinction: it should be clear that phenogrammar, in Curry's terms, is our syntactic logic, and tectogrammar is what we earlier referred to as derivational semantics. In being tectogrammar-first, however, they diverge in its operationalization. The computational pipeline they propose is sequential in nature, and follows the Aristotelian path from observable evidence to latent variables: the surface string is perceived as the yield of a (shallow) syntactic proof, from which a deep semantic proof is extracted. The operationalization closer to Curry would be inverted, placing phenogrammar at the top of the generative process, and following the Platonic information flow from deep and abstract to shallow and concrete. This perspective is embodied by abstract categorial grammars [de Groote, 2001] and their contemporary and closely related lambda grammars [Muskens, 2001]. Both are tectogrammar-first formalisms that make use of \mathbf{ILL}_{\circ} and the Curry-Howard isomorphism to obtain phenogrammatic realizations via homomorphic translations of the tectogrammatic parse.

5.2.1 Basic Definitions

The focus of our presentation will be on abstract categorial grammars, as they are closer in spirit to what is to come later. In its original definition, an abstract categorial grammar consists of two instantiations Σ, T of \mathbf{ILL}_{\circ} , and a map between them. The source instantiation Σ provides a set of base types Prop_0^{Σ} , and the so-called *abstract vocabulary*: a set of *abstract constants* Cons^{Σ} ,

each assigned a type from \mathcal{U}^Σ . The target instantiation T provides another set of base types Prop_0^T , and constants Cons^T with types from \mathcal{U}^T , called the *object vocabulary*. The map between them is once again a homomorphism $\llbracket \cdot \rrbracket$, defined on the basis of $\langle \eta_0, \theta_0 \rangle$. Not unlike before, η_0 is seen as implementing a mapping $\text{Prop}_0^\Sigma \rightarrow \mathcal{U}^T$, and θ_0 a mapping $\text{Cons}^\Sigma \rightarrow \text{Terms}^T$, both pointwise defined. Their homomorphic extension is trivially obtained by recursively defining their actions on implicational types, function terms and λ abstractions, where they simply mimic the source type- and term- structure. This formulation lends itself nicely to the notion of *grammar composition*, if one is to use the object logic of a grammar as the abstract logic of another. Each grammar is accompanied by two *languages*; the abstract language, i.e. the set of terms (of some *distinguished* type $p_d \in \text{Prop}_0^\Sigma$) derivable in the source logic, and the object language, i.e. the set of object terms the abstract language maps into. For the phenogrammar to tectogrammar picture to be made evident, the distinguished type needs to be mapped to the functional string type $p_d \mapsto \text{str}$, forcing terms of the object language to evaluate to strings. Note that, despite appearances, str is a first-order type $* \multimap *$ (where $*$ some arbitrary primitive) so as to permit the view of string concatenation as function composition, identical to (I.40):

$$+ := \lambda x_0^{\text{str}} x_1^{\text{str}} x_2^* . x_0 (x_1 x_2) \quad (\text{I.42})$$

5.2.2 Artificial Languages

Abstract categorial grammars are characterized by two measures of complexity: the maximal order of source constants' types, and the maximal order of the codomain of η_0 . The two together constitute the grammar's *class*, which concisely describes the sort of languages the grammar can model. This can prove effective in revealing a more granular stratification underlying the Chomsky hierarchy of formal grammars, when the latter are embedded into abstract categorial equivalents; as such, the framework has found extensive use as a meta-language for the study and formalization of formal grammars (as done by de Groote and Pogodalla [2004], *inter alia*).¹

To see this in practice, let's have some meta-fun pretty-printing the types of $(\mathbf{N})\mathbf{L}_{\diamond, \square}$ by modeling their type formation rules (which constitute a context-free grammar) using an abstract categorial grammar. First item on the agenda is the specification of our two logics Σ and T . The source logic Σ will provide the abstract backbone of the type grammar, containing a single base type, that of a well-formed "type" $\text{Prop}_0^\Sigma := \{\text{TYPE}\}$. The abstract vocabulary is then populated in Figure I.26 by all abstract constants denoting base "types"² and "type" constructors. The target logic T will be our phenogrammatic printer tasked with translating abstract terms ("types") to object terms (strings). We

¹There is a certain irony in formal grammars requiring or benefiting from formalization. If you're having trouble parsing this, consider that formal languages are essentially ad-hoc rules on strings; by formalization we mean giving these rules the type-theoretic treatment they deserve.

²In hindsight, that might have been an unfortunate choice of term² to overload.

$$\text{Cons}^\Sigma := \{n :: \text{TYPE}, np :: \text{TYPE}, pp :: \text{TYPE}, np :: \text{TYPE}, \\ \text{dia} :: \text{TYPE} \multimap \text{TYPE}, \text{box} :: \text{TYPE} \multimap \text{TYPE} \\ \text{ldiv} :: \text{TYPE} \multimap \text{TYPE} \multimap \text{TYPE} \\ \text{rdiv} :: \text{TYPE} \multimap \text{TYPE} \multimap \text{TYPE}\}$$
Figure I.26: Abstract lexicon for the language of $(\mathbf{N})\mathbf{L}_{\Diamond, \Box}$ types.

Abstract Constant	Object Term
n	\underline{N}
np	\underline{NP}
pp	\underline{PP}
s	\underline{S}
dia	$\lambda x_i. \underline{\Diamond} + x_j$
box	$\lambda x_i. \underline{\Box} + x_k$
ldiv	$\lambda x_i x_j. (+ x_i + \backslash + x_j +)$
rdiv	$\lambda x_i x_j. (+ x_j + / + x_i +)$

Table I.7: Object translation for the lexicon of Figure I.26.

will need a single object base type $\text{Prop}_0^T := \{*\}$, such that η_0 Some auxiliary object constants are necessary before we proceed: opening and closing brackets, a diamond and a box the two implications, and a unique match for each unique *constant* abstract constant (i.e. each abstract constant whose type is of order zero). The above – all of type str , and underlined to distinguish from functional symbols – are used by the abstract constant translation θ_0 defined in Table I.7: base constructors are mapped to their corresponding string representations, the two unary modalities simply concatenate their symbol to their single argument, whereas the two implications infix their arguments with the a slash or backslash, and wrap the result under brackets.

Figure I.27 presents the construction of the type previously assigned to

$$\frac{\frac{\text{rdiv}}{\text{TYPE} \multimap \text{TYPE} \multimap \text{TYPE}} \quad \frac{\frac{\text{dia}}{\text{TYPE} \multimap \text{TYPE}} \quad \frac{\frac{\text{box}}{\text{TYPE} \multimap \text{TYPE}} \quad \frac{\frac{\text{ldiv}}{\text{TYPE} \multimap \text{TYPE} \multimap \text{TYPE}} \quad \frac{\text{np}}{\text{TYPE}} \quad \frac{\text{np}}{\text{TYPE}}}{\text{ldiv np np} : \text{TYPE}}}{\text{box (ldiv np np)} : \text{TYPE}}}{\text{dia (box (ldiv np np))} : \text{TYPE}}}{\text{rdiv (dia (box (ldiv np np))) pp} : \text{TYPE}} \multimap E$$

Figure I.27: Constructing the type assignment of (I.33).

“litten”, $\Diamond\Box(\text{NP}\backslash\text{NP})/\text{PP}$ (contexts are intentionally left empty and axioms replaced by abstract constants for brevity). Applying the homomorphic translation to its abstract yields a printout in the form of the object term below (source function/argument brackets substituted with indendation levels for legibility):

$$\begin{aligned}
 & \lceil \text{rdiv} (\text{dia box} (\text{ldiv np np})) \rceil \text{pp} \rceil \\
 & = \lambda x_i x_j. \lceil + x_j + \lceil + x_i + \rceil \\
 & \quad \underline{\text{PP}} \\
 & \quad \lambda x_k. \lceil \Diamond + x_k \\
 & \quad \quad \lambda x_l. \lceil \Box + x_l \\
 & \quad \quad \quad \lambda x_m x_n. \lceil + x_m + \lceil + x_n + \rceil \\
 & \quad \quad \quad \underline{\text{NP}} \\
 & \quad \quad \quad \underline{\text{NP}} \\
 & \quad \quad \quad \xrightarrow{\beta^*} (\Diamond\Box(\text{NP}\backslash\text{NP})/\text{PP})
 \end{aligned} \tag{I.43}$$

Six reduction steps later and... *voilà* – our pretty printer works! The maximal order of the abstract constants is 1, and the maximal order of the translation is 2, making our grammar’s complexity class (1, 2), a proper subset of the (2,2) that encapsulates context free grammars.

5.2.3 Human Languages

Elegant and successful as they might be in their meta-theoretical enterprises, abstract categorial grammars have not fared as well with linguistic applications, in large part due to their computationally intractable nature. On the one hand, they stand out from the rest of the categorial family in not being lexicalized by default. The conceptual separation between lexicon and rules no longer holds: rules are fixed to the ones supplied by $\mathbf{ILL}_{\perp\circ}$, but inference is largely guided by the abstract constants. Abstract constants may contain lexical items that make their way to the final (object) derivation, or simply compositional recipes that leave no imprint whatsoever. At the same time, the framework is overly reliant on the constant map θ_0 (defined on a per-item basis) for the translation into the object language to take effect. Even in the lexicalized setup where the abstract lexicon is populated by words and words only, every abstract constant needs to be assigned both an abstract type and a unique object term for every phenogrammatic behavior it exhibits; two lexical dimensions, compared to the one of vanilla categorial grammars. Enforcing grammaticality while blocking overgeneration of the object language similarly requires a careful, parallel finetuning of both the abstract language and the translation – gone is the adage of words carrying their combinatorics on their sleeves. What’s worse, words triggering higher-order tectogrammatic phenomena will then need object translations of an even higher order for their

surface forms, making the design and population of a strict tectogrammatic translation $[\cdot]$ practically unfeasible. This part could in principle be partially mitigated by flattening complex syntactic phenomena into lower-order (alias) types in the source domain, and outsourcing their expansion to a parallel grammar for concrete semantics – this is less of a solution and more of a deferral, though. Beyond issues of practicality, there are also foundational problems at stake, as resorting to a lexical enumeration of phenogrammatic forms evidences inability to perform linguistic generalization – what Moot [2014] calls a problem of *descriptive inadequacy* revolving around *any* abstract replacement to a Lambek higher-order type. Last but not least, it is hard to imagine an abstract categorial grammar in action: it is unclear how to procure an abstract proof object from the *evaluated* yield of its object translation (i.e. the string form we are most likely encounter in the open) using traditional proof-theoretic disciplines – the two layers of function/argument structures (abstract- and object- level) and their interacting reductions would unnerve even the sturdiest of parsers (or so it seems).

As an artificial yet illustrative and down-to-earth example, let's brave the design of an abstract categorial grammar tasked with the production of an end-to-end linguistic example. Once more, we start with the specification of our two logics Σ and T . For efficacy and simplicity, we can have Σ coincide with the derivational semantics logic of Section 5.1.4, inheriting its terms and types for free. Doing so requires of course that we assume some high-level equivalence between the representations of what used to be abstract semantics before, and what now we call deep syntax – let's naively take this for granted. The object logic T , being responsible for the surface materialization of our derivational proofs, will again need to be a logic of strings and is thus populated by a single atomic type $\text{Prop}_0^T = \{*\}$; all abstract atoms are then sent to str . For each word, we will need an abstract constant $c \in \text{Cons}^\Sigma$, and a corresponding object constant $\llbracket c \rrbracket :: \text{str} \in \text{Cons}^T$, denoting the word's string form. Abstract constants will be sent to object terms via θ_0 , each of which must contain a single occurrence of a term constant, in order to preserve lexicalism and respect lexical transparency. Worth a special mention is the fact that the tectogrammatic image of words assigns them syntactic recipes, as they carry their own λ terms. The story is summarized in Table I.8 (object types are omitted for brevity – simply substitute all atoms of the corresponding abstract types with str to obtain them).

The equation below shows the computation of the homomorphic translation to the derivational term (I.39) inspected earlier.

Abstract Constant	Abstract Type	Object Term
eye	N	//eye//
suns	NP	//suns//
oceans	NP	//oceans//
the	N→NP	$\lambda x_i. //the// + x_i$
opiate	NP→NP	$\lambda x_i. //opiate// + x_i$
poured	NP→S	$\lambda x_i. x_i + //poured//$
behold	NP→NP→S	$\lambda x_i x_j. x_j + //behold// + x_i$
there	(NP→S)→NP→S	$\lambda x_i x_j. (x_i x_j) + //there//$
never	(NP→S)→NP→S	$\lambda x_i x_j. //never// + (x_i x_j)$
may	(NP→S)→NP→S	$\lambda x_i x_j. //may// + (x_i x_j)$
by	NP→PP	$\lambda x_i. //by// + x_i$
litten	PP→NP→NP	$\lambda x_i x_j. x_j + //litten// + x_i$
that	(NP→S)→NP→NP	$\lambda x_i x_j. x_j + //that// + x_i(//_//)$

Table I.8: Abstract lovecraftian lexicon abiding to the types of Table I.4.

$$\begin{aligned}
& \lceil \text{litten (by (that (\lambda x_i. (\text{may (never (behold } x_i))) (\text{the eye}))) (\text{suns}))} \rceil \\
= & (\lambda x_i x_j. x_j + //litten// + x_i) \\
& (\lambda x_k. //by// + x_k) \\
& (\lambda x_i x_m. x_m + //that// + x_i(//_//)) \\
& (\lambda x_n. \\
& \quad (\lambda x_o x_p. //may// + (x_o x_p)) \\
& \quad (\lambda x_q x_r. //never// + (x_q x_r)) \\
& \quad (\lambda x_s x_t. x_t + //behold// + x_s) \\
& \quad x_n \\
& \quad ((\lambda x_u. //the// + x_u) //eye//) \\
& \quad //suns// \\
& \stackrel{\beta^*}{\rightsquigarrow} \lambda x_i. x_i + //litten// //by// //suns// //that// //the// //eye// //may// //never// //behold// //_//
\end{aligned} \tag{I.44}$$

This already proves quite an endeavour; twelve reduction steps later, we are presented with a seemingly reasonable output. Upon closer inspection, though, an issue pops up: the end term is of type $\text{str} \rightarrow \text{str}$, making it an ill-fit for the discontinuous application to the *substring* $//opiate//oceans//poured//there//$. For the tectogrammatical form to be able surface correctly, some drastic adjustments are necessary. We could alter the derivational proof by including non-lexical abstract constants, or refine the source types to impose (or enable) structural variation, but either option would be undermining the cohesion between deep syntax and semantics we started from. The only alternative that does not resort to contaminating the original term – abstract and pure – with

vulgar restrictions of form would be to lift the complexity of the interpretation and design it anew.

It seems therefore that the appealing simplicity and elegance of the tectogrammatic logic is counterbalanced by an increasingly bulky and cumbersome transition to the (equally simple, yet far less elegant) phenogrammatic logic. The problem is of course more pronounced for natural languages, which overstep the strict confines of their formal counterparts [Moot, 2014]. If only we had a way to keep just the good part of a type-driven and semantically transparent deep syntax, without having to get involved with all the tedious labour of its surface materialization or the translation to it... Spoiler alert: we will in a bit.

5.3 Other Formalisms

Type-logical and abstract categorial grammars have monopolized our interest, yet are not the only members of the categorial grammar family. For the sake of completeness and impartiality, we will briefly discuss two other major flavours and contrast them to the ones so far presented.

5.3.1 Combinatory Categorial Grammars

A deviant from the categorial tradition are the broadly adopted combinatory categorial grammars [Ades and Steedman, 1982; Szabolcsi, 1989; Steedman, 2022]. These stray from the norm by rejecting the very idea of the syntactic variable (and with it, hypothetical reasoning), citing reasons of cognitive plausibility and parsing complexity. Obviously, a categorial grammar stripped of hypothetical reasoning would not amount to much on its own: it would only be able to resolve syntactically flat sentences. To regain some of the lost expressivity (ideally, exactly and only as much as needed), combinatory categorial grammars incorporate a collection of rules lent from the combinatory logic of Curry et al. [1958], albeit in restricted form. The first such rule is morpholexical in nature; it allows lexical items to raise their types once, before administering them to the syntactic derivation, forcing a flip in the local function/argument structure and allowing different semantic scopes to take effect as/when needed. The remaining rules are essentially four instances of function composition – one for each unique pair of directional implications considered. The absence of hypothetical reasoning means that these are no longer derivable theorems of some underlying type theory, but ad-hoc schemata, fixed *a priori* to fit their designated purpose. To counteract overgeneration, these rules are made available only to a pre-defined subset of the sum of lexical types, empirically specified. With respect to the interface, a combinatory derivaton can be cast into a semantic λ term the usual way; by assigning to each rule a corresponding term constructor. Note, however, that this procedure is a non-invertible *transformation* rather than an isomorphic correspondence; the purity of the Curry-Howard correspondence is lost, traded away

for the aforementioned decrease in parsing complexity.

In spite of their (non-minor) differences, the agendas of multimodal type-logical grammars and combinatory categorial grammars are quite aligned, at least at a high level: they both stipulate the presence of syntactic universals that guide structure formation, utilize them as a pathway to semantics *à la* Montague, and acknowledge the need for language-specific syntactic fine-tuning; one exercising proof-theoretic control via unary type operators and structural rules, the other controlling the applicability of the so-called combinatory rules via lexical adjustment. For better or worse, combinatory categorial grammars have taken the lion's share of the practitioners' focus: they boast an assortment of tools and annotated corpora across languages, the size of which far exceeds that of their less popular siblings – to the point where the term categorial grammars has become an almost synonym of combinatory categorial grammars.¹ I hope that, by its end, this thesis will have slightly adjusted the scales towards a healthier epistemological pluralism.

5.3.2 Hybrid Type-Logical Grammars

In the lands between type-logical grammars and their abstract siblings, there lives the strangeness of hybrid type-logical grammars. Originally proposed by Kubota and Levine [2012], hybrid type-logical grammars utilize a combination of the two slashes of traditional Lambek calculi with the non-directional linear implication, to give birth to a type grammar that combines the good aspects of both approaches while purportedly suffering the restrictions of neither. The types of a hybrid type-logical grammar are the result of two stages of induction. The first stage creates standard Lambek types, as in (I.16). The second stage is of the form

$$A, B, C := L \mid A \multimap B \quad (\text{I.45})$$

where L a valid Lambek type. The term calculus of hybrid type-logical grammars requires a translation of logical types into so-called *prosodic* types ST (for structure or string), such that all stage 1 (Lambek) types are sent to ST , and stage 2 types are inductively translated as (higher-order) functions over ST . The term constructors assigned to the elimination (resp. introduction) of the Lambek connectives is that of structure concatenation (resp. separation), whereas the term constructors assigned to the linear connective are standard function application and variable abstraction. The marriage of these two layers of abstraction in the same term calculus might seem unorthodox or at least aesthetically displeasing, but is not without merit. Concatenative terms allow the framework to relax the lexical pressure of an abstract categorial grammar by preserving the canonical categorial grammar treatment of local syntactic phenomena. Applicative terms constitute localized and controlled bursts of

¹At least if one is to consider the reviewers I get assigned a reliable statistical sample of the NLP population.

ad-hoc expressivity that can exceptionally allow the derivation of higher-order and non-local phenomena normally inaccessible to the OG Lambek calculi. Hybrid type-logical grammars are a relatively new addition to the family, and are subject to ongoing research, both from the linguistic and the proof-theoretic perspective [Kubota and Levine, 2020; Moot and Stevens-Guille, 2022]. As to why one would choose this setup over alternatives, your guess is as good as mine; hybrid proponents proclaim the system less obscure and more fit for linguistic applications [Kubota and Levine, 2020] – external validation is still pending.

6 Key References & Further Reading

Key references for this chapter were the Stanford Encyclopedia of Philosophy entry on type-logical grammars [Moortgat, 2014] and the tried-and-true extended introduction books on λ -calculi and type theories of Sørensen and Urzyczyn [2006] and Pierce [2004]. Moral credit is owed to my once faithful travel companion, the categorial grammar bible of Moot and Retoré [2012]; it provides an accessible yet detailed documentation of most of the concepts hinted at in this chapter. Sections 1 and 2 draw heavily, both in content and in style, from the excellent tutorial paper of Wadler [1993] on linear type theory – waning presentational influences might be discernible up to Section 4.

If unhappy about this chapter ending, or unsatisfied with the exposition provided, here's some extra reading material to keep you company. For a detailed inquiry on proof nets and their linguistic applications, or an exemplar of what an actual great dissertation looks like, take a look at my co-supervisor's one [Moot, 2002]. For a more mathematically eloquent presentation of modalities and their potential as tools of inferential and structural reasoning, refer to the (also superb) dissertation of Bernardi [2002]. For a slightly outdated but still very educative overview of abstract categorial grammars, the lecture notes of Kanazawa and Pogodalla [2009] should prove handy. If your eco-conscious side was moved by linear logic, but you find yourself lacking the bravery of facing the original manuscript of Girard [1987], the lecture notes of Troelstra [1991] would make for a good alternative. If on the other hand you were intrigued about the vast expanse of type theories beyond the tiny scope of this thesis, the entry point to the downwards descent into the rabbit hole should be the seminal work of Martin-Löf [1982]. A convincingly easy-to-swallow application of such type theories in the formal semantics world is extensively summarized by Chatzikyriakidis and Luo [2020]. If you do like formal semantics but big lambdas give you nausea, there's a broad selection of books to go for; I still find myself guiltily cross-checking definitions and examples with that of Winter [2016] at times. Finally, if what caught your attention was the historical drama at the beginning of the chapter, you will enjoy reading about the history of constructivism by Troelstra [2011].

Chapter Bibliography

- S. Abramsky. Computational interpretations of linear logic. *Theoretical computer science*, 111(1-2):3–57, 1993.
- A. E. Ades and M. J. Steedman. On the order of words. *Linguistics and philosophy*, 4(4):517–558, 1982.
- K. Ajdukiewicz. Die syntaktische konnexitat. *Studia philosophica*, pages 1–27, 1935.
- E. Bach. An extension of classical transformational grammar. 1976.
- C. F. Baker, C. J. Fillmore, and J. B. Lowe. The berkeley framenet project. In *COLING 1998 Volume 1: The 17th International Conference on Computational Linguistics*, 1998.
- Y. Bar-Hillel. A quasi-arithmetical notation for syntactic description. *Language*, 29(1):47–58, 1953.
- H. P. Barendregt et al. *The lambda calculus*, volume 3. North-Holland Amsterdam, 1984.
- R. A. Bernardi. *Reasoning with polarity in categorial type logic*. PhD thesis, 2002.
- S. Chatzikyriakidis and Z. Luo. *Formal Semantics in Modern Type Theories*. John Wiley & Sons, 2020.
- A. Church. A formulation of the simple theory of types. *The journal of symbolic logic*, 5(2):56–68, 1940.
- A. D. Correia. *Quantum distributional semantics: Quantum algorithms applied to natural language processing*. PhD thesis, Utrecht University, 2022.
- H. B. Curry. Functionality in combinatory logic. *Proceedings of the National Academy of Sciences*, 20(11):584–590, 1934.

- H. B. Curry. Some logical aspects of grammatical structure. *Structure of language and its mathematical aspects*, 12:56–68, 1961.
- H. B. Curry, R. Feys, W. Craig, J. R. Hindley, and J. P. Seldin. *Combinatory logic*, volume 1. North-Holland Amsterdam, 1958.
- V. Danos and L. Regnier. The structure of multiplicatives. *Archive for Mathematical logic*, 28(3):181–203, 1989.
- N. G. de Bruijn. Automath, a language for mathematics. In *Automation of Reasoning*, pages 159–200. Springer, 1983. Original manuscript from 1968.
- P. de Groote. On the strong normalization of natural deduction with permutation-conversions. In *International Conference on Rewriting Techniques and Applications*, pages 45–59. Springer, 1999.
- P. de Groote. Towards abstract categorial grammars. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, pages 252–259, 2001.
- P. de Groote and S. Pogodalla. On the expressive power of abstract categorial grammars: Representing context-free formalisms. *Journal of Logic, Language and Information*, 13(4):421–438, 2004.
- J.-Y. Girard. *Interprétation fonctionnelle et élimination des coupures de l'arithmétique d'ordre supérieur*. PhD thesis, Éditeur inconnu, 1972.
- J.-Y. Girard. Linear logic. *Theoretical computer science*, 50(1):1–101, 1987.
- S. Guerrini. A linear algorithm for mll proof net correctness and sequentialization. *Theoretical Computer Science*, 412(20):1958–1978, 2011.
- P. Hendriks. Ellipsis and multimodal categorial type logic. In *Proceedings of Formal Grammar*, pages 107–122. Citeseer, 1995.
- A. Heyting. Die formalen regeln der intuitionistischen logik. *Sitzungsbericht Preussische Akademie der Wissenschaften Berlin, physikalisch-mathematische Klasse II*, pages 42–56, 1930.
- W. A. Howard. The formulae-as-types notion of construction. *To HB Curry: essays on combinatory logic, lambda calculus and formalism*, 44:479–490, 1980. Original manuscript from 1969.
- T. Janssen. *Foundations and applications of Montague grammar*. PhD thesis, University of Amsterdam, 2014. Original publication date 1983.
- M. Kanazawa and S. Pogodalla. Advances in abstract categorial grammars: Language theory and linguistic modeling. *Lecture notes, ESSLLI*, 9, 2009.

- S. A. Kripke. Semantical analysis of modal logic i normal modal propositional calculi. *Mathematical Logic Quarterly*, 9(5-6):67–96, 1963.
- Y. Kubota and R. Levine. Gapping as like-category coordination. In *International Conference on Logical Aspects of Computational Linguistics*, pages 135–150. Springer, 2012.
- Y. Kubota and R. D. Levine. *Type-logical syntax*. MIT Press, 2020.
- N. Kurtonina and M. Moortgat. Structural control.
- J. Lambek. The mathematics of sentence structure. *The American Mathematical Monthly*, 65(3):154–170, 1958.
- J. Lambek. On the calculus of syntactic types. *Structure of language and its mathematical aspects*, 12:166–178, 1961.
- P. Martin-Löf. Constructive mathematics and computer programming. In *Studies in Logic and the Foundations of Mathematics*, volume 104, pages 153–175. Elsevier, 1982.
- R. Montague. Universal grammar. *Theoria*, 36(3):373–398, 1970.
- M. Moortgat. Multimodal linguistic inference. *Journal of Logic, Language and Information*, 5(3):349–385, 1996.
- M. Moortgat. Categorical type logics. In *Handbook of logic and language*, pages 93–177. Elsevier, 1997.
- M. Moortgat. Constants of grammatical reasoning. *Constraints and resources in natural language syntax and semantics*, pages 195–219, 1999.
- M. Moortgat. Typelogical Grammar. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2014 edition, 2014.
- R. Moot. Hybrid type-logical grammars, first-order linear logic and the descriptive inadequacy of lambda grammars. *arXiv preprint arXiv:1405.6678*, 2014.
- R. Moot and C. Retoré. *The logic of categorial grammars: a deductive account of natural language syntax and semantics*, volume 6850. Springer, 2012.
- R. Moot and S. J. Stevens-Guille. Logical foundations for hybrid type-logical grammars. *Journal of Logic, Language and Information*, 31(1):35–76, 2022.
- R. C. A. Moot. *Proof nets for linguistic analysis*. PhD thesis, 2002.
- G. Morrill. Type logical grammar: Categorical logic of signs. 1994.

- A. S. Murawski and C.-H. Ong. Dominator trees and fast verification of proof nets. In *Proceedings Fifteenth Annual IEEE Symposium on Logic in Computer Science (Cat. No. 99CB36332)*, pages 181–191. IEEE, 2000.
- R. Muskens. Lambda grammars and the syntax-semantics interface. 2001.
- B. Partee et al. Compositionality. *Varieties of formal semantics*, 3:281–311, 1984.
- B. C. Pierce. *Advanced topics in types and programming languages*. MIT press, 2004.
- D. Prawitz. A proof-theoretical study. *Uppsala: Almqvist & Wiksell*, 1965.
- J. C. Reynolds. Towards a theory of type structure. In *Programming Symposium*, pages 408–425. Springer, 1974.
- B. Russell. Mathematical logic as based on the theory of types. *American journal of mathematics*, 30(3):222–262, 1908.
- M. H. Sørensen and P. Urzyczyn. *Lectures on the Curry-Howard isomorphism*. Elsevier, 2006.
- M. Steedman. Combinatory categorial grammar. 2022.
- A. Szabolcsi. Bound variables in syntax (are there any?). *Semantics and contextual expression*, 295:318, 1989.
- A. S. Troelstra. Lectures on linear logic. 1991.
- A. S. Troelstra. History of constructivism in the 20th century. *Set Theory, Arithmetic, and Foundations of Mathematics*, pages 150–179, 2011.
- A. S. Troelstra and H. Schwichtenberg. *Basic proof theory*. Number 43. Cambridge University Press, 2000.
- J. van Benthem. *The semantics of variety in categorial grammar*. John Benjamins, 1988.
- W. Vermaat. *Controlling movement*. PhD thesis, 1999.
- P. Wadler. A taste of linear logic. In *International Symposium on Mathematical Foundations of Computer Science*, pages 185–210. Springer, 1993.
- H. Wansing. Formulas-as-types for a hierarchy of sublogics of intuitionistic propositional logic. In *Workshop on Nonclassical Logics and Information Processing*, pages 125–145. Springer, 1990.
- H. Wansing. Sequent systems for modal logics. *Handbook of philosophical logic*, pages 61–145, 2002.
- Y. Winter. *Elements of formal semantics: An introduction to the mathematical theory of meaning in natural language*. Edinburgh University Press, 2016.

CHAPTER II

Typing Dependency Structure

*“Predicates are functors,
complements – diamonds,
adjuncts – boxes;
Everything a type.”*

The previous chapter initiated us into the history-rich world of substructural logics in the intuitionistic tradition. Along the (artificially homogenized) story, we got to dip our toes into linguistic waters, where we saw these logics thrive and prosper, finding their place as the foundation for categorial grammars. The many flavours of categorial grammars all have a single common denominator: they treat syntax as a hierarchical structure that puts phrases together from small to big, starting from words and reaching up to the sentence, the imprint being a natural deduction tree (and perhaps a phrasal bracketing structure). This emphasis on the phrase, combined with the distinctive shape of the categorial parse, allows for a partial parallel to be drawn between categorial grammars and phrase structure grammars, despite their stark methodological and theoretical contrasts. Phrase structure grammars are rule-based systems that assign categories to phrases according to their syntactic function, and manipulate phrasal formation by specifying how their constituent parts combine – the produce being a bracketing structure, commonly visualized in tree format. A different approach to grammatical theory abandons the constituency relation, adopting the dependency relation in its stead. Dependency relations do not seem compatible with the categorial setup at a first glance: they are flat, and lack the notion of finite phrasal parts – in showing no attachment to iterative phrasal division, they are also not obviously compositional.

In this, chapter we will focus our efforts into bridging this gap between these two perspectives under a unified categorial grammar setup. We will motivate the incorporation of dependency relations into the categorial vocabulary by repurposing existing and well-studied tools that remain faithful to the type theory roots the previous chapter has established (hint: it's the modalities). We will then discuss how their inclusion alters the structural paradigms of the previous chapter, and the opportunities and problems this change comes with. In a show of poise and confidence, we will finally put the framework straight to the test, using it as a formal logic within which we will (attempt to) “prove” a large scale corpus of real-world Dutch written sentences¹. This will give us the chance to evaluate the linguistic utility of the logic, while also shedding light on its parallel treatment of function-argument structures and dependency relations in several linguistically interesting constructions. As a bonus, we get to not just keep the proofs for our future endeavours, but also share them for the whole world to do with them as they please.

7 Phrase vs. Dependency Structure

Before we get to theorycrafting, it would be useful to try and clarify what exactly is meant by constituency- and dependency- structure, and how the two differ.

7.1 Phrase Structure Grammars

Phrase structure grammars build on the observation that certain phrases seem to act as rigid and independent chunks, sometimes referred to as *constituents*. Viewed from within, these phrases may be rich in internal structure, but keep it sealed off to the outside. Viewed externally (i.e. in the context of a wider phrase that contains them), they are indivisible units, or at least for the purposes of phrasal composition. Phrases are inventorized according to their syntactic *categories*. If one so wishes, they can for the most part replace a phrase for another of the same category, with no effect to grammaticality or local structure, which suggests they are functionally indiscernible. The examples below testify to this¹; the underlined phrases can be freely interchanged – despite their wildly different internal structures, substituting one for another has no effect on the outer sentential structure:

- i. he · (beheld · (the · city))
- ii. he · (beheld · (the · ((glittering · minarets) · (of · (the · city)))))
- iii. he · (beheld · ((such · beauty) · (of · ((red · (and · white)) · flowers)))))
- iv. he · (beheld · ((some · (feature · (or · arrangement))) · (which · (he · ((had · known) · before)))))

¹~~todo~~: contradiction

¹Sourced from H.P. Lovecraft, *Celephaïs* (1922). In *The Rainbow*, vol. 2.

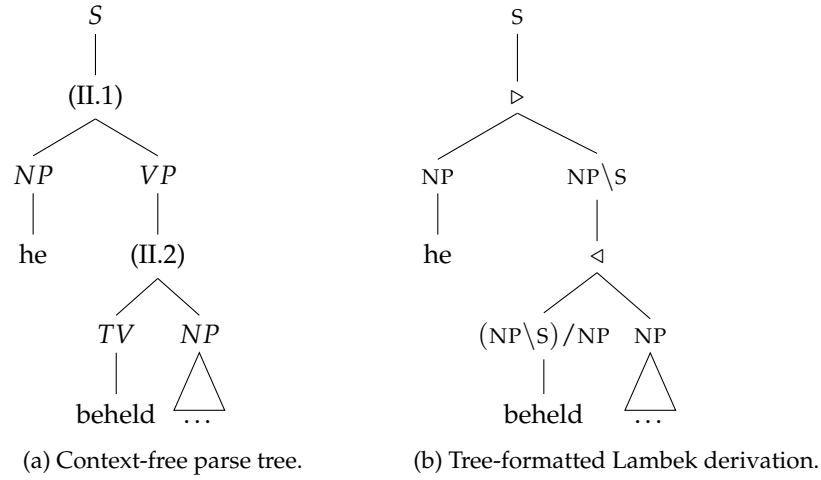


Figure II.1: A phrase-structure grammar parse tree (a) contrasted with a Lambek abstract syntax tree (b).

These so-called constituents interact, then, with one another depending not on their contents, but rather their categories. This perspective promotes a disciplined approach to grammar modeling, dating back to the formal grammars of Chomsky [1956], the archetypical example being context-free grammars [Chomsky, 1956; Backus, 1959]. There, the construction of complex expressions is guided by *production rules*, grammatical recipes that dictate what categories can sequentially combine, in what order, and what the category of their combination is. The above example would correspond, for instance, to production rules of the form:

$$S \rightarrow NP VP \quad (II.1)$$

$$VP \rightarrow TV NP \quad (II.2)$$

claiming that one of the ways to make a verb phrase VP involves concatenating a transitive verb TV with a noun phrase NP , which in turn can be plugged to the right of another NP to produce a declarative sentence S – each rule leaving a bracketing structure (or binary tree) in its wake (see Figure II.1a).

Even though context-free grammars are no longer seriously considered in the linguistic world, they have directly influenced most early attempts at grammar design – and by extension, their later successors and refinements. Notational evidence of this past are more than noticeable today, ranging from the wide adoption of tree-style notation for syntactic analyses to the conceptual syncretism of constituency grammars and phrase structure ones. More up-to-date frameworks expand upon the barebones context-free backend with niceties like a separation of functional dominance and linear precedence, added

rules that manipulate movement and discontinuity, the proclamation of a single category as the *head* of a production rule (or subtree), feature markings that carry semantic, morphological or phonological information, incorporation of dependency information, etc. [Gazdar et al., 1985; Jacobson, 1987; Pollard and Sag, 1994; Dalrymple, 2001, *inter alia*].

To our ill fortune, the field of formal syntax is actually an informal mess, more akin to a mine field rather than an academic one; it's best if I tread carefully and refrain from overextending myself here, in order to avoid setting off unseen traps or causing easily avoidable confusion. The point I want to make is that the focal center of the above formalisms is the *phrase* and its structure – as such, they are all referred to as phrase structure grammars, regardless of whatever extra fluff they carry or what their expressive capacity is. In that broader sense of the term, and removing any implicit connotations of intellectual lineage, categorial grammars can also be conceived as phrase-centric. Pure Lambek systems, for one, also explicate how phrases are combined, adhere to hierarchical forms similar to those of context-free grammars (except beautifully, see Figure II.1b), and have the exact same expressive capacity [Pentus, 1993]. They abstract away from the rule inventory by utilizing the smallest and purest set of rules possible – those of function application and variable abstraction – and internalize what used to be rule-imposed structure within the lexical categories themselves. Bracketing structure is now the footprint of function application, and the interface with semantics is naturalized by virtue of the Curry-Howard correspondence, as we saw earlier. Rather than a *VP* category and rule (II.1), we have the *type* $\text{NP} \backslash \text{S}$ – transparent with respect to both its syntactic combinatorics and semantic function. Performing the function application on a left-adjacent NP will then result to a local tree structure, not unlike the corresponding production rule – see Figure II.1 for a comparison. Note that in reality, categorial derivations in the type-theoretic tradition resemble trees only locally, since in high-order phenomena involving abstractions the unary, non-terminal λ nodes will either need to be uniquely named with the variable they are binding, or otherwise point to it with an additional edge – hence a directed acyclic graph could make for a more accurate representation format. Long story short, even without extensions to the logical core for managing discontinuity, calling deductive parsing constituency parsing would obviously not be doing the former justice; yet despite their methodological and theoretical divergences, their end yield is comparable – the antecedent structures of Figures I.20 to I.23 testify that the former may in fact be seen as subsuming the latter.

7.2 Dependency Grammars

The constituency tradition has co-evolved along the opposing view of dependency grammars [Tesnière, 2015; Sgall et al., 1986; Mel'cuk, 1988; Sleator and Temperley, 1995, *inter alia*]. Dependency grammars reject the binary phrasal division that constituency grammars abide by, and instead adopt a flatter

structural form, the only unit of which is the word. Words are connected with one another by dependency arcs, i.e. directed edges between word pairs. Each word can have arbitrarily many outgoing edges (dependents), but only a single incoming edge (head) – the exception is the root word which has no head of its own (i.e. the head of the matrix clause). A word is said to directly dominate its dependents, and indirectly dominate all words its dependents dominate (directly or otherwise) – e.g. the root indirectly dominates every other word in the sentence. This distinction between head and dependent is central to dependency grammars; broadly speaking, heads can be thought of as the words that decide the syntactic functionality of the collection of words (for fear of calling it a phrase) they indirectly dominate. The dependency structure of a sentence is once more a tree, with words now as both terminal and non-terminal nodes, glued together with dependency relations. A dependency tree is unconstrained by adjacency and word order: edges can penetrate or fly over other edges. This perspective is computationally appealing due to its simplicity and uniformity, as it allows a dependency grammar to argue about languages with wildly diverging syntactic and typological properties while remaining virtually unchanged. For the exact same reasons, it can also be seen as concealing – it sacrifices any potential of targeted analysis in the pedestal of universality. Finally, the semantically inclined might find a two-directional extension of dependency arcs enticing. In that setup, the added direction (which needs not agree with that of syntactic dominance) is devoted to semantic information flow, pointing from semantic predicates¹ to semantic arguments [Mel’cuk, 2003].

A dependency grammar that has gained significant traction over the last decade is the framework of universal dependencies [de Marneffe et al., 2021], claiming a broad collection of multi-lingual treebanks [Nivre et al., 2020] and tools. In universal dependencies, words are usually assigned a label pulled from a rudimentary set of part of speech tags and lexical identifiers and more importantly, dependency relations are also labeled according to their grammatical function, allowing the distinction of a words’ dependents according to the grammatical role they fulfill. Grammatical roles are typologically and thematically informed, and are inventorized with language universality as the prime goal. This inventorization upholds no semantic promises, but is not inconsistent with the aforementioned semantic view either. To obtain a semantic transcription of the dependency tree, one needs only specify whether the semantic flow of each grammatical role is co- or contra- directional to the edge’s syntactic flow, i.e. whether the arc marks its dependent as a *complement* (where syntactic head and semantic predicate coincide) or an *adjunct* (where the syntactic head is the semantic argument to its syntactic dependent)².

¹Apparently also an overloaded term. I will only ever use this in the strictly logical sense.

²Universal dependencies, a staunch pacifist, carefully and vocally refuses to make this claim, as complements and adjuncts are largely language-particular syntactic constructs and a notorious point of debate [Haspelmath, 2014]. I would like to believe I am not trespassing here, either – as will be made clearer in a bit, I employ the two terms in a purely semantic fashion.

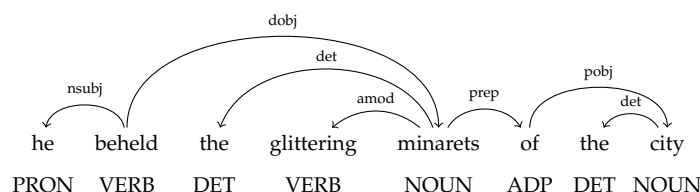


Figure II.2: A sample dependency parse in the universal dependencies format.

Figure II.2 shows an example dependency parse. Unlike before, we can not claim any semblance to the proofs that have occupied us thus far. At a first glance, dependency grammars have little in common with categorial grammars – structures are no longer binary nor made out of phrases, the axis of grammatical functions is completely new, and there seems to be little there reminiscent of the notions of induction and composition. Though on closer inspection and armed with some goodwill, we can recover from some of these divergences if we make a few concessions from both sides. We can start by treating any collection of words rooted in the same ancestor in the dependency graph as a constituent phrase, albeit possibly discontinuous – the result will give us at least some partial overlap with the categorial directive. Grammatical functions can then be thought of as being implicit, having been internalized in their positioning within a functor. For instance we do intuitively know that the Lambek transitive verb $(NP \backslash S) / NP$ requires an *object* NP to the right and a *subject* NP to the left – marking them as such is perhaps redundant, since the verbal meaning recipe places each syntactic argument into a distinct semantic slot. The binary bracketing structure is irrevocably lost, but this loss can be deemed as inconsequential if we “flatten” functor-induced phrasal boundaries by considering them only at these intermediary points where all of their arguments (however many) have been applied. It is not much further we can get with this mediatory role, though. No concession from the categorial side would be able to justify the underspecification of higher-order phenomena in a dependency graph, and no concession from the dependency side could make peace with the omission of the concept of headedness in an applicative natural deduction proof.

8 Modalities for Dependency Demarcation

In our new quest, we will seek to design a type logic that subsumes and rises above both phrase structure grammars and dependency grammars. As we saw in the previous section, the bar is not set particularly high for the first kind; the Lambek calculus can already do more than well enough. The challenge then is to integrate the added values of a dependency grammar in a type-theoretic framework. There’s two elements we are missing; distinguish-

ing between syntactic heads and syntactic/semantic predicates, and marking words and phrases according to their grammatical roles within some wider context.

8.1 Two Dimensional Predicates

Categorial grammars are inherently and by design biased towards predicate structures, primarily syntactic, but simultaneously also semantic (if one is to believe the story of Section 5.1.4, no distinction can be made between the two). Each phrase can be iteratively split apart into two subphrases (not necessarily contiguous), where one provides a functor, and the other the argument thereof. Note that this distinction does not preclude the possibility that the argument itself has a functional type – no assumption is made on the form of either subphrase’s type, other than the two being compatible. But what assumes the role of the functor in a local domain needs not always be the syntactic head of that domain. There’s a plethora of example cases. Quantifiers, for one, are inarguably predicates over the objects they quantify, yet they exactly obey the morphosyntactic characteristics prescribed by these objects (i.e. grammatical gender, case, number, etc.), evidencing that the latter are in fact the heads – a clear violation of any alignment between syntactic predicate and syntactic head we could ever hypothesize. A similar argument can be made for determiners and, more broadly speaking, any phrasal element that takes functional precedence without being the syntactically prominent part of its phrase, e.g. adjectival and adverbial modifiers.

This observation gives rise to a binary subcategorization of a binary predicate structure; to establish some risky terminology, it is either:

- i. an application of a head to its *complement*, or
- ii. an application of an *adjunct* to its head

where the distinction between complement and adjunct is made solely on the basis of their functional relation to the head.

The vanilla categorial vocabulary does not suffice to capture this extra dimension of function application – a problem also noticed by the intellectuals of proto-categorial civilizations, as archeological excavations reveal. In an unpublished manuscript, Moortgat and Morrill [1991] propose a two-dimensional implicational type operator and corresponding residuation laws: the first binary dimension is reserved for the usual left- vs. right- application distinction, whereas the second binary dimension specifies whether the head occurs to the left or to the right; the result is four unique ways of building up an implication. Congruent with the substructural trend of revealing structure that was once hidden, this division brings forth a two-valued structural binder, allowing the corresponding logic **DNL** to reason about *headed* binary trees. The authors refrain from committing to a specific linguistic application, but, translated into our terminology, their proposal can be schematically summarized by Figure II.3. Further away from syntax, Hendriks [1997] employs **DNL** to

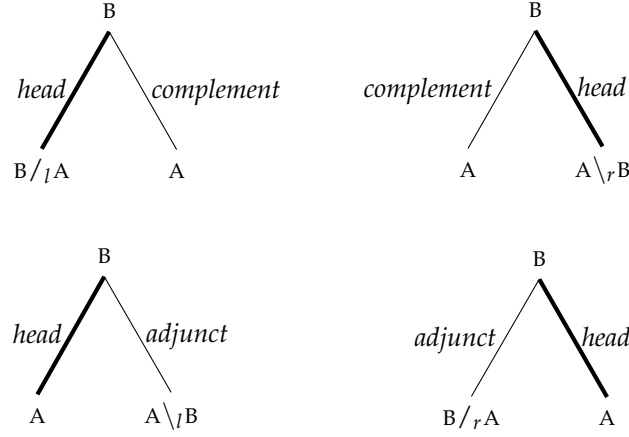


Figure II.3: The four implications of DNL.

account for prosodic intonation structures, where the head is assigned to vocally prominent elements.

8.2 Modal Dependents

As an alternative to introducing implicational (and by residuation, product) variants, we can instead opt for the more fashionable modal decomposition approach [Kurtonina and Moortgat, 1997]. This allows us to view a specialized (here: head-aware) implicational variant as a composition of its uniform base with a unary modality. The standard route would have used the modality to mark the head – we will instead mark the dependent. More than a petty act of rejection to establishment, this shall provide us with the means to further differentiate dependents according to the exact grammatical slots they occupy – after all, there’s quite a few different dependency labels, but only the one head.

8.2.1 Complements vs. Adjuncts

Sticking to our risky agenda, the first distinction we need to make is that of complements versus adjuncts. In the complement case, such a decomposition would look as follows:

$$B/_1A \equiv B/\Diamond A \quad (\text{II.3})$$

$$A_rB \equiv \Diamond A_rB \quad (\text{II.4})$$

The translation is straightforward: predicates in head position are functors requiring the same arguments as they would before, except now under a di-

amongd. In that sense, they *assign* diamonds to their complements by necessitating an application of the $\Diamond I$ rule of Figure I.15 prior to the function application. Recalling the structural imprint of the rule, this results in an extra layer of bracketing structure the delimits complement phrases and isolates them from their surroundings.

The adjunct case is slightly more obscure. Following the directives of the previous paragraph, we need to mark the dependent – this time a predicate in non head position – in a way such that its application on its argument leaves a bracketing imprint on the structure of the former rather than the latter. The solution manifests in the form of a box:

$$B/_r A \equiv \Box(B/A) \quad (\text{II.5})$$

$$A \backslash_l B \equiv \Box(A \backslash B) \quad (\text{II.6})$$

The translation is not much different: adjuncts are predicates wrapped by a box. To reveal the pure function contained therein and allow a proof to progress, we need to invoke the $\Box E$ rule of Figure I.15, the effect being a bracketing structure that now delimits adjunct phrases.

There is symmetry between the above two cases. The task of imposing dependency structure is always upon the functional predicate and its type. Head predicates mark their complements, whereas adjunct predicates mark themselves; in either case, it is the dependent structure that gets the brackets. The duality of predicate structure is thus mirrored in the innate distinction between function and argument of the applicative categorial backend, whereas the duality of syntactic headedness is captured by the unary modalities; type-checking in both dimensions.

8.2.2 Grammatical Functions

Let's take this a bit further. Universal dependencies may make no adjunct vs. complement distinction, but they go the extra mile of subspecifying dependents according to the exact grammatical roles they play. Extending our grammar logic accordingly is trivial. Rather than have a single diamond and box, we can consider a usecase where modalities are a *family* of unary residuals, i.e. a set of pairs, each labeled according to a single, unique dependency label. The generalization is a multimodal type system consisting of modal pairs:

$$\{(\Diamond_d, \Box_d) \mid d \in \text{Deps}\} \quad (\text{II.7})$$

where Deps the full set of dependency labels made available to each specific instantiation of the theory¹. The edge case of Deps being a singleton set col-

¹Note that the label set can vary depending on the designer's end goal; grammatical functions is just one of the possibilities. The setup is also more than compatible with frame semantics, where event-specific semantic structures (*frames*) are evoked by lexicalized syntactic heads to assign semantic/thematic roles to their dependents (*frame elements*) [Fillmore et al., 1976].

lapses to the previous exposition (whereby explicit labeling is redundant).

Each instance of a labeled modality will now come with its own introduction and elimination rules. Concomitantly, both the term calculus and the bracketing structures are extended with multiple labels; the modal rewrites and unary brackets of Section 4.1 are now differentiated on the basis of the dependency label that induced them. Unlike before, the structural effect is not a means to the end of structural reasoning, but the very purpose of the dependency modalities – as such, they are not necessarily associated with any structural rules (even though nothing precludes the possibility – it might even be reasonable to condition each dependency or combinations thereof to a unique set of structural irregularities, as we will see in a bit). Note, also, that the residuation properties and normalization routines apply only between diamonds and boxes of the *same label* – no interaction between mismatched types and terms is stipulated.

8.3 Inference with Dependency-Enhanced Types

Logical inference in the setup envisaged here is not dissimilar conceptually to the standard type-logical pipeline of Section 5.1, but there’s some crucial differences that require explication, plus a few critical gotchas to beware of.

8.3.1 Initial Lexical Adjustments

For starters, functors previously involved with simple applicative phenomena will now need to abide to either of the type patterns below:

$$A, B := \diamond_d A \backslash B \mid B / \diamond_d A \mid \square_d (A \backslash B) \mid \square_d (B / A) \quad (\text{II.8})$$

For these dependency-enhanced types to appear and take effect, the lexicon needs to be adjusted accordingly.

It is the lexicon’s first duty then to discriminate between head and non-head functors by decorating them or their arguments with the appropriate modalities. An intransitive, for instance, would now be typed as $\diamond_{\text{su}} \text{NP} \backslash S$ – to produce a sentence, the type demands to its left not just any noun phrase, but rather one marked as a subject. The story is no different with more than one complements – i.e. a transitive would be $(\diamond_{\text{su}} \text{NP} \backslash S) / \diamond_{\text{obj}} \text{NP}$, and so on. A determiner, however, would be typed as $\square_{\text{det}} (\text{NP} / N)$ – it recognizes the right-adjacent noun as its head, but still takes functional precedence over it, licensing the function application by dropping its determiner box. Similarly, a prenominal modifier would be typed as $\square_{\text{mod}} (\text{NP} \backslash \text{NP})$ – to apply to its unmarked head, the type would need first liberate itself of its box, being a modifying adjunct.

Atomic type assignments will remain for the most part unchanged, as propositional constants are necessarily complements (or heads of a singleton phrase, to be pedantic) and their grammatical role cannot be decided *a*

priori, anticipating a phrasal head to enforce it instead. Plural nouns like the “dolphins” and “whirlpools” of Table I.4, for instance, would still be typed as NP – there’s no telling in advance whether they will occur as subjects, direct objects or something else. Exceptionally for words whose morphological characteristics already confine them to a single possible grammatical role, we can consider an alternative typing that restricts them to exclusively that grammatical role. The straightforward thing to do would be to lexically mark them with a diamond – e.g. for the nominative version of the third person singular personal pronoun, “he”, we might assign the type $\diamond_{su}NP$, denoting it must necessarily occur in subject position. However, this would create a structural asymmetry between a nominal *assigned* the subject role via the $\diamond_{su}I$ rule (inducing corresponding brackets) versus the pronoun *carrying* the subject role (and thus remaining bracket-free). To break this asymmetry, a better alternative would be to use a lexical assignment that rests on the *closure* operator of (I.24) instead, i.e. $\Box_{su}\diamond_{su}NP$. Now for the verbal head to find its subject-marked argument, the pronoun would need to reveal its diamond via the $\Box_{su}E$ rule, independently bracketing *itself* in the process, while excluding any potential for grammatical misuse.

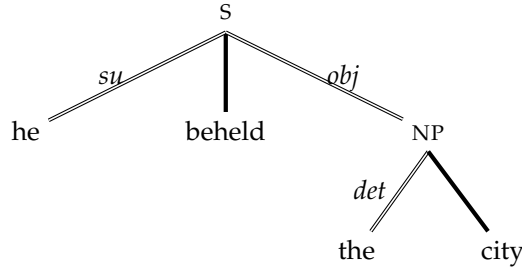
8.3.2 Dependencies and Structural Reasoning

Lambek vs. Lambek (vs. van Benthem) The next thing to consider is how the inclusion of dependency modalities alters the structural core of each base logic. In any case, modalities induce a multi-labeled unary bracketing structure. For the non-associative **NL**, the result is trees of mixed but consistent arity – our subclassing of functors in (II.8) means that each binary branch (imposed by a function/argument structure) will contain a normal branch that corresponds to the local head *phrase* and a distinguished unary branch that labels the non-head phrase – brackets and parentheses galore. Opting for the more traditional associative base of **L** changes the scenery to one of shorter, wider and less exuberant trees. Since the vanilla structure is now just a sequence, treeness is imposed solely by the modal brackets; the result is a variadic but uniform tree structure, where each subtree contains a single local head *word* and multiple dependent phrases, each of them in turn wrapped under a unary branch (or simply just having its edge labeled, to make things easier to the eye). This visual paradigm also applies to the even laxer **NLP** – the difference being that the yield of each local tree is recursively equivalent under bracket-preserving permutation, i.e. commutativity now holds between constituent phrases (subtrees) rather than words (terminal nodes).

The above points hint to the fact that dependency modalities introduce structural constraints that may (to some extent) obviate the need for a strict structural binder. From the linguistic perspective, **L** seems to hit the sweet spot – it has constituents live happily together in horizontal, non-binary clusters set upon lush trees, with each constituent given a role to fulfill. The systematic ordering of arguments according to their obliqueness order is made

$$\begin{array}{c}
\frac{\frac{\text{he} : \square_{su} \diamond_{su} NP}{\langle \text{he} \rangle^{su} \vdash \diamond_{su} NP} \text{lex} \quad \frac{\frac{\text{beheld} : (\diamond_{su} NP \backslash S) / \diamond_{obj} NP}{\text{beheld}, \langle \langle \text{the} \rangle^{det}, \text{city} \rangle^{obj} \vdash \diamond_{su} NP \backslash S} \text{lex}}{\langle \text{he} \rangle^{su}, \text{beheld}, \langle \langle \text{the} \rangle^{det}, \text{city} \rangle^{obj} \vdash S} \backslash E \\
\frac{\frac{\frac{\text{the} : \square_{det} (N / NP)}{\langle \text{the} \rangle^{det} \vdash N / NP} \text{lex} \quad \frac{\text{city} : N}{\langle \langle \text{the} \rangle^{det}, \text{city} \rangle^{obj} \vdash NP} \text{lex}}{\langle \langle \text{the} \rangle^{det}, \text{city} \rangle^{obj} \vdash \diamond_{obj} NP} \diamond_{obj} I \\
\frac{\langle \langle \text{the} \rangle^{det}, \text{city} \rangle^{obj} \vdash \diamond_{obj} NP}{\langle \langle \text{the} \rangle^{det}, \text{city} \rangle^{obj} \vdash \diamond_{su} NP \backslash S} /E
\end{array}$$

(a) Simple applicative derivation in dependency-enhanced L.



(b) Corresponding tree structure, read off the antecedent. As before, heavy edges denote heads, and double edges telescope unary branching (now labeled).

Figure II.4: The structural effect of dependency-enhanced functors.

redundant by their labeling; the positional explication of **NL** is replaced by the denominational explication of the modal brackets. What's more, headedness is not proliferated among functionally incomplete constituents, i.e. each complete phrase (read: one typed as a propositional constant) is flat among its arguments, requiring only a single head and implicitly disallowing heads from being phrases in themselves (in line with the mandates of dependency grammars). Figure II.4 presents a simple first example to illustrate the point.

How about **NLP** though? We have so far dismissed the syntactic utility of the logic as being overly permissive, presenting it as meaningful only from a semantic perspective. With dependency brackets in the picture, the explosive combinatorics of global commutativity are somewhat tamed; it is now permitted only within the context of subtrees. In programming language terms, this can be paralleled to a tree-shaped variable scoping strategy, where scopes are identified by their names (unary labels) and those of their ancestors, and the order of variable declaration is irrelevant, but the nestedness of embedded trees is not. From a linguistic perspective, this would be akin to a natural language that exhibits quasi-free local word order but makes heavy use of overt morphological case marking to disambiguate.

This is still not very realistic, but presents an interesting opportunity. Rather than commit to commutativity in general, we are invited to step into the cross-roads between the two logics, employing **L** as the global base while inventor-

izing commutative scrambling- and topicalization- like behaviors on the basis of structural postulates informed by dependency roles. Utilizing the now explicit boundaries of dependency domains, we can repurpose the notion of context to denote subtrees (despite being in an associative calculus!), obtaining the means to formulate rules like:

$$\frac{\Gamma[\langle\Delta, \langle\Phi\rangle^{\text{obj}}\rangle^d] \vdash A}{\Gamma[\langle\langle\Phi\rangle^{\text{obj}}, \Delta\rangle^d] \vdash A} \text{obj-top} \quad (\text{II.9})$$

which can be read as saying that an object can be preposed within its local d-labeled clause, if one such is nested within Γ^1 . In principle, this could simplify the categorial treatment of such phenomena: one can always start with a canonical derivation, e.g. one where all arguments are in their expected positions, and proceed by shuffling them around given the structural rule inventory. As a bonus, adorning these rules with non-void term rewrites would allow them to upkeep their relevance for pragmatics. Exciting (or not) as this might sound, it was only ever meant to incentivize the use of dependency modalities; it won't be something we will be pursuing presently, for we have another kind of beast to face.

Crossing Boundaries The structural rule format hinted at would be capable of dealing with the movement of phrases *within* a dependency domain, conditionally relaxing the word order constraints of L under certain dependency configurations. Yet it fails to provide any insights on how this could work in the case of structures that are misplaced not in terms of linear order, but of nestedness level. That is, beyond the standard question of word order, dependency modalities import previously invisible structural brackets that can pose a challenge when it comes to traversing *along* dependency domains – a challenge external to the grammar's implicational core. To make this clearer, let us revisit the keystone achievement of the previous chapter, namely the derivation of the object-relative clause of Figure I.22a (the phrase in question is “*that the eye may never behold*”, in case you were confident enough to skip the chapter). Conforming to our routine, let's first adapt some of the lexical assignments of Table II.1 (and add a few new ones for good measure).

The propositional constants for nouns N and noun phrases NP are unmarked, plain and boring; let's not speak of them any further. The first-order types of the determiner DET , adjectival modifier ADJ , and verbal types, ITV and TV , should by now also be familiar. The higher-order types of the adverb and the modal auxiliary, ADV and AUX , might, however, require some elucidation. Note, first, that despite seemingly divergent, the two types are identical when stripped of their modalities. This is perfectly in line with our agenda

¹In reality, we would need to mark complete clauses to remove the possibility of arbitrary shuffling and accidental overgeneration, but this can be trivially accomplished by boxing the phrasal end-result, e.g. $\Diamond_{su}NP \setminus \Box_{cl}S$ for an intransitive, where cl would mark a complete clause and assume the role of d in (II.9).

eye, city	::	N	
walls, twilight	::	NP	
high, sterile	::	ADJ/	$:= \square_{\text{mod}}(\text{NP}/\text{NP})$
a, the	::	DET	$:= \square_{\text{det}}(\text{NP}/\text{N})$
reigned	::	ITV	$:= \diamond_{\text{su}}\text{NP}\backslash\text{S}$
behold	::	TV	$:= \text{ITV}/\diamond_{\text{obj}}\text{NP}$
never	::	ADV/	$:= \square_{\text{mod}}(\text{ITV}/\text{ITV})$
may	::	AUX	$:= \diamond_{\text{vc}}\text{ITV}/\text{ITV}$

Table II.1: Dependency-enhanced lovecraftian lexicon.

of revealing previously coalescent diversity. The negation “*never*” functions like an adverbial adjunct: it is an endomorphism of an intransitive phrase that marks itself as a modifier in the process. The modal auxiliary “*may*”, on the other hand, heads its local phrase by assigning to an intransitive dependent the role of a verbal complement, *vc*.

Next, we need to turn our attention to the relativizer – let’s first address the missing dependencies:

$$\square_{\text{mod}}(\text{NP}\backslash\text{NP})/\diamond_{\text{body}}(\text{S}/\diamond_{\text{obj}}\text{NP}) \quad (\text{II.10})$$

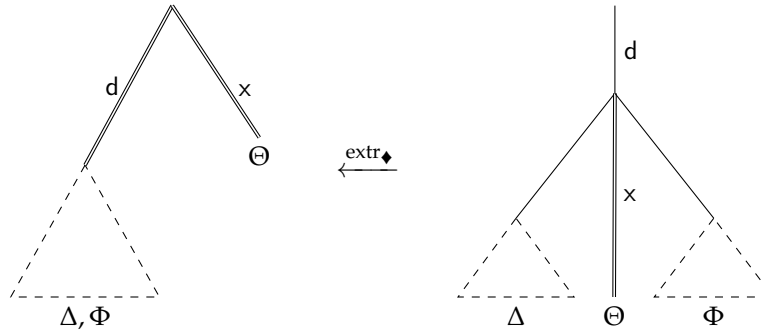
The functor now states the following: it requires first a relative clause body, namely a sentence missing an object-marked noun phrase in its rightmost border, in order to produce a postnominal adjectival phrase, i.e. a modifying adjunct.

Naively, we would assume that the absence of binary bracketing structure in **L** would allow us direct access to the gap within the relative clause body, counteracting the need for the control modalities of (I.35). But the gap is still enclosed, except this time under layers of impenetrable dependency domains! Seems like we need to reinstate control – both kinds of modalities must be employed in tandem; although in truth, the two do not really constitute distinct kinds per se: they only differ insofar as their linguistic purposes do. Nevertheless, it might be handy to make a notational distinction between them, just for the sake of reading comprehension. From now on, we will use filled symbols to denote control modalities (i.e. ones whose purposes are confined to rebracketing and movement), and white symbols to denote dependency modalities (i.e. ones whose purpose is linguistic annotation, and the brackets of which we expect to see in the antecedent of the proof’s end yield). Even when having a single control pair, assigning it an explicit label *x* is still useful, since it will allow us to tell its brackets apart from the rest. In this regime, our relativizer’s type assignment becomes:

$$\text{that} :: \text{REL}_0 := \square_{\text{mod}}(\text{NP}\backslash\text{NP})/\diamond_{\text{body}}(\text{S}/\blacklozenge_{\text{x}}\blacksquare_{\text{x}}\diamond_{\text{obj}}\text{NP}) \quad (\text{II.11})$$

$$\frac{\Gamma[\langle\Delta, \langle\Theta\rangle^x, \Phi\rangle^d] \vdash A}{\Gamma[\langle\Delta, \Phi\rangle^d, \langle\Theta\rangle^x] \vdash A} \text{extr}_\blacklozenge$$

(a) Controlled extraction rule.



(b) Corresponding tree transformation.

Figure II.5: Controlled extraction in the dependency-bracketed setting.

which is faithful to both (II.10) and (I.35), as it conveys that the missing object is now *movable*.

Mobility, however, also means something different now. Taking inspiration from the established structural vocabulary (see Figure I.19 if you need to jog your memory), we need to concoct a novel structural rule: one that allows the *extraction* of a nested substructure under the appropriate bracketing conditions. The magical concoction is presented in Figure II.5a. Within arbitrary context Γ , it looks for a d -labelled unary tree enclosing a x -labelled unary tree Θ wrapped by sequences Δ to the left and Φ to the right. There, it allows us to pull the Δ out, casting the outermost unary tree into a binary sequence and assigning d to the concatenation of Δ and Φ alone. If this makes little sense, see Figure II.5b for a visual rendition. If still unclear, move your mental cursor four sentences back (this one included) and try again.

Equipped with this missing bit of alchemical understanding, we're at long last able to produce analyses for some less contrived linguistic examples; Figure II.6 presents the dependency-enhanced derivation we set out to deliver. Before we move on, though, some important observations are in order. First, the presentation makes an implicit quantification over the outer label, d – if we were to be really pedantic, we'd need a unique instantiation of that rule for each dependency (but not control!) modal label in the logic. On a less bureaucratic note, the contextual formulation of the rule carries the usual problem of complicating syntactic equality checking. In practice, we can employ a strictly localized version, i.e. one where Γ is an empty context (just a hole),

$$\begin{array}{c}
\frac{\frac{\text{never} : \text{ADV}_/ \quad \text{lex}}{\langle \text{never} \rangle^{\text{mod}} \vdash \text{ITV}/\text{ITV}} \quad \frac{\frac{\text{behold} : \text{TV} \quad \text{lex}}{\langle \text{behold} \rangle^{\text{mod}} \vdash \text{ITV}/\text{ITV}} \quad \frac{\frac{x_i : \blacksquare_x \diamond_{\text{obj}} \text{NP} \quad \text{id}}{\langle x_i \rangle^x \vdash \diamond_{\text{obj}} \text{NP}} \quad \blacksquare_x E}{\text{behold}, \langle x_i \rangle^x \vdash \text{ITV}} /E}{\langle \text{never} \rangle^{\text{mod}}, \langle x_i \rangle^x \vdash \text{ITV}} /E \\
\vdots \\
\frac{\frac{\text{may} : \text{AUX} \quad \text{lex}}{\langle \text{may} \rangle^{\text{mod}} \vdash \text{ITV}/\text{ITV}} \quad \frac{\frac{\langle \text{never} \rangle^{\text{mod}}, \langle x_i \rangle^x \vdash \text{ITV}}{\langle \text{never} \rangle^{\text{mod}}, \text{behold}, \langle x_i \rangle^x \vdash \diamond_{\text{vc}} \text{ITV}} \quad \diamond_{\text{vc}} I}{\langle \text{never} \rangle^{\text{mod}}, \text{behold}^{\text{vc}}, \langle x_i \rangle^x \vdash \diamond_{\text{vc}} \text{ITV}} \text{extr}_\blacklozenge}{\langle \text{may} \rangle^{\text{mod}}, \langle \text{never} \rangle^{\text{mod}}, \text{behold}^{\text{vc}}, \langle x_i \rangle^x \vdash \text{ITV}} /E \\
\frac{\frac{\langle \langle \text{the} \rangle^{\text{det}}, \text{eye} \rangle^{\text{su}} \vdash \diamond_{\text{su}} \text{NP} \quad \text{lex}}{\langle \langle \text{the} \rangle^{\text{det}}, \text{eye} \rangle^{\text{su}}, \text{may}, \langle \text{never} \rangle^{\text{mod}}, \text{behold}^{\text{vc}}, \langle x_i \rangle^x \vdash \text{ITV}} \quad \backslash E}{\langle \langle \text{the} \rangle^{\text{det}}, \text{eye} \rangle^{\text{su}}, \text{may}, \langle \text{never} \rangle^{\text{mod}}, \text{behold}^{\text{vc}}, \langle x_i \rangle^x \vdash \text{S}} \quad \frac{\frac{x_j : \blacklozenge_x \blacksquare_x \diamond_{\text{obj}} \text{NP} \quad \text{id}}{\langle x_j \rangle^x \vdash \diamond_{\text{obj}} \text{NP}} \quad \blacklozenge_x E}{\langle \langle \text{the} \rangle^{\text{det}}, \text{eye} \rangle^{\text{su}}, \text{may}, \langle \text{never} \rangle^{\text{mod}}, \text{behold}^{\text{vc}}, x_j \vdash \text{S}} \quad \blacklozenge_x E \\
\frac{\frac{\text{that} : \text{REL}_o \quad \text{lex}}{\langle \langle \langle \text{the} \rangle^{\text{det}}, \text{eye} \rangle^{\text{su}}, \text{may}, \langle \text{never} \rangle^{\text{mod}}, \text{behold}^{\text{vc}} \rangle^{\text{body}} \vdash \diamond_{\text{body}} (S/\diamond_{\text{obj}} \text{NP})} \quad \frac{\langle \langle \text{the} \rangle^{\text{det}}, \text{eye} \rangle^{\text{su}}, \text{may}, \langle \text{never} \rangle^{\text{mod}}, \text{behold}^{\text{vc}}, x_j \vdash \text{S}}{\langle \langle \text{the} \rangle^{\text{det}}, \text{eye} \rangle^{\text{su}}, \text{may}, \langle \text{never} \rangle^{\text{mod}}, \text{behold}^{\text{vc}} \rangle^{\text{body}} \vdash S/\diamond_{\text{obj}} \text{NP}} /I}{\langle \langle \langle \text{the} \rangle^{\text{det}}, \text{eye} \rangle^{\text{su}}, \text{may}, \langle \text{never} \rangle^{\text{mod}}, \text{behold}^{\text{vc}} \rangle^{\text{body}} \vdash \diamond_{\text{body}} (S/\diamond_{\text{obj}} \text{NP})} \quad \diamond_{\text{body}} I \\
\frac{}{\text{that}, \langle \langle \langle \text{the} \rangle^{\text{det}}, \text{eye} \rangle^{\text{su}}, \text{may}, \langle \text{never} \rangle^{\text{mod}}, \text{behold}^{\text{vc}} \rangle^{\text{body}} \vdash \square_{\text{mod}} (\text{NP} \backslash \text{NP})} /E
\end{array}$$

Figure II.6: Dependency-enhanced adaptation of the proof of Figure I.22a, exemplifying the interaction between dependency and control modalities.

which means that extractions need to be preemptively applied immediately after every bracketing operation, rather than deferred to the future. Finally, it is important to remember that the rule is supplementary to (and *not* a substitute for) controlled associativity and/or commutativity: altering the linear order of substructures in **L**, or their binary brackets in **NL**, still calls for different structural rules with possibly different control modalities that will need to coexist with the $\text{extr}_\blacklozenge$ rule for structures to find their intended positions.

Ever higher order The example just inspected hides a crucial wisdom: variable abstraction applies to variables – *not* complex structures thereof. Control modalities may impose transient bracketing structure upon hypotheses, temporarily hindering their abstraction, but it is always retroactively redacted with the $\diamond E$ rule (after all the necessary structural operations have taken effect). The bracketing imposed by dependency modalities, however, is built to last, posing a potential roadblock to hypothetical reasoning and higher-order types.

Hypothetical complements are relatively easy to tackle: they just come packed with their diamonds at variable instantiation time. Excluding the presence of the irrelevant control box, this is exactly the strategy followed in the example under scrutiny (see the id rule instantiating x_i in Figure II.6). Upon closer inspection, we can verify that this is in fact just the η normalized version of hypothesizing a plain type, assigning it the desired dependency brackets via the $\diamond I$ rule, and then performing a substitution of the bracketed variable for a logical equivalent (wrapped under the necessary control brackets) via the $\diamond E$ rule – consult Figure II.7 and contrast with Figure I.16b if unconvinced.

Hypothetical adjuncts are less forgiving. A hypothesized adjunct will seek

$$\frac{\frac{\overline{x_k : \text{NP}} \text{ id}}{\langle x_k \rangle^{\text{obj}} \vdash \diamond_{\text{obj}} \text{NP}} \diamond_{\text{obj}} I \quad \frac{\frac{\overline{x_i : \blacksquare_x \diamond_{\text{obj}} \text{NP}} \text{ id}}{\langle x_i \rangle^x \vdash \diamond_{\text{obj}} \text{NP}} \blacksquare_x E}{\langle x_i \rangle^x \vdash \diamond_{\text{obj}} \text{NP}} \diamond_{\text{obj}} E \quad \eta \equiv \frac{\overline{x_i : \blacksquare_x \diamond_{\text{obj}} \text{NP}} \text{ id}}{\langle x_i \rangle^x \vdash \diamond_{\text{obj}} \text{NP}} \blacksquare_x E$$

Figure II.7: η long form of the \diamond_{obj} connective of hypothesis x_i in Figure II.6.

to apply itself to some phrasal head, which is impossible unless it first drops its box. But in dropping its box, it becomes enclosed in structural dependency brackets that prohibit its eventual abstraction. We will need to once more resort the $\diamond E$ rule to remove \blacksquare , except this time it will be the interior combination of Figure I.17d that we are invoking, not subject to η contraction. To see this in action, let us once more consider an excerpt from our go-to source: “a city of high walls where sterile twilight reigned”¹. The relative adverb “where” heads yet another a relative clause, “where sterile twilight reigned”, acting as a postnominal modifier to the noun phrase “a city of high walls”. The relative clause differs to the ones so far inspected, in missing from its embedded subordinate “sterile twilight reigned $_$ ” not a complement, but an adjunct: a gap-equivalent to the lexical “there” of Figure I.20e, except dependency-enhanced, i.e. $\square_{\text{mod}}(\text{ITV} \setminus \text{ITV})$ in shorthand. As Figure II.8a illustrates, the need for bracket erasure necessitates prefixing the gap type with a residual diamond, steering us towards our first ever fourth order type:

$$\text{where} :: \text{REL}_{\text{loc}} := \square_{\text{mod}}(\text{NP} \setminus \text{NP}) / \diamond_{\text{body}}(S / (\diamond_{\text{mod}} \square_{\text{mod}}(\text{ITV} \setminus \text{ITV}))) \quad (\text{II.12})$$

This beast of a type plays the starring role in the derivation of Figure II.8; it promises to provide a postnominal modifier, if presented with a (third order) relative clause body, that being a sentence missing to its right a diamond-marked (second order) postverbal modifier.

But why go through all this trouble of producing the dependency brackets only to then immediately cancel them out? The alternative of hypothesizing a plain functor without any dependency markings would be logically valid but grammatically suboptimal and contrary to our agenda, as it would give us no insights on what the dependency function of the gap is; the modalities stay. Another, more pressing question is that of the compatibility between hypothetical adjuncts and the control modality of the previous paragraph, i.e. how could we deal with the nested gap in e.g. “where sterile twilight *may* reign $_$ ”. Fortunately, we need not worry: our previous treatment still holds with only the most minor of adjustments. Same as before, we have to alter the typing of the gap by prepending a boundary crossing permit, the interior pair $\blacklozenge_x \blacksquare_x$. As the modal chains are no longer η contractable, the gap will manifest as three distinct variables, sequentially substituting one for another via two $\diamond E$ rules, as shown in Figure II.9. The first substitution of x_i for x_j is responsible for re-

¹H.P. Lovecraft, *Azathoth* (1938). In *Leaves* (2).

(a) Structurally freeing a hypothesized adjunct...

(b) ...to provide a gap in the higher-order argument of the relative adverb “where”.

moving the mod brackets for x ones, allowing any applications of the extr \blacklozenge rule in the telescoped subproof s . The bracketed variable will eventually be positioned at the outermost branch of the antecedent tree, at which point it can be substituted for x_k , the “true” variable of the gap. Don’t give in to despair; this is peak complexity – with these type patterns at our disposal we should be able to tackle any linguistic phenomenon we might encounter from this point on. Note, finally, that complement and adjunct gaps are not that different to one another after all: they both mimic their corresponding lexical assignment (a plain type or a boxed functor, respectively), except prepended by a diamond of whichever dependency role they will assume and (if needed) a structural extraction licenser.

Having completed our tour of dependency-decorated proofs, it is time for respite and reflection. Let us take a moment to internalize where we started from and where we are now.

Our first ambition was to provide categorial type logics with the means necessary to argue about dependency relations, ideally subsuming dependency

$$\begin{array}{c}
\frac{\frac{x_i : \square_{\text{mod}}(\text{ITV} \setminus \text{ITV})}{\langle x_i \rangle^{\text{mod}} \vdash \text{ITV} \setminus \text{ITV}} \text{id}}{\square_{\text{mod}} E} \quad \frac{\frac{x_j : \blacksquare_x \diamond_{\text{mod}} \square_{\text{mod}}(\text{ITV} \setminus \text{ITV})}{\langle x_j \rangle^x \vdash \diamond_{\text{mod}} \square_{\text{mod}}(\text{ITV} \setminus \text{ITV})} \text{id}}{\blacksquare_x E} \\
\hline
\frac{\langle x_j \rangle^x \vdash \text{ITV} \setminus \text{ITV} \quad \diamond_{\text{mod}} E}{\langle x_j \rangle^x \vdash \text{ITV} \setminus \text{ITV}} \\
\vdots \\
\frac{\langle \Gamma \rangle^d, \langle x_j \rangle^x \vdash A \quad \frac{x_k : \blacklozenge_x \blacksquare_x \diamond_{\text{mod}} \square_{\text{mod}}(\text{ITV} \setminus \text{ITV})}{\blacklozenge_x E} \text{id}}{\langle \Gamma \rangle^d, x_k \vdash A} \\
\vdots
\end{array}$$

Figure II.9: Schematic pattern for extracting a nested hypothetical adjunct.

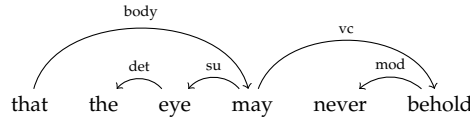
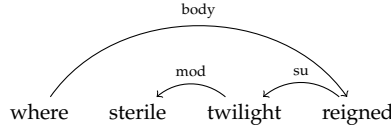
(a) Dependency tree of $\llbracket \text{that}, \langle \langle \langle \text{the} \rangle^{\text{det}}, \text{eye} \rangle^{\text{su}}, \text{may}, \langle \langle \langle \text{never} \rangle^{\text{mod}}, \text{behold} \rangle^{\text{vc}} \rangle^{\text{body}} \rrbracket$.(b) Dependency tree of $\llbracket \text{where}, \langle \langle \langle \text{sterile} \rangle^{\text{mod}}, \text{twilight} \rangle^{\text{su}}, \text{reigned} \rangle^{\text{body}} \rrbracket$.

Figure II.10: Trees extracted from the derivations of Figures II.6 (a) and II.8 (b).

trees within the logic's antecedent structures. To claim our endeavour a success, we need to actually *show* how these dependency trees can be extracted.

Let's begin with some trivial observations. In the domain of dependency annotated proofs, endsequents will contain neither variables nor any control brackets. Collapsing any notion of structure imposed by the functional core (i.e. drop constituency and word order from the builtin structural binder, since dependency trees are agnostic to those), we end up with structures made exclusively of constants (terminal nodes), multisets of structures (unordered variadic branches) and dependency enclosed structures (unary branches). This typological description can be further refined by noticing that the non-zeroary tree operators alternate in turns: there's no chaining of unary brackets (each phrase is only assigned at most one dependency role), nor a multiset of multisets (as the simplified structural binder is flat). Finally, consider that each multiset will coincide with a dependency domain, containing exactly one (unmarked) head constant (necessarily a word), some n adjuncts and some m complements.

With these in mind, we arrive at the following (modulo permutation) in-

ductive definition of a dependency induced structure:

$$\Gamma := \underbrace{\langle \Gamma_0 \rangle^{d_0} \dots \langle \Gamma_{n-1} \rangle^{d_{n-1}}}_{\text{adjuncts}} \underbrace{\langle \Gamma_n \rangle^{d_n} \dots \langle \Gamma_{n+m-1} \rangle^{d_{n+m-1}}}_{\text{complements}} \underbrace{\kappa}_{\text{head}} \quad (\text{II.13})$$

Converting such a structure to a dependency tree akin to the one of Figure II.2 is pleasantly easy: just apply the function below to it (plug an invisible “root” node and label to get things going)¹:

$$\begin{aligned} \text{deptree} &:: \text{Struct} \rightarrow \text{Head} \rightarrow \text{Deps} \rightarrow \text{Set}[\text{Arc}] \\ \text{deptree } (\langle \Gamma_0 \rangle^{d_0} \dots \langle \Gamma_{n+m-1} \rangle^{d_{n+m-1}} \kappa) \text{ root label} &= \\ &\{ \text{root} \xrightarrow{\text{label}} \kappa \} \cup \bigcup_{i=0}^{n+m-1} \{ \text{deptree } \Gamma_i \ \kappa \ d_i \} \end{aligned}$$

Figure II.10 shows the function’s yield applied to two of the section’s example derivations. Given a partition of Deps into adjuncts and complements, the conversion can trivially be extended with the bidirectional dependency arcs described in Section 7.2. Word order is also straightforward to capture, if we just keep the structural order provided by a non-commutative calculus intact. In any case, our efforts go to show that the general structure of a dependency tree can easily be captured by a dependency-enhanced type logic using the type assignment patterns discussed. Aligning the type logic to a specific flavour of a dependency grammar was never our main intention, but should be fairly easy to accomplish by reverse-engineering the annotational specifics of the target: a task for future generations.

8.4.2 Semantics

The trees of Figure II.10 leave something to be desired: by backpedaling towards dependency grammars, higher-order phenomena have been completely dismissed – wasted are all our efforts to tame and manage the bracketing structures of hypotheses. Perhaps importing residuated modalities just for the sake of cracking some dull and flat dependency relations was an overkill after all? The answer is no. Antecedent bracketing structures is but the most superficial aspect of our grammar logic’s proofs – the function argument relations of the implicational core are retained, and in fact coexist with the dependency annotations (including higher-order ones) in the logic’s term calculus, which we have been unjustly ignoring.

The far richer type and term structure of the syntactic calculus creates, in turn, ample opportunity for the passage to semantics. We are of course still presented with the cautious option of simply forgetting about dependencies, retracting to the same dependency agnostic **ILL** we did earlier. But the radical

¹For the more verbally inclined, we need to simply enter every dependency domain and establish an arc from the local head to (the head of) each of its dependents.

path would have us preserve dependency operators in the intermediate station of derivational semantics¹, in hope of them finding downstream applications later on. A first merit would be the availability of richer type assignments in the lexical semantics domain, where previously identical function types become distinguishable by virtue of their modal decorations. There, syntactic modalities can be translated to semantic operators, lifting the target signature accordingly – *monads* make for a natural choice [Kobayashi, 1997]. Furthermore, dependents are delineated and identified by their syntactic roles, allowing distinct semantic treatments, if so desired. Even in the modest setting of a non-inflated translation, modalities can help tell which of the (possibly many) combinations of semantic slots are occupied in a given construction, allowing the correct compositional recipe to be retrieved from the semantic lexicon. In the edge case, they open the possibility for a homomorphic translation that “forgets” parts of the implicational directives of the logic, relying instead on the source dependency markings for the construction of its semantic terms.

9 Æthel

With our theorycrafting over, we have in our hands a descriptive model of syntactic and semantic composition, promising to capture dependency relations while keeping both its feet set firmly in type theory. Unfortunately, it is well attested by now that “all models are wrong...” [Box, 1976]. Any promise of theoretical universality, cognitive plausibility, linguistic intrinsicness or what have you, would require some degree of handwaving and conjecturing that I am not comfortable with. What is undisputable, however, is the *omniversality* of the λ calculus [Wadler, 2015] – we’re at least certain that our modeling tools are not some ad hoc machinery designed to tackle a highly localized problem, but the one and only programming language ever worth writing. Beyond our tooling, and having *a priori* given up any ideation of truth, the only measure of success for our model is that of its utility (“...but some are useful”, [Box, 1976]). This sets up a new research imperative: we must prove the model useful!

“Oof, that’s a hard one”, you might say, and you’ll be right. Thankfully, there are two tried and tested ways to proceed. The traditional way involves fiercely advocating the idea, bashing the competition at workshops and conferences, making bold claims and generating traction at any given chance, going all out on grant applications, eventually building a cult of personality, and finally resting on your laurels as the hype becomes self-sustaining and you find yourself being cited for papers you have contributed little to other than your name; at long last, utility affirmed by popular approval. This first methodology, sometimes called the *scientific method*, is a rather involved and painstakingly long process that starts yielding profits in the long run.

¹And thus necessarily also the extraction pair $\blacklozenge_x, \blacksquare_x$.

The modern way, more befitting the temporary and precarious contract of the mobile, adaptive, multi-purpose academic laborer, is a short-term investment: simply do everything you can do yourself within the timespan you're given and hope for the best. We'll go for this one.

This choice has some methodological repercussions.

Chapter Bibliography

- J. W. Backus. The syntax and semantics of the proposed international algebraic language of the zurich acm-gamm conference. In *Proceedings of the International Conference of Information Processing UNESCO Paris June, 1959*.
- G. E. P. Box. Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799, 1976. doi: 10.1080/01621459.1976.10480949. URL <https://www.tandfonline.com/doi/abs/10.1080/01621459.1976.10480949>.
- N. Chomsky. Three models for the description of language. *IRE Transactions on information theory*, 2(3):113–124, 1956.
- M. Dalrymple. *Lexical functional grammar*. Brill, 2001.
- M.-C. de Marneffe, C. D. Manning, J. Nivre, and D. Zeman. Universal Dependencies. *Computational Linguistics*, 47(2):255–308, 07 2021. ISSN 0891-2017. doi: 10.1162/coli_a.00402. URL https://doi.org/10.1162/coli_a_00402.
- C. J. Fillmore et al. Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the origin and development of language and speech*, volume 280, pages 20–32. New York, 1976.
- G. Gazdar, E. Klein, G. K. Pullum, and I. A. Sag. *Generalized phrase structure grammar*. Harvard University Press, 1985.
- M. Haspelmath. Arguments and adjuncts as language-particular syntactic categories and as comparative concepts. *Linguistic Discovery*, 12(2):3–11, 2014.
- H. Hendriks. The logic of tune a proof-theoretic analysis of intonation. In *International Conference on Logical Aspects of Computational Linguistics*, pages 132–159. Springer, 1997.

- P. Jacobson. Phrase structure, grammatical relations, and discontinuous constituents. In *Discontinuous constituency*, pages 27–69. Brill, 1987.
- S. Kobayashi. Monad as modality. *Theoretical Computer Science*, 175(1):29–74, 1997.
- N. Kurtonina and M. Moortgat. Structural control. *Specifying syntactic structures*, pages 75–113, 1997.
- I. A. Mel’cuk. *Dependency syntax: theory and practice*. State University of New York Press, 1988.
- I. A. Mel’cuk. Levels of dependency in linguistic description: Concepts and problems. *Dependency and Valency. An International Handbook of Contemporary Research*, 1:188–229, 2003.
- M. Moortgat and G. Morrill. Heads and phrases: Type calculus for dependency and constituent structure. *Manuscript, Universiteit Utrecht*, pages 429–450, 1991.
- J. Nivre, M.-C. de Marneffe, F. Ginter, J. Hajič, C. D. Manning, S. Pyysalo, S. Schuster, F. Tyers, and D. Zeman. Universal dependencies v2: An evergrowing multilingual treebank collection. *arXiv preprint arXiv:2004.10643*, 2020.
- M. Pentus. Lambek grammars are context free. In [1993] *Proceedings Eighth Annual IEEE Symposium on Logic in Computer Science*, pages 429–433. IEEE, 1993.
- C. Pollard and I. A. Sag. *Head-driven phrase structure grammar*. University of Chicago Press, 1994.
- P. Sgall, E. Hajicová, E. Hajicová, J. Panevová, and J. Panevova. *The meaning of the sentence in its semantic and pragmatic aspects*. Springer Science & Business Media, 1986.
- D. D. Sleator and D. Temperley. Parsing english with a link grammar. *arXiv preprint cmp-lg/9508004*, 1995.
- L. Tesnière. *Elements of structural syntax*. John Benjamins Publishing Company, 2015. Original publication date 1959.
- P. Wadler. Propositions as types. *Communications of the ACM*, 58(12):75–84, 2015.