The Semantic Architecture of the Faculty of Language: Compositional Operations and Complex Lexical Representations

Guillermo Del Pinal

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY
2015

ABSTRACT

The Semantic Architecture of the Faculty of Language: Compositional Operations and Complex Lexical Representations

Guillermo Del Pinal

This dissertation defends two controversial claims about the types of operations and representations used by the Faculty of Language (FL) to compute the meaning of expressions. (I) FL is compositional in the sense that it computes the meaning of complex expressions from their structure and the meanings of their immediate constituents. Compositionality, I argue, is an innate constraint on the meaning-determining operations of FL. (II) FL uses lexical items which, in most cases, have internal semantic structure. This internal structure has extension-determining and non-extension determining components. I argue that the non-extension components are part of linguistic meaning because they play a key role in certain compositional linguistic constructions. These claims lie at the center of important debates in Philosophy of Mind, Language, and Cognitive Science. Theorists who approach the study of language using formal tools tend to be sympathetic to the view that FL is compositional. However, this position has recently come under serious attack from many directions. In contrast, thesis (II) regarding the internal complexity of lexical items has been increasingly supported by Psychologists and Cognitive Scientists, but tends to be rejected by formal Philosophers and Formal Semanticists. In this dissertation, I argue that both claims are independently plausible and, despite the consensus to the contrary, mutually reinforcing. In addition, I present an implementation of a compositional type-driven formal semantic theory that has access to the non-extension determining components of lexical items, and discuss some crucial advantages of this approach over more traditional formal semantic theories.

Contents

Acknowledgements

This dissertation benefited enormously from the help and guidance of the members of my committee, Akeel Bilgrami, Daniel Rothschild and Achille Varzi. Akeel Bilgrami brilliant knack for coming up with clear and succinct versions of my arguments lead to various substantial improvements. Achille Varzi forced me to address important foundational issues that I sometimes, unwisely, tried to avoid. Daniel Rothschild has helped me in more ways that I can single out: he helped me frame the main problems, steered me clear of countless fruitless paths, raised general and specific challenges and helped me develop potential solutions. I also discussed various parts of this dissertation with various colleagues at Columbia University. I am especially grateful to Michael Brent, Sebastian Watzl, Anubav Vaseduvian, Brian H. Kim, Marco Nathan, Nathan Brie, and Matt Moss. I would also like to thank Zoltan Szabo for some very interesting and helpful discussions of compositionality, which lead to substantial improvements to Chapters 2 and 3. Finally, I would also like to thank Louise McNally and Barbara Partee for their sharp and constructive comments in a version of Chapter 4 which was submitted for publication.

for Corina Grethel Saenz Lehnhoff

Chapter 1

General Introduction

The Faculty of Language enables us to learn languages that we can use to express and understand an unbounded number of thoughts. What is the basic structure of this faculty of the mind such that it can perform that task? According to a distinguished tradition that runs through psychology, linguistics and philosophy, the first step to answer this question is to assume that the Faculty of Language is a cognitive *computational* system. This computational approach is generally adopted in investigations of our phonetic and syntactic competence. But some prominent theorists have argued that our semantic competence, our capacity to understand the meaning of expressions, is not sufficiently *systematic* to call for a computational explanation.

This dissertation begins with a defense of the view that our semantic competence is systematic and productive in just the way we would expect if the Faculty of Language were a recursive computational system, part of whose function is to compute not only the phonetic and syntactic structure of expressions but also their meaning (Chapters 1-2). Once this move is accepted, we can ask a more specific question about the basic structure of the Faculty of Language:

What are the basic operations and representations which it uses to compute the meaning of expressions?

The main goal of this dissertation is to defend two claims about the kinds of operations and representations used by the Faculty of Language to compute the meaning of expressions:

(I) The Faculty of Language is compositional in the sense that it computes the meaning of complex expressions from their structure and the meanings of their immediate constituents. Compositionality is an innate constraint on the operations of the Faculty of Language (Chapter 3).

(II) The Faculty of Language uses lexical items which, in most cases, have internal semantic structure. This internal structure has a part which determines their extension and a part which does not determine their extension but that plays a role in certain compositions (Chapter 4-5).

(I) and (II) lie at the center of heated debates about the architecture of the Faculty of Language. My strategy for defending them flows from a general dissatisfaction with many of the current discussions: despite the occasional lip-service, theorists of language too often forget that questions about compositionally and lexical semantics are paradigmatic questions of cognitive science, which require a genuine interdisciplinary approach. For example, many theorists argue, against (I), that even if we assume that the Faculty of Language has access to non-compositional operations to compute the meaning of expressions, we can still account for the systematicity of our semantic competence. While this is true, it is not a good reason to reject (I). For the assumption that the Faculty of Language can use non-compositional operations, while locally acceptable, has unacceptable consequences for language acquisition (Chapter 3).

Similarly, many linguists and philosophers try to account for a whole range of problematic expressions without significantly enriching the formal semantic framework which we inherited from Frege and Montague. So they are committed, against (II), to the view that lexical items do not have, or at least that the compositional operations are not sensitive to, non-extension determining internal semantic structure. As a result, some of their accounts look like implausible ad hoc attempts to save the original formal semantic framework. In the meantime, cognitive psychologists are actively developing sophisticated theories about the complex structure of our lexical concepts. If we reformulate our semantic theories so that they can use the most plausible of these complex structures, we can provide more principled compositional accounts for a whole range of problematic expressions (Chapter 4-5).

When tackling questions about the basic architecture of the Faculty of Language we have to face some classic debates in the philosophy of mind and language: nativism vs empiricism, descriptivism vs essentialism, and the priority of thought vs that of language. To defend (I) is to defend a form of nativism about our linguistic competence, for (I) says that compositionality is an innate constraint on the basic semantic operations of the Faculty of Language. To defend (II) is to defend a mix between descriptivism and essentialism about lexical semantics, for (II) says that lexical terms have a dual structure with an extension-determining component and a descriptive, non-extension determining component. Finally, to defend (I) and (II) is to defend the priority of thought over language, for both claims suggest that we have substantial and domain-specific conceptual and inferential capacities before we can acquire natural languages.

Part I

On the Semantic Operations of *FL*

Chapter 2

Why is *FL* Compositional?

This chapter defends the view that the Faculty of Language is compositional, i.e., that it computes the meaning of complex expressions from the meanings of their immediate constituents and their structure. I first argue that compositionality and other competing constraints on the way in which the Faculty of Language computes the meanings of complex expressions should be understood as hypotheses about innate constraints of the Faculty of Language. I then argue that, unlike compositionality, most of the currently available non-compositional constraints predict incorrect patterns of early linguistic development. This supports the view that the Faculty of Language is compositional. More generally, this paper presents a way of framing the compositionality debate, focusing on its implications for language acquisition, that can lead to its eventual resolution, so it will hopefully also interest theorists who disagree with its main conclusion.

2.1 Introduction

The human Faculty of Language (FL)—the mental faculty which plays a central role in the acquisition and processing of natural languages—enables us to systematically produce and understand an unbounded number of novel expressions. More precisely, if 'FL_S' is the FL of an arbitrary competent speaker S of some natural language E (English, Spanish, etc.), then:

(PRODUCTIVITY) FL_S can generate *correct* interpretations (relative to E) for complex expressions which S has never encountered before. FL_S has this capacity for indefinitely many distinct complex expressions, generating a distinct meaning for an indefinite number of these expressions.

(SYSTEMATICITY) The generative capacity of FL_S is structured in the following way: if it can generate *correct* interpretations (relative to E) for complex expressions $e_1, \ldots, e_n$, it can generate correct interpretations for all other complex expressions constructed from: (i) constituents of $e_1, \ldots, e_n$ and (ii) syntactic structures employed in any of the complex expressions $e_1, \ldots, e_n$.

For a long time, it was held that the best way to explain PRODUCTIVITY and SYSTEMATICITY (henceforth '$P\&S$')[1] is to assume that FL includes a recursive computational system with a compositional semantics.[2] Roughly, a computational system is 'compositional' if (i) it contains both primitive and syntactically complex symbols, and (ii) it is constrained to determine the semantic properties of its complex symbols only from their structure and the semantic properties of their immediate constituents.

[1]Here we will assume that $P\&S$ are essentially correct. Although widely accepted, esp. in discussions of the non/compositionality of FL, $P\&S$ are not completely uncontroversial (see e.g. Pullum and Scholz [177], [147]). For an extensive defense of $P\&S$, see Del Pinal [138].

[2]The most famous version of this argument was presented by Fodor to defend the compositionality of both natural languages and thought (see Fodor and Pylyshyn [51], Fodor [42], Fodor and Lepore [48]). However, Fodor recently presented some arguments to the effect that natural languages don't seem to be compositional (Fodor [45]).

Recently, however, several prominent linguists, philosophers and cognitive scientists have criticized the view that FL is compositional, and especially the argument from $P\&S$.[3] There is some truth to these criticisms. The traditional argument for compositionally is indeed substantially incomplete. No one doubts that assuming that FL is compositional is *one* way of explaining $P\&S$; but there are now other reasonable explanations which assume that FL is *not* compositional.[4] This presents an important challenge. For the assumption that FL is compositional has shaped the way we theorize about basically every central aspect of our linguistic competence.

Despite these criticisms, I will argue that assuming that FL is compositional is still the best explanation of $P\&S$. I begin by proposing that we should frame debates about the non/compositionality of FL as debates about the fixed, innate structure of the part of FL which computes the meanings of complex expressions (§2.2-2.3). Given this framework, we can determine, for each competing explanation of $P\&S$, its broad empirical consequences for language acquisition and development. I then show that, unlike assuming that FL is compositional, the non-compositional accounts of $P\&S$ entail that, in the course of acquiring a natural language, speakers should go through certain stages of early linguistic development which, it turns out, speakers never seem to go through (§2.4-2.5). This strongly suggests that FL is compositional. In the final section I discuss some objections to this argument (§2.6).

[3] Some theorists argue on general grounds that our linguistic competence is (probably) not compositional (see Jonsson [82]; Hampton and Jonsson [67]; Johnson [81]; Travis [202]; Lahav [97]). Others argue that although it might be compositional, the general arguments usually taken to establish this are not very persuasive (see Szabo [190], [193]; Dever [37]; Pagin and Westersthl [123]; Baggio et al. [2]).

[4] Jonsson's illuminating [82], which we will discuss in detail in what follows, presents various non-compositional explanations of $P\&S$.

2.2 *FL* as a Cognitive Computational System

Following the tradition of computational psychology, we will assume that FL is a language-processing cognitive computational system. This approach has been famously advocated by Chomsky [23], [24], [25]. The view I present below is Chomskian, but in a weak sense that can be welcomed by theorists of different theoretical orientations. This approach revolves around two basic theoretical notions: 'Ψ-languages' and 'languages'. Both terms require some explication.[5]

By a 'language' I mean a set of pairs of acoustic/visual signals and meanings or interpretations which characterizes a natural language. For example, 'English' is a set consisting of certain pairs of acoustic/visual signals and meanings, e.g., (*red*, RED), (*John is happy*, HAPPY (JOHN)), (*red wall*, RED WALL), and so on. A 'language', as I am using the term, is 'extensionally defined' but *not* an external or mind-independent abstract structure of the sort Chomsky [23] argues is of no relevance to the study of FL. Specifically, 'languages' consist of the input/output pairs of representations which 'Ψ-languages' compute, i.e., they specify the main cognitive task which 'Ψ-languages' solve.[6]

Most linguistic theorists agree that, to compute languages, Ψ-languages need to carry-out *at least* three main cognitive tasks: (i) map acoustic/visual signals into expressions (phonetics), (ii) map expressions into syntactic structures (syntax), (iii) map syntactic structures into meanings or interpretations (semantics). An 'Ψ-language' then is a cognitive computational system that can generate

[5]The term 'Ψ-language' was introduced by Ludlow [103]. The following discussion is indebted to his discussion of the small but important distinction between 'Ψ-languages' and Chomsky's 'I-languages'.

[6]'English', as used here, does not exactly correspond to what we ordinarily mean by 'English'. For example, 'English', as ordinarily used, does not include the expression *the child seems sleeping*. But the meaning of this expression is arguably computed by Ψ-English, so it is part of English, as here defined (see Pietroski and Crain (forthcoming)). Another complication which we can ignore for now is that the semantics does not generate interpretations for all the outputs of syntax. A putative example of this is Chomsky's *colorless green ideas sleep furiously* (but see Camp [19] for rejection of the view that these sorts of sentences don't have a literal interpretation).

phonetic structures, syntactic structures, and semantic structures or interpretations. For example, think of 'Ψ-English' as a cognitive computational system that, given certain signals, outputs certain meanings (and vice-versa), thereby computing English. For our purposes, we can remain neutral about the 'nature' of the representations used by Ψ-languages, but in principle this framework can be paired with a Chomskian internalist view (see e.g., Pietroski [134], [135]) or with any of the externalist views more commonly adopted by philosophers and formal semantic theorists (see e.g., Ludlow [103]).

Ψ-languages, as here defined, have three basic properties, crucial to debates about the non/compositionality of FL: (i) they are idiolects, (ii) they have an unbounded generative capacity; and (iii) some of their properties are innate.

Ψ-languages are idiolects. In principle there can be as many distinct Ψ-languages and languages as there are individuals with a FL; or even more, since on the way to its mature and stable state, each FL goes through various developmental stages. However, general cognitive constraints, including some particular to FL, together with general properties of language acquisition environments, reduce the differences between individual Ψ-languages, at least to the extent required for successful communication between speakers of what we informally call a 'linguistic community'.[7] Here we are concerned with general properties of Ψ-languages, so we can ignore variations between the Ψ-languages of members of the same 'linguistic community', when compared at the same stage of development. For example, we will assume that the community of what we informally call 'English speakers' is a linguistically homogenous community, so that 'English' captures (not perfectly) the set of <acoustic/visual signal, meaning> pairs which members of this community use to communicate. Relative to this idealization, Ψ-English is the cognitive computational structure that

[7] That idiolectical conceptions of language can be wedded to successful accounts of communication has been shown by Larson and Segal [100] and Higginbotham [72].

an arbitrary competent speaker of this homogenous linguistic community uses to compute English.

Ψ-languages have an unbounded generative capacity. Ψ-languages can assign interpretations to an unbounded number of novel expressions, following the patterns specified in $P\&S$. For this reason, theorists studying languages often define or describe (fragments of) them intensionally, rather than by listing their <acoustic/visual signal, meaning> pairs. Different functions-in-intension can define the same set of pairs, the same language, in which case we call them 'extensionally equivalent'. It is sometimes said that if two or more functions-in-intension are extensionally equivalent, the claim that one is the 'correct' one doesn't make sense. This is correct if we assume that the only task of a linguistic theory is to describe or define a language; but incorrect if we assume that part of the task of a linguistic theory is to discover (properties of) Ψ-languages (Chomsky [23]; Evans [41]; Davies [35]). We can discriminate between (at least some) extensionally equivalent models on the grounds that one is, or seems to be, closer to the function actually computed by the Ψ-language's computational processes. Whether a particular piece of psychological data can be used to make this discrimination is usually controversial. Still, there is widespread agreement that relevant evidence can come from data about patterns of language loss, acquisition or revision, and from any neurological data that reveals properties about the computational capacities or structure of the mind/brain.

Some properties of Ψ-languages are innate. FL undergoes development from an initial state prior to exposure to linguistic data, through various intermediate states, to the 'mature' and stable state in which it incorporates Ψ-languages that can fully compute 'natural languages' such as English, Spanish, etc. 'Mature' Ψ-languages consist of certain semantic and syntactic rules and principles, some of which have to be acquired in the course of linguistic development. For example, speakers acquiring Ψ-English have to acquire lexical rules such as $[\![red]\!] = \textsc{red}$, and syntactic principles such as that heads precede their complements and that null subject sentences such as *is raining* are not allowed. Other rules and principles, which do not seem to be learned in the course of language development, are more plausibly seen as innate and often fixed properties of FL common to all Ψ-languages. Some candidates for innateness are the syntactic and semantic primitives, the constraints that all syntactic principles are structure-depedent and all syntactic branching is binary.

Precisely which rules and principles are innate, and which of these are unique to FL, is a matter of ongoing debate between nativists who propose a substantial base of innate and language-specific structures (e.g. Berwick et al. [158], Baker [3]), empiricists who propose a minimal base of innate and no language-specific structures (e.g. Elman et al. [40], Pullum and Scholz [146], Perfors et al. [132]), and theorists who defend mixed or intermediate positions (e.g., Xu [212]). But what is not controversial—and what we will assume—is that FL has *some* innate structure, common to all Ψ-languages, except those affected by unusual genetic or developmental conditions. If we drop this assumption, it is impossible to explain how FL can represent and interact with linguistic data to begin to develop into a mature Ψ-language.[8]

To close this brief presentation of our operating conception of FL, I should

[8]There is a heated debate about the learning mechanisms used in language acquisition. Linguists tend to emphasize that acquiring an Ψ-language, as Szabo [191] puts it, 'requires little or no explicit instruction', follows a certain developmental sequence, and 'tends to yield a remarkably uniform level of competence'. This suggests that, for the most part, acquiring an Ψ-language is quite unlike learning social conventions (Szabo [191]) or scientific theories (Chomsky [28]). Unlike knowledge of a scientific theory, speakers are not conscious of—and if prompt cannot state—most of the rules and principles which they 'acquire' as part of theirΨ-languages. When 'acquiring' most rules and principles, speakers do not seem to make the sorts of mistakes they would make if they were constantly testing reasonable but incorrect 'hypotheses' against language data. However, some developmental psychologists disagree with this picture and argue that language acquisition essentially relies on our innate, but domain general, 'science forming' mechanisms (e.g., Gopnik [61]). As will become clear, the argument we present for the compositionally of FL does not *assume* any one of these competing views on the nature of the language learning mechanisms.

add that I will *not* assume that FL is an informationally encapsulated cognitive module. This is important because some advocates of compositionality assume that FL is informationally encapsulated, and then appeal to this property to defend its compositionality (see e.g. Borg [14] and Larson and Segal [100]). Critics are justifiably skeptical of this way of defending compositionality (Robbins [168]; Jonsson [82]: chap. 6). Ψ-languages exhibit *some* degree of modularity—they are domain specific, have mandatory operations which are (for the most part) fast, have limited central accessibility, and characteristic patterns of breakdown and development. If to account for these features of Ψ-languages we assume their (almost) total informational encapsulation, we move towards the view that they are likely compositional. For most non-compositional accounts require that, to determine the meanings of complex expressions, the semantics have access to some subset of non-linguistic information. However, as most theorists rightly point out, modularity comes in degrees. Even if we can hold that FL is modular to some non-trivial degree, there is currently no good reason to assume that it is informationally encapsulated to the degree that would be required to make it incompatible with most reasonable non-compositional accounts (Robbins [168]).

2.3 Compositionality as a functional constraint of *FL*

In our framework, to say that FL is compositional is to say that there is a particular constraint on the way in which it generates the meanings of complex expressions: the algorithms which generate semantic interpretations for complex expressions can only use semantic information provided by their immediate constituents and information about their combinatorial structure. This does not tell us, for a particular type of complex expression (e.g., $[_{NP} A\ N]$), what particular algorithm determines its meaning; it only tells us that the algorithm computes a compositional function. We will call general semantic constraints (such as compositionality and other competing constraints) which range over all types of complex expressions, 'meaning-determination constraints' (MDCs). MDCs should be distinguished from particular 'semantic rules' (SRs) which determine the meanings of *particular types* of complex expressions (e.g., $[\![[_{NP}A\ N]]\!] = f_{NP}([\![A]\!], [\![N]\!])$). MDCs range over and constraint the general form of particular SRs.

This distinction between MDCs and SRs raises an important question which has been neither sufficiently nor adequately raised in the literature. Should we think of compositionality as a principle that we learn when we acquire some Ψ-language (so that we could have acquired a different MDC)? Or should we think of it more like an innate and fixed property of FL, hence present in all Ψ-languages? The latter option is closer to the way in which I suggest we should understand compositionality and other competing, non-compositional MDCs. Specifically, we should understand MDCs as constraints on what, following Pylyshyn [156], we'll call the 'functional architecture' of the semantics of FL.

We can understand the notion of 'functional architecture' by analogy to the way in which it is used in computer science (Dawson [36]; Pylyshyn [156], [157]). The functional architecture of a computational system M is the fundamental programming language used to write the algorithms that M computes. This programming language is fundamental in the sense that its primitive operations or functions must be built into the (possibly virtual) machine M. Similarly, the functional architecture of a cognitive computational system C (e.g., an Ψ-language) is something like the basic set of representations and operations available to C. The particular rules and algorithms which can be represented and computed by C are those which can be defined in terms of Cs basic programming language. So if we specify Cs functional architecture we thereby implicitly specify Cs cognitive capacity, i.e., the set of cognitive rules and algorithms which can be represented and processed by C. A 'functional constraint' on C is a way of (partially) specifying Cs functional architecture, hence (implicitly) Cs cognitive capacity.

To further clarify this notion, consider a rule which is clearly *not* a functional constraint, e.g., the lexical semantic rule $[\![red]\!] = \text{RED}$. This rule might be part of some Ψ-languages—e.g., Ψ-English—but it is obviously not a functional constraint. Firstly, particular lexical rules are optional features of Ψ-languages. In certain conditions, FL can acquire the rule $[\![red]\!] = \text{RED}$; but in other conditions, FL can acquire different rules for $[\![red]\!]$, e.g., $[\![red]\!] = \text{BLUE}$, or $[\![red]\!] = \text{ANGRY}$. Secondly, the processes of learning lexical rules such as $[\![red]\!] = \text{RED}$ can be usefully understood as a rational learning process in which different hypotheses about the meaning of *red* (e.g., $[\![red]\!] = \text{RED}$, $[\![red]\!] = \text{MAROON}$, $[\![red]\!] = \text{DARK ORANGE}$) can be tested and rejected or accepted. Thirdly, acquired in roughly this way, FL must be capable of explicitly representing the contents of lexical rules.

Functional constraints are fundamentally unlike such *optional* and rationally acquired cognitive rules. Functional constraints specify the *fixed* representational and computational capacity of a cognitive system, i.e., the basic representations and operations used by the system. Hence functional constraints (i) are not acquired via cognitive processes (esp., via processes that can be properly modeled as inferential, or more broadly rational, responses to information), and (ii) we need not assume that they are explicitly represented by cognitive systems. A good example of a functional constraint is the putative informational encapsulation of some modular cognitive systems. A module M is not informationally encapsulated because M learned a rule which specifies that, in its computational operations, M should not use information from other cognitive systems. Rather, M's informational encapsulation is explained by a constraint on its fixed functional architecture: M is implemented in a way that blocks operations of information exchange with other cognitive systems. M's inter-modular information restriction is a constraint on M's cognitive capacity; it is not something M can cognitively learn or alter.

MDCs are more like constraints on the exchange of information between some cognitive modules than like optional lexical rules such as $[\![red]\!] = \text{RED}$. At no point in language development does it seem that speakers are trying or have to learn a general rule or principle which, like compositionality, structurally constrains the kind of information which their Ψ-languages can use to determine the meanings of different types of complex expressions (see §2.6.6 below). Furthermore, it does not seem possible to specify a counterfactual acquisition scenario in which speakers would acquire, for cognitive/rational reasons, a different MDC. This suggests that the claim that FL or Ψ-languages satisfy some particular MDC should be understood as a proposal about how to constraint the functional architecture of the semantics of FL.

Taking MDCs as functional constraints ties each competing proposal to a set of characteristic consequences for language acquisition. The reason for this should be clear. To specify the functional architecture of a cognitive system is to implicitly specify the system's cognitive capacity, i.e., the set of cognitive rules and algorithms which the system can represent and process. To hold that FL is constrained by a compositional MDC entails that FL is not *cognitively capable* of instantiating, hence of acquiring, Ψ-languages with non-compositional SRs. In contrast, to hold that FL is constrained by a non-compositional MDC entails that FL is cognitively capable of acquiring Ψ-languages with compositional *and* non-compositional SRs. These differences in the SRs they can 'see' determine the consequences for acquisition of the competing MDCs.

For example, assume that the FL_1 of speaker S_1 has compositional MDC M_1, that the FL_2 of speaker S_2 has non-compositional MDC M_2, and that S_1 and S_2 are beginning the process of acquiring Ψ-language L, compatible with both M_1 and M_2. S_2 has to consider a hypothesis space that includes both compositional and noncompositional SRs. This difference should be manifest in at least slightly different patterns of linguistic development (e.g., in the sorts of mistakes they could make), even if S_1 and S_2 eventually converge at L. Hence even if both M_1 and M_2 can explain $P\&S$ (in the sense that all Ψ-languages compatible with either MDC satisfy $P\&S$) and are compatible with L (in the sense that under certain conditions both speakers could eventually acquire L), we can still prefer one MDC if it predicts patters of development which better fit or explain the course of actual linguistic development. In what follows, I will argue that this is the reason why compositionality is more plausible than the non-compositional MDCs.

2.4 Compositionality as a MDC

This section presents the notion of compositionality I will defend. The next section present the non-compositional MDCs. To clearly state and compare the competing MDCs, I will use the following terminology:

- A 'lexical rule' is an expression of the form '$[\![x]\!] = m$', where x ranges over particular expressions, e.g., '$[\![dog]\!]$ = DOG' and '$[\![brown\ dog]\!]$ = BROWN DOG'.

- A 'semantic rule' (SR) is an expression of the form '$[\![[_Z X\ Y]]\!] = m$', where '$[_Z X\ Y]$' stands for any arbitrary *type* of syntactic structure (e.g., $[_{NP} A\ N]$), including the most general one, where Z is any branching node with {X, Y} as its immediate constituents.

Compositionality, interpreted as a MDC, amounts to the following constraint:

(CO) If L is an Ψ-language which FL can represent, then:

1. L cannot use lexical rules to determine the meanings of complex expressions.
2. Each SR in L is of the form '$[\![[_Z X\ Y]]\!] = f_Z([\![X]\!], [\![Y]\!])$', were '$f_Z$' is a humanly computable function defined on the set of meanings.

Condition 1 of CO excludes all Ψ-languages which assign meanings to syntactically complex expressions in a list-like way. To see why this condition should be part of any adequate MDC, including non-compositional ones, consider the consequences for acquisition of dropping it. FLs with MDCs without condition 1 would have to consider, for any particular complex expression, if its meaning is determined through a lexical rule. For example, assume S knows, $[\![brown]\!]$, $[\![dog]\!]$, and $[\![[_{NP} A\ N]]\!] = f_{NP}([\![A]\!], [\![N]\!])$. If S's FL has a MDC without condition 1, S would still have to consider (without triggering from any special feature of the learning data, e.g., repetition) whether $[\![[_{NP}[_A brown][_N dog]]\!]$ is not given by any of a set of lexical rules, which yield not $[\![[_{NP}[_A brown][_N dog]]\!]$ = BROWN DOG, but rather $[\![[_{NP}[_A brown][_N dog]]\!]$ = ANGRY BROWN DOG or LAME BROWN DOG or any other direct meaning assignment consistent with the learning data. This cognitive 'flexibility' substantially complicates language acquisition and predicts patterns of linguistic development which we never find.[9]

[9] As will become clear, condition 1 is *not* the point of contention between CO and the non-compositional MDCs; but one might still object to it on the grounds that it seems incompatible with idioms. Explaining idioms is everyone's problem, but some influential recent accounts are consistent with and even support condition 1. Idioms are ambiguous expressions: they have a literal phrasal and a idiomatic meaning. The literal meaning of e.g. *kick the bucket* is KICK THE BUCKET, and its idiomatic meaning is TO DIE. There is substantial evidence that the literal meaning of idioms is automatically processed in parallel with their idiomatic meaning (Tabossi [196], Glucksberg [56]). This suggests, as predicted by condition 1, that I-languages are constrained to determine the literal meaning of complex expressions,

Condition 2 guarantees that all Ψ-languages compatible with *CO* assign meanings to complex expressions through *SR*s that have access only to the (syntactic) mode of composition of expressions and the meanings of their immediate constituents. If we hold *CO*, conditions 1 and 2 hold in general—i.e., of all Ψ-languages which *FL* is cognitively capable of instantiating. This ensures that *CO* provides an adequate structural explanation of *P&S*.

To further clarify *CO* let us consider its relation to the syntax and pragmatics interfaces, beginning with the former. *CO* is compatible with a "strongish" compositional view, in the sense of Jacobson [79]. According to this view, the syntax and the semantics work in tandem: there is no intermediate level (such as the *LF* of early transformational grammars) that is first built from surface structures—using syntactic operations that have no corresponding semantic operations—and serves as input to the semantics. All the *CO* compatible solutions to problematic expressions that I discuss later in the paper respect strongish compositionally, but my arguments for *CO* are compatible with weaker views on Jacobson's scale, e.g., a view according to which there are some syntactic operations (with no corresponding semantic operations) that create *LF* structures from surface structures. As will become clear, the debate about whether we can hold on to strongish compositionality (which depends

even idiomatic ones, via *SR*s. To explain how the idiomatic meaning of idioms is determined, we have to make a distinction between two types of idioms, based on their syntactic flexibility. Some idioms are syntactically inflexible (except for negation) and behave like words, e.g., *by and large*. There is evidence that the idiomatic meaning of syntactically inflexible idioms is computed directly, as syntactically simple expressions (Glucksberg [56]); so their meaning is determined, consistently with condition 1, in a list-like way via lexical rules. Other idioms are syntactically flexible and behave like phrases, e.g., *spill the beans* can be used as *the terrorist didn't spill a single bean during the interrogation*, or as *John was weak, he spilled all the beans during the interrogation*. There is also evidence that the idiomatic meaning of syntactically flexible idioms is computed in the ordinary compositional way, except that their simple parts are polysemous or ambiguous and, in the idiomatic context, take on the relevant idiomatic meaning (McGlone, Glucksberg, and Cacciari [111]). The idea is that most mature English speakers know, e.g., not only that *spill* means FALL FROM CONTAINER and *beans* means EDIBLE LEGUMES, but also that in some special (idiomatic) contexts they can also mean, respectively, REVEAL and SECRETS. The assumption that parts of the idiomatic phrase correspond to parts of the idiomatic meaning explains why flexible idioms can be internally modified (Nunberg et al. [119]), as in *the investigator spilled some of the beans* or *the suspect quickly spilled all the beans*, with predictable and systematic changes to the meaning of the idiomatic phrase.

on issues like whether we need quantifier/auxiliary raising rules), is independent of the debate between compositional and non-compositional MDCs as I frame it here, and would still arise, mutatis mutandis, even if we hold a non-compositional MDC.

Another important issue at the syntax interface concerns the relation between syntactic rules and types of SRs. Some Montague-style theories use particular phrase-structure rules such as $S \rightarrow NP\ VP$ and pair them with construction-specific compositional SR such as (2.1):

(2.1) $[\![[_S NP\ VP]]\!] = f_S([\![NP]\!], [\![VP]\!])$

where 'f_S' is a function which given $[\![NP]\!]$ and $[\![VP]\!]$, outputs $[\![S]\!]$. As stated, CO is compatible with those views; but we will make a stronger assumption, namely, that CO requires general, not construction-specific, SRs. So we will assume that, in the formulation of CO above, Z stands for any branching node with $\{X, Y\}$ as its immediate constituents. This is in any case how we would have to interpret CO if it is paired with a syntactic theory, such as Minimalism, that does *not* have category-specific phrase-structure rules. A famous theory along these lines is the type-driven theory presented in Heim and Kratzer [70]. Type-driven theories do not require category-specific syntactic phrase-structure rules. Heim and Kratzer assume that the syntax delivers to the semantics bare-phrase structures. Translated into our terminology, this means that the semantics sees only the most general type of syntactic structure, a branching node and its immediate constituents. Given this assumption about what the syntax delivers to the semantics, CO entails that the SRs have to be general, ranging over all (types of) complex expressions (including NPs, VPs, Ss, etc.). An example of a general SRs is Functional Application (FA):

(FA) If α is a branching node, $\{\beta, \gamma\}$ is the set of α's daughters, and $[\![\beta]\!]$ is a function whose domain contains $[\![\gamma]\!]$, then $[\![\alpha]\!] = [\![\beta]\!]([\![\gamma]\!])$

However, CO is compatible with various accounts of how the compositional operations work, i.e., of the nature of the general SRs. According to recent Neo-Davidsonian accounts, composition is a uniform operation such as predicate conjunction over monadic concepts (Pietroski [137]). Other linguists, closer to Montague's original framework, use rules such as FA, predicate modification, and various type-shifting rules (Jacobson [79], Heim and Kratzer [70]). The important point, for our purposes, is that we could state the whole dialectic between compositional and non-compositional MDCs by assuming either view.[10]

Consider now the interface with pragmatics. CO is compatible with at least two kinds of context-sensitivity. Firstly, CO allows the meaning of some, most, or all lexical items to be characters. We can represent this by saying that, for any expression e:

- $[\![e]\!]_c = f_e(c)$

where f_e is the character of e and $f_e(c)$ is the occasion meaning of e in c. If e has no free parameters, then for all c's, $f_e(c) = m$, where m is the standing meaning of e. Secondly, CO allows SRs to take the modulated (instead of the standing or occasion) meanings of the immediate constituents of complex expressions. Following Recanati [163], we can represent the modulated meanings of an expression e, $[\![e]\!]_{M,c}$, as follows:

- $[\![e]\!]_{M,c} = mod(e,c)([\![e]\!]_c)$

mod takes as an argument an expression e and context c in which e occurs and returns as value the modulation function $f_{M,e}$, which takes $[\![e]\!]_c$ and returns the meaning that is salient/relevant/appropriate for e in c.

[10] This is not to deny, of course, that the outcome of this debate affects the interpretation of CO. For example, neo-Davidsonians often avoid type-shifting rules by positing covert syntactic elements (see e.g. Pietroski's account of proper names in [137]); so assuming that composition is predicate conjunction might entail that we abandon strongish compositionality.

There are two main ways of implementing the context-sensitive *mod* function to get general SRs that determine the meanings of complex expressions in terms of the modulated instead of the standing or occasion meanings of their constituents. On an unconstrained view, *mod* is generalized to apply at every level of interpretation. On a constrained view, which is the one we will adopt here, *mod* applies only on lexical items. To illustrate, let us implement this constrained version of *mod* in a type-driven framework. Focusing again on FA, our interpretation should be formulated as follows (assume for brevity that all non-branching nodes are terminal nodes):

(TN_M) If α is a terminal node, then $[\![\alpha]\!]_{M,c} = mod(\alpha, c)([\![\alpha]\!]_c)$, where $[\![\alpha]\!]_c$ is specified in the lexicon.

(FA_M) If α is a branching node, $\{\beta, \gamma\}$ is the set of α's daughters, and $[\![\beta]\!]_{M,c_1}$ is a function whose domain contains $[\![\gamma]\!]_{M,c_2}$, then $[\![\alpha]\!]_{M,c} = [\![\beta]\!]_{M,c_1}([\![\gamma]\!]_{M,c_2})$

On this account, *mod* does not operate on the outputs of FA_M (or other rules for interpreting the meaning of complex expressions), but only on terminal nodes/lexical items. This allows a constrained form of meaning modulation. Since the compositional step (i.e., the determination of the meaning of complex expressions), in this sort of framework, corresponds to the FA_M rule, we can say that meaning modulation is pre-compositional.

Pragmatic processes also modify the outputs of the semantics, but there is no good reason to model post-compositional pragmatic processes as part of FL. Many linguists now think that syntactic/semantic computations work in phases that are sent off for pragmatic interpretation *before* full sentences or clauses are processed by FL. In Minimalist theories, the main phases are vPs and CPs, but due to the 'left edge condition' (Chomsky [26, 27]), the phases that are sent out for pragmatic processing are more fine grained (Cook and Newson [31], Radford [159]). Theorists who adopt Categorical Grammars also usually assume that the outputs to pragmatics are sub-sentential phrases (Jacobson [79]). If interpretation proceeds in such phases, which are inputs to (primary) pragmatic processes, then there is no reason why we should incorporate into the semantics a generalized version of *mod*, i.e., a function which modulates both the inputs and outputs of the compositional operations. Such output modulations would be redundant; indeed, in actual case studies (as in the CO compatible accounts we present below), most of the modulation operations operate on lexical items.

To conclude, let me emphasize the most important consequences, for this discussion, of interpreting CO as excluding construction-specific SRs. Firstly, assuming that FL can only represent general SRs has the advantage of substantially diminishing the amount of rules which we have to assume speakers either acquire or innately possess. Secondly, that assumption also coheres nicely with an important current trend in generative linguistics, namely, to generalize or eliminate phrase-structure rules (see Chomsky [23]; Heim and Kratzer [70]). But most importantly, assuming that FL cannot represent construction-specific SRs has distinctive empirical consequences for language acquisition and development. For example, Heim and Kratzer's type-driven theory entails that speakers need not acquire or innately posses a construction-specific SR like (1) for each type of syntactic construction. If speakers know, about a complex expression, (i) the meanings of its parts and (ii) its structure, this theory predicts that they should have the linguistic competence to adequately determine its meaning. In other words, once they know (i) and (ii), there is no space for speakers to make a mistake that leads to an incorrect understanding of a complex expression. Non-compositional theories make different predictions about the sorts of mistakes speakers can make. For as we will now see, each non-compositional MDC has to assume that speakers can acquire construction-specific SRs. This entails that, despite knowing (i) and (ii), speakers could, early in development, systematically assign incorrect meanings to tokens of certain types of complex expressions, for they could assign an incorrect construction-specific SR to any type of complex expression.

2.5 CO vs Non-compositional MDCs

The claim we will defend is that assuming a MDC approximately like CO is currently the best explanation of $P\&S$. CO seems correct insofar as it requires that the meaning of complex expressions be determined via SRs. But we might suspect that CO is too restrictive insofar as it requires that SRs determine the meaning of complex expressions *only* from information derived from their immediate constituents and their structure.

Szabo [193] eloquently elaborates this suspicion. The fact that S, an arbitrary competent speaker of English, understands some token of a novel complex expression e shows only that the information necessary to determine e's meaning is available to S in the context and information state in which S processed e. Part of the information S has in this state is information about e's structure and about the meaning of its constituents. But as Szabo reminds us, S also has access to other information—e.g., other linguistic information, general features of the context and certain general beliefs. This information may partly determine the meaning of tokens of e and other complex expressions, in which case structural and constituent information is not generally sufficient to determine the meaning of complex expressions. Since the opposite is assumed by CO, we might conclude that we should replace CO with a 'weaker' MDC compatible with the possibility that there is a set of non-constituent-derived but generally available information which partly determines the meaning of certain types of complex expressions.

This suspicion against CO derives most of its initial plausibility from its generality. To show this, we will now examine *particular* proposals for types of information that could, via non-compositional SRs, partly determine the meaning of certain types of complex expressions. The non-compositional proposals most commonly presented appeal to certain types of (i) contextual information and (ii) general beliefs. From our perspective, the often overlooked point to note is that to allow information of type (i)-(ii) to partly determine, via non-compositional SRs, the meanings of certain types of complex expressions, we have to assume, in each case, that the MDC of FL weakens condition 2 of CO to allow the desired non-compositional SRs.[11] However, in doing that each non-compositional MDC is also made compatible with many other (unintended) SRs. As a result, unlike CO, each non-compositional MDC predicts patterns of early linguistic development that seem obviously incorrect.

2.5.1 Context-Sensitive Non-compositional MDCs

The first proposal we will consider is to replace CO with a MDC which allows non-constituent contextual information to partly determine the meaning of (certain types of) complex expressions. The motivation for adopting this non-compositional MDC is that there seem to be complex expressions with 'unarticulated semantic constituents': their meaning is determined by the meanings of their parts, their structure, and certain contextual information which is not the meaning of any of its constituents.

The paradigmatic examples of expressions with 'unarticulated constituents' are simple 'meteorological expressions' like:

(2.2) It is raining.

[11] There are various reasons why this is often overlooked. One is that critics often focus only on a particular type of complex expression, and on SRs for that type of expression, and fail to consider the *general* consequences of adopting a non-compositional MDC that is weak enough to permit the particular SR they are considering.

In most contexts, tokens of (2.2) seem to express the proposition that it is raining at a certain time and place. One can hold that the relevant time is signified by the tense of the auxiliary verb, so that the time is represented at LF. But it seems that no constituent of (2.2) signifies or indexically encodes the relevant place. At the same time, it seems that competent speakers who understand the parts of (2.2) can understand (2.2), i.e., it seems that competent speakers can productively and systematically arrive at these interpretations. This suggests that: (i) the proposition expressed by (2.2) includes information of a location, and (ii) this information is not determined by either the structure or the constituents of (2.2). According to this account, then, the meaning of simple meteorological expressions like (2.2) is not determined compositionally.

Note that unlike the unarticulated constituency account just sketched, most accounts of the meaning of meteorological expressions respect CO. For example, Borg [14] argues that at LF (2.2) has a time but not a location variable; but she denies that the proposition literally expressed by (2.2) has a location specification. Recanati [161] defends an account similar to Borg's: (2.2) can be used to assert a proposition that is indefinite with respect to location, which suggests that location definite uses of (2.2) involve primary pragmatic enrichments.

Since there are CO compatible accounts of simple meteorological expressions, why replace CO with a MDC that allows non-compositional accounts of meteorological expressions? The issue in this discussion is not whether there are accounts of meteorological expressions compatible with CO; the issue is whether we should take CO as the MDC of FL. We might question this if adopting CO *forces us*, a priori, to dismiss otherwise plausible accounts of the meaning of certain types of complex expressions, such as the unarticulated constituency account of simple meteorological expressions. It *seems* preferable to adopt a MDC which allows both compositional and non-compositional SRs.

Basically for these reasons, Jonsson [82] argues that we could replace CO with LOC, which we here reformulate as a MDC:

(LOC) If L is an Ψ-language which FL can represent, then:

1. L cannot use lexical rules to determine the meanings of complex expressions.
2. Each SR in L is of form (a) or (b):
   (a) '$[\![[_Z X\ Y]]\!] = f_Z([\![X]\!], [\![Y]\!])$', where '$f_Z$' stands for a humanly computable function defined on the set of meanings
   (b) '$[\![[_Z X\ Y]]\!] = f_{Z'}([\![X]\!], [\![Y]\!], g)$' where '$g$' stands for a location function (functions from contexts to places) and '$f_{Z'}$' stands for a humanly computable function defined on the set of meanings and location functions.

LOC is weaker than CO in the sense that condition 2 allows, on the right-hand side of SRs, reference to location functions which are not the meaning of a constituent of the complex expressions whose meaning they determine.[12] This opens space for unarticulated constituency accounts of simple meteorological expressions, via construction-specific SRs. To illustrate, assume that (2.2) has the following simple structure:[13]

(2.2) $[_S[_E\ It][_{VP}[_{Aux}\ is][_V\ raining]]$

The unarticulated location function could be introduced, via a construction-specific non-compositional SR, at the level of the S or VP. For our purposes

[12]As stated, LOC is logically weaker than CO: every Ψ-language compatible with CO is compatible with LOC but not vice-versa. However, I do not emphasize this because we can impose additional constraints on LOC (some of which we will discuss below) which entail that there is no logical strength ordering between LOC and CO. For our purposes what is crucial is only that LOC has to weaken condition 2 of CO to allow the desired type of non-compositional SRs.

[13]This structure is obviously not the one that would be assigned by a serious syntactic theory. For a more realistic structure, see footnote 14. However, none of the points I will make depend on the particular structure assigned to simple meteorological expressions.

this choice does not matter, but assume it is introduced at the level of the VP:

(2.3) $[\![[_{VP} Aux\ V]]\!] = f_{VP'}([\![Aux]\!], [\![V]\!], g_1)$

Rule (2.3) contains the function g_1, a location function which is not the meaning of any of the constituents of the left-hand side of the rule.

As Jonsson [82] argues, adopting LOC does not affect the explanation of $P\&S$. Firstly, LOC, like CO, prohibits lexical rules to determine the meaning of complex expressions. Secondly, the sorts of unarticulated meanings which LOC allows—i.e., location functions from contexts to places—are constituted by information which speakers generally have access to, and there is no reason to deny that FL can access this kind of contextual information. For these reasons, we might be tempted to conclude that the non-compositional LOC is a better choice of MDC than CO.

However, LOC entails that speakers have to face certain choices in language acquisition that, judging from the general patterns of early linguistic development, speakers never seem to face.

To see that, note, first, that if we assume LOC, then to generate English (using non-compositional SRs) speakers would have to acquire an Ψ-language with construction-specific SRs such as (2.3), and not one with general SRs such as FA. For only in the case of *some types* of complex expressions—e.g., meteorological expressions—is it plausible to assume that unarticulated location functions partly determine their meanings. For example, the meanings of most NPs of the form $[_{NP} A\ N]$—*black cat*, *angry cow*, *pretty dolphin*, etc.—do not include an unarticulated constituent that, given a context, determines a location. The same is true of most Ss, e.g., *John is thinking*, *Einstein's idea is fantastic*, *Empiricism is dead*, etc. Indeed, even construction-specific SRs like (2.3) would need to be reformulated in terms of more fine-grained syntactic categories, for (2.3) incorrectly assigns a location specification to *all* tensed VPs with the syntactic structure $[_{VP}\ Aux\ V]$, e.g., *is thinking* and *is happy*.[14] So if we accept LOC, speakers would have to either acquire or innately posses (fine-grained) construction-specific SRs.

Consider the first option, that construction-specific SRs (a fortiori, phrase-structure rules) are innate. If we take this option, we would have to attribute *substantially* more innate knowledge to speakers than if we adopted CO, while gaining no descriptive coverage. In any case, this option is empirically implausible. Assuming a syntax that uses phrase-structure rules, there is substantial cross-linguistic evidence that at least some of these rules have to be acquired (Roeper [169]). For example, compounds are recursive in Germanic languages but not in Romance languages. Possessives are recursive in English but not in German. Prenominal As are recursive in English but not in French, and the opposite holds for post-nominal As. There are plenty of other examples like this. If phrase-structure rules have to be acquired, then the construction-specific SRs for such rules cannot plausibly be innate. So assuming that construction-specific SRs are innate is not a viable option for defenders of LOC.

[14] This point does not depend on assuming that simple meteorological expressions have that structure: regardless of the particular syntactic structure we assign to simple meteorological expressions, there are other expressions which are syntactically identical *at the level of structure where the non-compositional rule applies*, but which do not have a location specification. For example, assume that simple meteorological expressions have the syntactic structure they are assigned in most P&P syntactic theories: $[TP\ [[PRN]\ [\bar{T}\ [T\ VP]]]]$ (see Radford [159]). The non-constituent contextual function g could be introduced at the level of the complex TP, via an SR for $[_{TP} PRN\ \bar{T}]$, or at the level of the complex $\bar{T}$, via an SR for $[_{\bar{T}} T\ VP]$. As a result, expressions like *he is thinking*, *he is depressed*, *it is said that life is short*, *it is sad to think about suffering* would be incorrectly assigned a location specification.

To avoid this problem of implausibly over-saturating all sorts of complex expressions with unarticulated location specifications, the SRs acquired by LOC constrained FLs have to use construction-specific SRs with fine-grained syntactic categories. For example, to give an unarticulated constituency account of expressions like *it is raining* in a way that doesn't over-saturate with location specifications the meanings of similarly structured sentences, we can include a new category for a subset of verbs, call it V_e. Assume V_e includes verbs of physical events like *snows* and *rains*, but not verbs of mental events or states like *loves* and *thinks*. A model which includes fine-grained category V_e/VP_e can include fine-grained construction-specific SRs like (4.7)-(4.8):

(2.4) $[\![[_{VP_e} Aux\ V_e]]\!] = f_{VP'}([\![V_e]\!], [\![Aux]\!], g_1)$

(2.5) $[\![[_{\bar{T}_e}\ T\ VP_e]]\!] = f_{\bar{T}_e}([\![T]\!], [\![VP_e]\!], g_1)$

The other option is to assume that construction-specific SRs are learned or acquired. In itself, this is not a problem. Assume that S knows (i) that '$\gamma\ \beta$' is an expression of the form $[_{VP} Aux\ V]$, and (ii) $[\![\gamma\ \beta]\!]$, $[\![\gamma]\!]$, and $[\![\beta]\!]$. For S to acquire a rule like (2.3) from (i) and (ii), we have to assume that S is able to use information (i) and (ii) to test hypotheses like (2.6) and (2.7):

(2.6) $[\![\gamma\ \beta]\!] = f_{VP'}([\![\gamma]\!], [\![\beta]\!], g_1)$

(2.7) $[\![\gamma\ \beta]\!] = f_{VP}([\![\gamma]\!], [\![\beta]\!])$

If (2.7) fits the data, S assumes that $[\![\gamma\ \beta]\!]$ is determined from its parts using compositional rule (2.8):

(2.8) $[\![[_{VP} Aux\ V]]\!] = f_{VP}([\![Aux]\!], [\![V]\!])$

If (2.6) fits the data, S assumes that $[\![\gamma\ \beta]\!]$ is determined from its parts using non-compositional rule (2.3), here repeated:

(2.3) $[\![[_{VP} Aux\ V]]\!] = f_{VP'}([\![Aux]\!], [\![V]\!], g_1)$

Suppose that S concludes that hypothesis (2.6) fits the data better than hypothesis (2.7), then S can generalize to all structures of the form $[_{VP} Aux\ V]$ and so acquire non-compositional SR (2.3).

Note, however, that if we replace CO with LOC analogous learning procedures enter each case of acquiring a construction-specific SR. For unlike CO, LOC allows the *general* possibility that construction-specific non-compositional SRs determine the meaning of *every syntactic type* of complex expression. Assuming only LOC, each time S acquires a construction-specific SR, S would be open to consider compositional and non-compositional alternatives. For example, assume that (9)-(10) are correct SRs, in the sense that they output the correct meaning-assignments, relative to English. En route to acquiring (9)-(10), S is cognitively 'open' to consider a hypotheses space like (9*)-(10*), where each g_i stands for a location function:

(2.9) $[\![[_S NP\ VP]]\!] = f_S([\![NP]\!], [\![VP]\!])$

(2.10) $[\![[_{NP} AP\ NP]]\!] = f_{NP}([\![AP]\!], [\![NP]\!])$

(2.9*) $[\![[_S NP\ VP]]\!] = f_S([\![NP]\!], [\![VP]\!])$ or $f_{S'}([\![NP]\!], [\![VP]\!], g_1/g_2/.../g_n)$

(2.10*) $[\![[_{NP} AP\ NP]]\!] = f_{NP}([\![AP]\!], [\![NP]\!])$ or
$f_{NP'}([\![AP]\!], [\![NP]\!], g_1/g_2/.../g_n)$

This entails that, at some early stage in the acquisition of Ψ-English, S *could* make mistakes like (2.11)-(4.13), even if eventually S ends up acquiring the correct *SR*s (2.9)-(2.10) (in the examples below, assume that S knows the relevant syntax and the correct meanings for the lexical items, relative to Ψ-English).

(2.11) $[\![[_{S_{ms}} NP\ VP_{ms}]]\!] = f_{S'}([\![NP]\!], [\![VP_{ms}]\!], g_1/.../g_n)$, where '$VP_{ms}$' is a subcategory of VPs headed by Vs of mental states. So S takes $[\![John\ is\ happy]\!]$ to be JOHN IS HAPPY HERE/CLOSE TO HOME/EVERYWHERE/... (note that LOC allows location functions other than g_1), $[\![Mary\ is\ sad]\!]$ to be MARY IS SAD HERE/ CLOSE TO HOME/EVERYWHERE/... and so on.

(2.12) $[\![[_{S_{mt}} NP\ VP_{mt}]]\!] = f_{S'}([\![NP]\!], [\![VP_{mt}]\!], g_1/.../g_n)$, where '$VP_{mt}$' is a subcategory of VPs headed by Vs of mental traits. So S takes $[\![John\ is\ silly]\!]$ to be JOHN IS SILLY HERE/CLOSE TO HOME/EVERYWHERE/..., $[\![Mary\ is\ courageous]\!]$ to be MARY IS COURAGEOUS HERE/CLOSE TO HOME/EVERYWHERE/..., and so on.

(2.13) $[\![[_{NP_c} AP_c\ NP)]\!] = f_{NP'}([\![AP_c]\!], [\![NP]\!], g_{10}/.../g_n)$, where '$AP_c$' is a subcategory of APs headed by color As. So S takes $[\![gray\ shark]\!]$ to be GRAY SHARK WHEN UNDER WATER/WHEN I SEE IT/WHEN OUTSIDE WATER/..., $[\![red\ fish]\!]$ to be RED FISH WHEN UNDER WATER/WHEN I SEE IT/WHEN OUTSIDE WATER/..., and so on.

The meaning assignments in (2.11)-(4.13) are consistent with the learning-data encountered by most speakers during early language acquisition. For example, S can reasonably assume that color As, when combined with common Ns, result in color attributions that have location restrictions, resulting in cases like (4.13). The non-obvious mistakes in (4.13) would take time to correct, since the location-restricted assertion-conditions of each color A and common N complex expression is a commonly used subset of their (mature English) non-location restricted assertion-conditions. However, speakers do not make mistakes about the meaning of color A and common N compounds analogous to those presented in (4.13). Similarly, S can reasonably assume that expressions which attribute mental states or traits are location restricted, like (2.11)-(2.12). The non-obvious mistakes in (2.11)-(2.12) would also take time to correct, since the location-restricted assertion-conditions of each expression attributing a mental state is a commonly used subset of their (English) non-location restricted assertion conditions. However, speakers do not make mistakes about the meanings of expressions attributing mental states or traits analogous to (2.11)-(2.12).

That speakers, even early in development, never go through states like (2.11)-(4.13) suggests that they never consider a hypothesis space like (2.9*)-(2.10*), i.e., a hypothesis space which includes, for each type of complex expression, a set of possible non-compositional SRs, each involving some unarticulated location function. However, this is the hypothesis space that would be open to speakers if their MDC was LOC. So either we reject LOC, or it is a mystery why speakers acquiring Ψ-English never adopt or even consider 'reasonable but mistaken' construction-specific non-compositional SRs like those presented in (2.11)-(4.13).

To sum-up, although LOC does not undermine the explanation of $P\&S$, it seems to entail false predictions about patterns of early linguistic development. LOC entails that language acquisition partly consists in acquiring phrase-structure rules and construction-specific SRs. Each time speakers acquire a construction-specific SR, they would be open to consider a whole set of competing non-compositional SRs. Given the type of information encountered in early language acquisition scenarios, the fact that speakers do not adopt, at least temporarily, some reasonable but 'incorrect' non-compositional SRs is left completely unexplained. In addition, replacing CO with LOC buys our linguistic theories no additional descriptive coverage. For these reasons, CO is a more plausible MDC than the weaker, non-compositional LOC.

2.5.2 General Beliefs Sensitive Non-compositional MDCs

Another recent and more influential proposal is to replace CO with a MDC that allows general beliefs to partly determine the meanings Ψ-languages assign to (certain) types of complex expressions (Prinz [145], Jonsson and Hampton [67]). One reason to propose such a MDC is that there are complex expressions that seem to have semantic features neither present in nor determined by the semantic features of their constituents. For example, consider complex NPs such as *black cat* and *brown cow.* For some speakers, $[\![black\ cat]\!]$ seems to include the information that their presence brings bad luck; and for others $[\![brown\ cow]\!]$ seems to include the information that brown cows produce bad milk. To linguistically account for the 'free-enrichment' of complex NPs, some theorists propose that we allow general beliefs about the extension of complex NPs—usually called 'extensional feedback' beliefs—to partly determine their meaning, even if such beliefs are not part of the meaning of their constituents (Murphy [117], Prinz [143, 145], Hampton [63], Jonsson and Hampton [67]). There are other types of 'free-enrichments' of complex expressions, including NPs not plausibly modeled as computations involving the use of extensional feedback. For now we will focus on the extensional feedback class.

There are accounts of 'free-enrichment' complex NPs compatible with CO. One proposal assumes that the meaning of tokens of common Ns is enriched online to include some (relevant) encyclopedic information (Barsalou [6]; Carston and Wilson [210]). According to this view, in some contexts, $[\![cow]\!]$ is enriched to include information like 'produces bad milk if brown', $[\![cat]\!]$ to include information like 'brings bad luck if black', etc. There is no a priori reason to reject the claim that the meaning of some lexical items can be enriched online to include such information. Another proposal is to hold that the free-enrichments of the meanings of complex NPs are post-linguistic pragmatic enrichments, even if they are often sub-personal and automatic. The intuitive meanings of many kinds of expressions are affected by pragmatic enrichments, which are often sub-personal and automatic (Recanati [161]; Glucksberg [57]). There is no a priori reason to reject the claim that such post-compositional processes account for the 'intuitive' meanings of free-enrichment complex NPs.

Since there are CO compatible accounts of free enrichment complex NPs, why replace CO with a MDC that allows non-compositional accounts of the meanings of these complex NPs? Again, the issue in this discussion is *not* whether there are CO compatible accounts of free-enrichment complex NPs; the issue is whether we should take CO as the MDC of FL. We might question this if adopting CO $forces$ us, a priori, to dismiss apparently reasonable accounts of the meaning of certain types of complex expressions, such as the extensional feedback account of free-enrichment complex NPs.

To allow non-compositional accounts of free-enrichment complex NPs, Jonsson [82] proposes GEN, which we here reformulate as a MDC:

(GEN) If L is an Ψ-language which FL can represent, then:

1. L cannot use lexical rules to determine the meanings of complex

expressions.

2. Each SR in L is of form (a) or (b):

   (a) '$[\![[_Z X\ Y]]\!] = f_Z([\![X]\!], [\![Y]\!])$', where '$f_Z$' stands for a humanly computable function defined on the set of meanings

   (b) '$[\![[_Z X\ Y]]\!] = f_{Z+}([\![X]\!], [\![Y]\!], b)$' where '$b$' stands for a set of general beliefs and 'f_{Z+}' stands for a humanly computable function defined on the set of meanings and general beliefs.

GEN is weaker than CO in the sense that it allows, to partly determine the meaning of complex expressions, general beliefs which are not part of the meaning of their constituents.[15] This opens space for non-compositional accounts of free-enrichment complex NPs, including extensional feedback accounts, via non-compositional SRs:

(2.14) $[\![[_{NP} A\ N]]\!] = f_{ex}(f_{NP}([\![A]\!], [\![N]\!]), b)$

(4.14) refers to the extensional feedback belief set b, which is not the meaning of any of the constituents of the complex expressions whose meaning it partly determines. According to (4.14), the meaning of complex NPs is a function f_{ex} from the value of the ordinary compositional function f_{NP} for NPs, which applies to the meaning of its constituents, and from the extensional-feedback belief set b. To illustrate f_{ex}, assume that b stands for S's extensional feedback, which includes the belief that black cats bring bad luck:

(2.15) $f_{ex}(f_{NP}([\![black]\!], [\![cat]\!]), b) =$ BLACK CAT AND BRINGS BAD LUCK

What f_{ex} does is to incorporate into the meaning of complex NPs, as compositionally determined, whatever beliefs about the NP there are in the extensional

[15] As stated, GEN is logically weaker than CO: all Ψ-languages compatible with CO are also compatible with GEN, but not vice-versa. However, what I said with respect to LOC also applies in this case: what is crucial, for our purposes, is only that GEN has to weaken condition 2 of CO to allow for the desired type of non-compositional SRs. This is compatible with there being some additional constraints on GEN (some of which we discuss below) that entail that there is no strict logical strength ordering between GEN and CO.

feedback belief set.

As Jonsson [82] argues, adopting GEN does not seem to affect the explanation of $P\&S$. Firstly, GEN, like CO, prohibits lexical rules for complex expressions. Secondly, the beliefs which GEN allows to partly determine the meaning of free-enrichment complex NPs are beliefs which speakers have access to, at least at the personal level. If we assume that FL also has access to these general beliefs—i.e., if we assume that FL is *not* an informationally encapsulated module—Ψ-languages constrained by GEN would satisfy $P\&S$.

Why should we stick with CO, which blocks the possibility that speakers can acquire Ψ-languages with some non-compositional SRs like (4.14) to deal with extensional feedback complex NPs, when we can, apparently without violating $P\&S$, adopt a weaker non-compositional MDC which allows that possibility? The problem faced by GEN is quite similar to the problem faced by LOC.

Note, first, that SRs like (4.14) are actually problematic, relative to English. If S's Ψ-language has (4.14), it follows that, if α is an expression of the form $[_{NP}A\ N]$, its meaning is partly determined by *all* of S's extensional feedback beliefs about the referent of α. For example, if S believes not only that black cats bring bad luck, but also that cats are felines and felines are never underwater, that cats are animals and that animals are robots controlled by Martians, then:

(2.16) $f_{ex}(f_{NP}([\![black]\!], [\![cat]\!]), b) =$ BLACK CAT AND BRINGS BAD LUCK AND IS NEVER UNDER WATER AND IS A ROBOT CONTROLLED BY MARTIANS

However, whatever peculiar beliefs about cats, black cats, black animals, colored animals, etc., S has, not *all* of them are part of $[\![black\ cats]\!]$, as determined by S's Ψ-language. Some theorists have defended the idea that beliefs about the referent of an arbitrary complex expression can, in principle, affect the meaning speakers assign to it; but with the exception of radical holists, no one defends the view that *all* such beliefs affect the meaning speakers assign to it, as (4.14) entails.[16] Most theorists who assume that extensional feedback beliefs about the entity denoted by a complex expression partly determine its meanings implicitly assume that only *some* of those beliefs play such a role. Of course, GEN is *compatible* with constrained versions of (4.14), such as (4.14+):

(4.14+) $[\![[_{NP}A\ N]]\!] = f_{ex}(f_{NP}([\![A]\!], [\![N]\!]), b^{+})$

(4.14+) is just like (4.14) except that the extensional belief set b^{+} is a subset of the extensional belief set b. What subset? One option, assuming that we can represent something like degrees of belief, is that b^{+} only includes extensional feedback beliefs that pass a threshold. In this way, not *all* of S's extensional feedback beliefs about e.g. black cats would be included in the meaning which S's Ψ-language assigns to *black cat*, but only the ones that S 'really' believes.

Still, although rule (4.14+) might work for cases like *black cat* and *brown cow*, it is incorrect for English. For example, take *gray shark* and *fierce lion*, which have the form $[_{NP}A\ N]$, so that (4.14+) applies to both of them. Suppose again that at some early stage of linguistic development, S 'really' believes that lions are only fierce in their territory, and that color attributions to fish are restricted to the way they look when underwater. In this case, rule (4.14+) would have the result that $[\![$*fierce lion*$]\!]$ is something like FIERCE LION IN HIS TERRITORY, and that $[\![$*gray shark*$]\!]$ is something like GRAY SHARK WHEN UNDERWATER. In short, just as in the case of LOC, if we adopt GEN, we have to assume that speakers acquire construction-specific SRs which can use fine-grained syntactic categories, such as the following:

[16] The most famous holist about linguistic meaning is probably Block [8]. However, I agree with Block [9] that holistic inferential role theories of meaning are compositional. The reason for this is simple. According to these theories, the meaning of an expression is given by all of its inferential roles. Hence extensional feedback beliefs such as that black cats bring bad luck are part of $[\![$*cat*$]\!]$. From this perspective, to hold that extensional feedback beliefs are incorporated into the meanings of tokens of complex expressions via non-compositional SRs would be entirely superfluous, since they are already part of the meaning of the constituents. For further discussion, see Szabo's [189] response to Fodor and Lepore's [48] claim that total inferential roles are not compositional.

(2.17) $[\![[_{NP_a} A_c\ N_t]]\!] = f_{ex}(f_{NP_a}([\![A_c]\!], [\![N_t]\!]), b^+)$

'NP_a' is a category of NPs formed out of color As and terrestrial animate beings. In this case, (4.28) applies to *brown cow* and *black cat*, but not to *fierce lion* and *gray shark*.

However, if we adopt GEN, S could, early in linguistic development, test 'mistaken' construction-specific SRs like (4.14) and (4.14+) for at least some types of complex expressions, in cases in which they output reasonable but incorrect meaning assignments. Consider the following examples (assume that S knows the relevant syntax and the correct meanings for the lexical items, relative to Ψ-English):

(2.18) $[\![[_{S_m} NP\ VP_m]]\!] = f_{S_m}(f_S([\![NP]\!], [\![VP_m]\!]), b+)$, where '$b+$' stands for the set of highly weighted extensional feedback beliefs, and 'S_m' is a category of Ss formed by a NP and a VP_m which predicates some mental trait. In this case, if S assumes that attributions of mental traits are restricted to certain locations, S would take $[\![John\ is\ silly]\!]$ to be JOHN IS SILLY IN HIS HOUSE (or some other reasonable location restriction), and so on.

(2.19) $[\![[_{NP_m} A_m\ N_{an}]]\!] = f_{NP_m}(f_{NP}([\![A_m]\!], [\![N_{an}]\!]), b^+)$, where '$NP_m$' is a category of NPs formed by mental trait As and Ns that stand for animate beings. In this case, if S believes that certain mental traits of animate objects are restricted to certain locations, S would take $[\![fierce\ lion]\!]$ to be something like FIERCE LION WHEN IN HIS TERRITORY (or some other location restriction), $[\![silly\ student]\!]$ to be SILLY STUDENT WHEN IN HIS SCHOOL (or some other reasonable location restriction), and so on.

(2.20) $[\![[_{NP_p} A_m\ N_{an}]]\!] = f_{NP_p}(f_{NP}([\![A_m]\!], [\![N_{an}]\!]), b^+)$, where '$NP_p$' is a category of NPs formed by physical trait As and Ns that stand for animate beings. In this case, if S believes that certain physical traits of animate objects are strongly correlated with certain mental traits, S could take $[\![strong\ cat]\!]$ to be something like STRONG AND MEAN CAT, $[\![strong\ student]\!]$ to be something like STRONG AND BULLYING STUDENT, and so on.

Again, we do not find patterns of early linguistic development in which speakers test reasonable but incorrect SRs such as (4.29)-(4.31). This suggests that FL is not as unconstrained as it would be if GEN was its MDC. Note, in addition, that GEN faces basically the same problems faced by LOC. For if assuming GEN, speakers would have to test and eliminate reasonable but incorrect construction-specific SRs which have the same output as those we considered in (2.11)-(4.13), which we used to criticize LOC. To illustrate this, consider (2.12) again:

(2.12) $[\![[_{S_{mt}} NP\ VP_{mt}]]\!] = f_{S'}([\![NP]\!], [\![VP_{mt}]\!], g_1/.../g_n)$, where '$VP_{mt}$' is a subcategory of VPs headed by Vs of mental traits. So S takes $[\![John\ is\ silly]\!]$ to be JOHN IS SILLY HERE/CLOSE TO HOME/EVERYWHERE/..., $[\![Mary\ is\ courageous]\!]$ to be MARY IS COURAGEOUS HERE/CLOSE TO HOME/EVERYWHERE/..., and so on.

GEN does not allow SRs like the one used in (2.12), but it allows SRs that have, under similar conditions, very similar outputs, as is illustrated by (4.29) and (4.30). As we argued before, if the MDC of FL doesn't exclude these options, it would be 'reasonable', early in linguistic development, for S to *believe* that assertions using certain types of complex expressions have location restrictions, as in (2.11)-(4.13), even if these beliefs are eventually abandoned.

So GEN inherits most of the problems faced by LOC, and introduces some of its own. If we hold that GEN is the MDC, we have to explain how speakers acquire rules closer to (4.28) than to (4.14) or (4.14+), how they acquire something like (2.9)-(2.10) from an initial hypotheses space that is much wider than (2.9*)-(2.10*), and so on. But even if that can be explained, the crucial point is that we should find, early in linguistic development, mistakes that reveal the use of construction-specific SRs like (4.29)-(4.31), which result in various kinds of reasonable but incorrect free-enrichments of certain types of complex expressions.

This objection to GEN is important because *many* critics of compositionality hold that what we should infer from $P\&S$ is only that FL respects some weak constraint along the lines of GEN—a constraint which allows the meaning of complex expressions to be compositionally determined, but also allows the meaning of some types of complex expressions to be partly determined by general beliefs (Murphy [117], Prinz [143,145], Hampton and Jonsson [67]). But if we change our perspective and take MDCs not as convenient methodological assumptions but as empirical hypotheses about the functional architecture of FL, the problematic consequences for language acquisition of weak constraints such as GEN are intuitively easy to see.

Suppose that, using any general learning strategies at your disposal, you are given the task of acquiring the semantics (S_T) of a target Ψ-language (Ψ_T). To do this, you are given subsets of the language (L_T) generated by Ψ_T and some hints about what S_T *cannot* be like. Suppose that the only 'hint' about S_T—the only hint about the form of the SRs—that you are given is that it satisfies something like GEN. This amounts to the following hint: for a complex expression e of any type generated by Ψ_T, $[\![e]\!]$ is determined by the meanings and structure of its immediate parts, and possibly *any* other general beliefs which are consistent with the data. Given only this hint, you would begin the task of acquiring the SRs of Ψ_T with a substantially unconstrained hypothesis-space. Even if you eventually acquire S_T, you would have to consider and reject many reasonable but 'incorrect' construction-specific SRs such as (4.29)-(4.31). You would have to learn, as you encounter more L_T data, that only some construction-specific SRs that involve extensional feedback and/or some subsets of general beliefs are correct for Ψ_T.

This process of testing and rejecting reasonable construction-specific SRs is not a process which we ever, at least systematically, observe in actual patterns of early linguistic development, which suggests that normal speakers do not begin the acquisition of target Ψ-languages with a MDC as unconstrained as GEN.

It is natural to think that we can, if not avoid, at least weaken the force of this objection by defending instead a constrained version of GEN. For example, consider a MDC, call it 'GEN^+', which allows non-compositional SRs but only of the extensional feedback type. We need not state GEN^+ in detail, since it is obvious how to do this. GEN^+ might seem like an ad hoc MDC, proposed merely to reject CO. However, extensional feedback beliefs have a special place amongst our general beliefs, e.g., they are explicitly stored in memory, are for the most part easily retrievable, and can be incorporated into meaning assignments without having to appeal to complicated inferential computations. So it is not implausible to suggest that FL has selective access to extensional feedback beliefs, but not to other types of general beliefs. Another option is a MDC, call it GEN^*, which allows non-compositional SRs that involve only highly-weighted beliefs. Highly-weighted beliefs also have a special place amongst our beliefs. There is no known reason to deny that FL has selective access only to highly weighted beliefs, but not to other types of general beliefs. However, note that the construction-specific non-compositional SRs used in (4.29)-(4.31) are compatible with both GEN^+ and GEN^*, since they refer only to sets of highly-weighted extensional feedback beliefs. So the same sort of general objection raised against LOC and GEN can be raised against GEN^+ and GEN^*.

2.6 Objections and Open Issues

All the MDCs we examined can account for $P\&S$, but the non-compositional ones predict incorrect patterns of early linguistic development. This strongly suggests that CO is the most plausible MDC currently on the table. Let us now consider some objections to this argument.

2.6.1 Objection 1: Against the weak Chomskian conception of FL

We presented a relatively concrete, computational conception of FL. This might worry some theorists, especially philosophers, who usually build their arguments for or against compositionality on radically unconstrained conceptions of FL and natural languages. I think this is part of the reason why these arguments are generally found unconvincing.[17] In any case, we tried a different approach. In addition to overall plausibility, we adopted a conception of FL that meets two conditions: (i) it is flexible enough to be compatible with the competing compositional and non-compositional MDCs, and (ii) it is concrete enough so that, when paired with one of the competing MDCs, it issues in clear empirical consequences, esp., for language acquisition. This conception of FL can be accepted by theorists on both sides of this debate.

In contrast, some famous advocates of compositionality assume a conception

[17] For example, consider an often cited version of the traditional 'learnability' argument for compositionality, which says that, if some language L is productive and systematic, it follows that, for L to be learnable, L must be compositional. As various critics point out, this argument is clearly invalid as stated, and unconvincing if reformulated as an inference to the best explanation. For L to be learnable in principle it is necessary and sufficient that it have a computable syntax and semantics, and a function's being computable doesn't guarantee that it is compositional (Dever [37]; Pagin and Westerstahl [123]). Without further constraints on L and the learning process, nothing else can be inferred.

of FL that cannot be accepted by critics. For example, some theorists assume that FL is an informationally encapsulated module, and then appeal to this property to defend its compositionality (see e.g. Borg [14] and Larson and Segal [100]). Critics are justifiably skeptical of this way of defending compositionality (Robbins [168]; Jonsson [82]: chap. 6). Ψ-language exhibit *some* degree of modularity—they are domain specific, have mandatory operations which are (for the most part) fast, have limited central accessibility, and characteristic patterns of breakdown and development. If to account for these features of Ψ-language we assume their (almost) total informational encapsulation, it follows that they are likely compositional. For non-compositional MDCs require that, to determine the meanings of complex expressions, the Ψ-language have access to some subset of non-linguistic information. However, as most theorists rightly point out, modularity comes in degrees. Even if we can safely hold that FL is modular to some non-trivial degree, there is currently no good reason to assume that it is informationally encapsulated to the degree that would be required to make it incompatible with non-compositional MDCs such as LOC and GEN.[18] In short, defending the compositionality of FL by assuming its (almost) total informational encapsulation is an ad hoc move based on a currently unsupported assumption.[19]

The conception of FL we presented is more plausible and, relative to this

[18]Of course, some non-compositional MDCs might be in tension with the assumption that FL is *weakly* modular. However, except for GEN, all the non-compositional MDCs we examined require only limited information exchanges between FL and other modules. A related point comes-up in a discussion between Robbins and Borg. A central premise in Borg's [14] defense of the view that semantic processing is compositional is the claim that semantic processing exhibits the basic signs of modularity, from which Borg infers that semantic processing is informationally encapsulated. Robbins [168] convincingly argues that semantic processing does not seem to be modular to the extent that Borg needs to infer its radical informational encapsulation.

[19]Language processing is 'for the most part' fast; but there are many exceptions. As Chomsky [23] notes, for most speakers, it takes some time to determine that *John is too stubborn to visit anyone who talked to* is gibberish; or that *John is too stubborn to expect anyone to talk to* means that John is so stubborn that an arbitrary person would not expect anyone to talk to John; or that in *his sister loves her brother*, *his* can co-refer with *her brother*, and *her* with *his sister*, if their reference is not somehow contextually supplied.

debate, deliberately neutral. We assumed that Ψ-language are cognitive systems that compute E-languages; that some parts of Ψ-language have to be acquired, e.g., lexical semantic rules and, under some views about the MDC, some construction-specific SRs. We also assumed that other parts of Ψ-language are part of their fixed functional architecture, e.g., general MDCs. We did *not* assume that FL is informationally encapsulated. We also did *not* impose strict conditions on the learnability of Ψ-language. Specifically, when examining the non-compositional MDCs, we did *not* assume that correct SRs could not be acquired given a non-compositional MDC, the generally available linguistic data, and some general learning procedure. Clearly, this conception of FL does *not* entail that Ψ-language are compositional. This conception can be accepted by theorists defending a non-compositional MDC and might even give them some hints about how they could support their position. For example, they could look for evidence that early in development children do try some non-compositional SRs, or that some cases of language loss suggest the selective loss of particular construction-specific SRs. Currently there is no such evidence. But the only point I'm making here is that the conception of FL we assumed has properties which are, unlike assuming its information encapsulation, relatively uncontroversial, independently motivated, and can be fruitfully adopted even by theorists sympathetic to non-compositional MDCs.

2.6.2 Objection 1: Is CO as descriptively adequate as the non-compositional MDCs?

Assume that a MDC satisfies the condition of 'descriptive adequacy' relative to L if an Ψ-language compatible with it generates L. One might question the claim that, in terms of descriptive adequacy relative to English and other natural languages, CO and the non-compositional MDCs are on equal footing. We did show, for each type of complex expression which motivated the introduction of a non-compositional MDC (meteorological expressions for LOC and extensional feedback complex NPs for the versions of GEN), that there are plausible accounts of how their meaning is determined compatible with CO. However, these types of expressions are only a subset of the types of expressions often considered problematic for compositionality, which include conditionals, genitives, nominal compounds, etc. If some of these types of expressions can only be given a non-compositional account, then relative to descriptive adequacy CO is in worse shape than some of the non-compositional MDCs. Indeed, several critics have argued that the view that FL is compositional is empirically false because the meanings of various types of complex expressions don't seem to be compositionally determined (see e.g. Fodor [45], Lahav [97] and Hampton [66])

This is a reasonable worry, but as is clearly illustrated in most recent surveys on this issue, it has been substantially addressed by the collective effort of theorists who have proposed various compositional accounts for each of the problematic constructions.[20] These compositional accounts generally use the tools we used to show that there are CO compatible accounts of the meaning of simple meteorological expressions and extensional feedback complex NPs: lexical context-sensitivity, primary pragmatic processes, including meaning modulation, and lexical entries which in occasions of used are informationally enriched. The moral we should draw from this is that, as things currently stand, there are no types of expressions that can be taken as *direct* empirical counter-examples to CO.[21] This is why to resolve debates about the MDC of FL we need to

[20] Jonsson [82]: ch. 5 presents an up-to-date review of compositional accounts of many *problematic* linguistic constructions. See also Dever [37], Szabo [188], [190], Recanati [163], Pagin [121], [123], and Partee [125].

[21] At the end of a survey of 'problem cases' for compositionality, Jonsson concludes—echoing other theorists—that 'whether semantic theories in the end should be compositional...cannot be settled by attempting to provide examples that cannot be handled by a compositional (explicated in terms of CO) account since there does not seem to be any such cases' (Jonsson [82]: ch. 5).

move beyond descriptive adequacy.

To be clear, a consequence of adopting CO is that some data and intuitions about the meaning of certain expressions have to be dealt with at the level of pragmatics, hence not by appealing only to the workings of FL. With some important differences, most theorists accept this. For example, Borg [14]and other minimalists propose that we substantially 'clean' the data coming from meaning intuitions; while Recanati [161] and other contextualists try to account for a wider range of our pre-theoretical meaning intuitions. In discussions of competing MDCs what is important, when one of the accounts deals with some pre-theoretic intuition only in conjunction with non-linguistic cognitive processes, is that such decisions follow in a principled way from the assumed division between semantics and pragmatics, a division which everyone, including non-compositionalists, have to accept.

2.6.3 Objection 2: Why doesn't CO over-generate meanings in ways that parallel the non-compositional MDCs?

Assume CO is descriptively adequate in the sense just specified—and specifically that Ψ-languages compatible with it can account for meteorological expressions and extensional feedback complex NPs. Why, if a CO constrained FL can account for that phenomena, can't it also over-generate meanings in a way that parallels the reasonable but unattested mistakes allowed by FLs constrained by non-compositional MDCs such as LOC and GEN?

Consider a way in which CO could allow mistakes that *seem* to mirror the sorts of mistakes which we used to object to non-compositional MDCs. Here is an obvious candidate: if for some speaker S, $[\![cat]\!]$ = CAT AND MEAN IF STRONG—i.e., if it includes such conditional information—then $[\![strong\ cat]\!]$ would mean STRONG AND MEAN CAT. We can model this as a result of such conditional information being included in S's lexical entry for *cat*, or of its arising from particular meaning modulations (enrichments) in certain occasions of use of *cat*. Similar observations can be used to generate incorrect locative restrictions in cases such as *gray cat*. Suppose such (incorrect) conditional information could, in certain circumstances, be part of the occasion meaning of lexical items; you might then suspect that a CO constrained FL could give rise to the same incorrect patterns we invoked against the non-compositional MDCs.

In response, note that the mistake when S determines, under CO, $[\![strong\ cat]\!]$ can be traced to an overly enriched standing or occasion meaning for *cat*, hence does not systematically affect the meaning-assignments to other expressions, including complex expressions that do not have *cat* as a constituent. So mistakes which can be traced to lexical items (which are allowed by CO and by the non-compositional MDCs) are quite different from those allowed only by the non-compositional MDCs. Consider (4.30) above, which involves the SR "$[\![[_{NP_m} A_m\ N_{an}]]\!] = f_{NP_m}(f_{NP}([\![A_m]\!], [\![N_{an}]\!]), b^+)$", where '$NP_m$' is a category of NPs formed by mental trait As and Ns for animate beings. In (4.30), S is testing a non-compositional construction-specific SR, allowed by GEN, which systematically assigns to expressions of the form $[_{NP_m} A_m\ N_{an}]$—such as *fierce lion, courageous soldier, angry pit-bull, loving father*—a meaning which includes an incorrect (but reasonable) locative restriction. The mistake in (4.30) is not due to a particular lexical item, and affects a much wider range of expressions than the subset of expressions of which that item is a constituent. The mistake is due to S's 'trying out' an incorrect non-compositional SR for expressions of the form $[_{NP_m} A_m\ N_{an}]$. The objections which I raise against LOC, GEN, and its variants are of that form. They apply even under the assumption that S assigns the correct meanings to all lexical items.

In short, it is not true that the relevant patterns of mistakes that can occur assuming LOC or GEN can also occur assuming CO, which blocks speakers from considering construction-specific SRs. The reason why CO, which allows meaning modulation of lexical items, does not allow patterns of mistakes analogous to (11)-(13) and (18)-(19) is due to the case-by-case nature of pragmatic modulations: whether S takes an utterance of *gray shark* to mean GRAY SHARK WHEN UNDERWATER or *strong cat* to mean STRONG AND MEAN CAT, under this account, depends on particular features of the contexts of utterances. In other words, it is a case-by-case decision which does not systematically affect S's literal meaning assignments to expressions of the same form in other contexts, for S's FL, constrained by CO, cannot acquire construction-specific non-compositonal SRs. In contrast, if S, constrained by one of the non-compositional MDCs, is trying out a construction-specific non-compositional SR, then the inclusion of the locative restrictions or general beliefs would result automatically—i.e., from the automatic processing of FL—and would range over all expressions (including novel ones) of the form over which the adopted SR ranges.[22]

2.6.4 Objection 3: Aren't we ignoring crucial trade-offs between CO and the non-compositional MDCs?

One might argue that the previous argument for CO depends on ignoring the full set of trade-offs between the competing MDCs. Take the arguments against LOC and GEN. We said that we can account for the meaning of simple meteorological expressions and extensional feedback complex NPs in ways that are consistent with CO. But to do that we have to assume that there is a distinction between the standing and the occasion meaning of expressions, that modulation functions can further modify the occasion meaning of lexical items, and that some of these modifications consist of enriching their meaning (see section

[22] I am grateful to an anonymous reviewer for a detailed discussion of this objection.

2.4). However, except for dealing with an over generation worry in Objection 2, we didn't carefully consider other costs of using those tools.

The key to address this worry is to note that, as far as we know, these are tools which we need to incorporate into any plausible linguistic model, including those that adopt a non-compositional MDC, to account for expressions other than simple meteorological expressions or extensional feedback complex NPs. This is obvious in the case of having to appeal, as Borg, Recanati, and Glucksberg do, to (primary) pragmatic effects to account for the intuitive meaning of tokens of certain types of complex expressions. Furthermore, assuming that lexical items are semantically non-atomic and include rich arrays of information is also becoming popular in accounts of genitives, possessives, privative As and adverbial modifications (see Vikener and Jensen [206], Coulson and Fauconnier [32], and Wunderlich [211]). Finally, (although here we did not appeal to this tool) assuming that some lexical items, which are not obvious indexicals, have context-sensitive parameters is common in recent accounts of words like *tall*, $flat$, *green*, *home*, $faraway$ etc. (see Szabo [188], Segal and Rothschild [173], Kennedy and McNally [88], and Recanati [163]: ch. 3). Non-compositional MDCs also have to account for these sorts of expressions, and to do so have, in many cases, to use these same tools, for the non-compositional SRs which are allowed by each non-compositional MDC do not provide a general way to deal with all or most of these expressions. So whatever the cost of introducing these tools, it is also incurred by models which adopt non-compositional MDCs. In addition, the non-compositional MDCs, but not CO, also introduce other technical tools—non-compositional, construction specific SRs—which have problematic consequences for language acquisition.

2.6.5 Objection 4: Why not take CO as just a default or cognitive bias?

An interesting response to the claim that CO is more plausible than any of the non-compositional MDCs is to propose a middle-ground. The idea is that we interpret CO as a default or cognitive bias of FL, such that under certain conditions language learners can drop this default and acquire non-compositional construction-specific SRs. There are various ways of implementing this proposal. For example, we can formulate a 'mixed' MDC, call it '$CO(LOC)$', that has CO as a default and LOC as a secondary constraint on SRs. Following the same recipe, we can construct a MDC, call it '$CO(GEN)$', that has CO as a default and GEN as a secondary constraint on SRs. To fully determine if mixed MDCs with a compositional default are more plausible than CO would require an extensive discussion; here I will only briefly explain why it seems unlikely.

The challenge for mixed MDCs is to specify the properties of the learning data that would trigger the use of the non-default part of the constraint, i.e., the search for a non-compositional SR. The triggering conditions would have to include data that can be reliably used to infer that a compositional SR does not output the correct meanings for tokens of some type of complex expression, e.g., for tokens of simple meteorological expressions or extensional feedback complex NPs. For concreteness, let us focus on $CO(LOC)$ (the points I make apply equally well to $CO(GEN)$). The point of proposing that we replace CO with $CO(LOC)$ is to allow speakers to acquire construction-specific non-compositional SRs that assign location restrictions to meteorological expressions, while keeping a general bias for compositional SRs for at least most other types of complex expressions. So the triggering conditions have to meet two requirements: (i) token assertions of meteorological expressions trigger the search for a non-compositional SR, and (ii) token assertions of complex expressions that do not have a location restriction do not trigger the search for a non-compositional SR (which would explain why we don't observe patterns of mistakes like those in (2.11)-(4.13)). What could the triggering conditions be?

One might think that the search for a non-compositional SR should be triggered when S concludes that the meanings of complex expressions of some type seem to involve location restrictions, and that these meanings do not match the compositionally determined (default) meanings, which do *not* involve location restrictions. Things are not so simple, however, for data about the meanings of complex expressions, which comes mostly from linguistic interchanges, is usually noisy and plagued with mismatches between the compositionally determined and the asserted content of tokens of complex expressions. Speakers have tools to deal with these mismatches without having to revise the relevant default compositional SRs (see Objection 3 above). For example, in the case of assertions of meteorological expressions, speakers could assume that the location restrictions are due to primary pragmatic effects, such as the free-enrichment of the meanings of tokens of lexical items such as *rains* and *snows*. The question then is this: under what conditions should language learners resolve the mismatches between the asserted and the compositionally determined meanings of tokens of complex expressions of some type without using any of these available tools, but by revising instead the relevant construction-specific SRs?

This question reveals the problem with $CO(LOC)$ and other mixed MDCs with a compositional bias: simply put, all the ways of specifying the triggering conditions either render the mixed MDCs superfluous, or entail incorrect patterns of linguistic development similar to those entailed by pure non-compositional MDCs. On the one hand, if the conditions that trigger the search for a construction-specific, non-compositional SR are too demanding, this search would never begin. This is because, as we just said, most cases of mismatches between the asserted and the compositionally determined meanings of tokens of complex expressions of some type (including those involving meteorological expressions) can be resolved *without* revising the default compositional SRs, e.g., by revising instead the relevant lexical entries or factoring-in systematic primary pragmatic effects. In this case, replacing CO with a more complex mixed MDC such as $CO(LOC)$ would be an entirely superfluous theoretical move. On the other hand, if we weaken the conditions that trigger the search for a non-compositional SR, this search would be triggered by mismatches (involving location restrictions) between the asserted and the compositionally determined meanings of tokens of expressions of various types, in addition to meteorological expressions. In this case, we would expect to find some speakers that test 'incorrect' SRs such as those in (2.11)-(4.13). For if we assume these weak triggering conditions, why would speakers resolve location restriction mismatches by revising the relevant SR in the case of meteorological expressions but *not* in the case of other types of expressions such as those in (2.11)-(4.13)? However, as we argued before, there is not much, if any, testing of non-compositional construction-specific SRs going on in early language acquisition, contrary to what would be predicted by mixed MDCs such as $CO(LOC)$ when paired with weak conditions that trigger the search for non-compositional SRs.

2.6.6 Objection 5: Why can't MDCs be learned?

We argued that we should understand MDCs as innate constraints on the functional architecture of FL, specifically, as innate 'over-hypotheses' on SRs, i.e., constraints on the general form of the SRs which FL is cognitively capable of representing. But are we really forced to hold that MDCs are innate?

Learning from experience requires some innate constraints. In particular, FL must have some innate constraints which help speakers learn the semantics of target Ψ-languages. However, maybe the innate constraint on FL is more abstract than particular MDCs such as CO, LOC, or GEN, and is more like an 'over-over-hypothesis' which constrains possible MDCs in a way analogous to the way in which MDCs constrain possible SRs. This more abstract constraint, call it '$O(MDC)$', could constraint possible MDCs so that each allows only SRs that, for any type of complex syntactic structure $[_Z X\ Y]$, determine $[\![[_Z X\ Y]]\!]$ as a function of the meaning of its immediate constituents {X, Y}, and *possibly* something else. Although weak, $O(MDC)$ does restrict the set of possible MDCs. Speakers would still have to learn which of CO, LOC, GEN, etc., is correct for the target Ψ-language. If constrained only by something like $O(MDC)$, then early in linguistic development S could consider, among others, SRs such as:

(2.21) $[\![[_Z X\ Y]]\!] = f_Z([\![X]\!], [\![Y]\!])$

(2.22) $[\![[_Z X\ Y]]\!] = f_{Z'}(f_Z([\![X]\!], [\![Y]\!]), b)$

(2.23) $[\![[_Z X\ Y]]\!] = f_{Z''}(f_Z([\![X]\!], [\![Y]\!]), g)$

where 'b' stands for a set of extensional feedback and 'g' for a location function. According to this view, when testing hypotheses like (4.32)-(2.23) against the data (i.e., subsets of the language), S not only selects the best fitting SR, but at the same time, and prior to acquiring other particular SRs, S also selects the best fitting MDC. For example, if S determines that (4.32) generates the correct meaning assignments for tokens with the form $[_Z X\ Y]$, this in turn suggests to S that the MDC is probably closer to CO than to LOC or GEN—assuming that S applies a learning mechanism which selects the logically strongest and simplest MDC, among those allowed by $O(MDC)$, which is consistent with the selected SR.

This sketch of how MDCs could be learned is especially interesting because in other cognitive domains similar learning processes—of acquiring over-hypotheses of the sort usually assumed to be innate by nativists—have been modeled using Hierarchical Bayesian Models (HBMs) (Kemp et al. [87]). To be clear, the reason we suggested that MDCs are likely innate is *not* that, given developmentally plausible subsets of the target language and any powerful domain-general learning mechanism, it is otherwise impossible to explain how someone could acquire a target Ψ-language such as Ψ-English; the reason is that there seems to be no learning of the relevant sort going on in actual language acquisition—i.e., no learning of MDCs and even of construction-specific SRs. Of course, this does not conclusively show that although MDCs could in principle be learned they in fact are innate. For without considering specific HBMs, it is very hard, if not impossible, to determine whether the sorts of patterns Bayesian learners would have to go through to acquire MDCs—which given certain ways of modeling the problem, could be quite minimal—are consistent with the general patterns of early linguistic development. As far as I know, no HBMs that can acquire particular MDCs have been tested, although in principle such models can certainly be constructed.[23] Still, even if we can construct cognitively plausible HBMs that can acquire MDCs and so seriously consider the view that MDCs are acquired early in language acquisition, HBMs have some general properties which suggest that this result would be consistent with the main claim defended here—namely, that CO is the MDC on Ψ-languages.

[23] There are two models that might be thought to bear on this issue. The first models the cultural evolution of natural languages, and shows that 'compositional' languages would be selected over 'non-compositional' languages (Kirby [181]). This model is not relevant to our problem—whether a Bayesian learner, given developmentally plausible bits of the language, would acquire CO over the non-compositional MDCs—because the 'non-compositional' languages considered by the model are implausible extremes, e.g., they do not satisfy any of the non-compositional MDCs considered here (they are even more unconstrained). The second model acquires a 'compositional semantics' (Tenenbaum et al. [133]). However, the main task faced by this model is only to pair simple expressions with lambda-types. The model is set-up to calculate the meaning of all complex expressions in the same general way, compositionally via FA. So this model's functional architecture instantiates CO—it can't even represent non-compositional SRs.

We can illustrate this last point by considering in broad outline a hypothetical HBM ('HBM_L') that, constrained by an $O(MDC)$, acquires a particular MDC in the processes of acquiring the semantics (i.e., the SRs) of Ψ-English. We know something about the general conditions that would have to be met by 'HBM_L'. The process of acquiring a target Ψ-language requires MDCs of some sort; so HBM_L has to select a MDC early in the process of acquiring the full set of SRs, even if this selection is later revised. This is not a problem. As we just said, HBMs can be set-up to acquire the relevant over-hypotheses before they acquire most of the specific lower-level hypotheses. Now, assume that, given some subset E_1 of English, HBM_L selects the MDC from the set of CO, LOC and GEN with the highest conditional probability. HBM_L computes that as a function of the prior probability of each MDC—$P(CO)$, $P(LOC)$ and $P(GEN)$—and of the likelihood of E_1 given each MDC—$P(E_1|CO)$, $P(E_1|LOC)$ and $P(E_1|GEN)$. HBMs assign the highest prior probability to the simplest over-hypotheses. CO is the simplest MDC, since it has the fewest free parameters (for the same reason, compositional SRs are simpler than non-compositional SRs), so $P(CO) > P(LOC)$ and $P(CO) > P(GEN)$. The likelihoods partly depend on what is included in E_1. We are modeling the earliest stages in the acquisition of Ψ-English, when speakers are beginning to learn how to determine the meaning of very simple complex expressions, so E_1 includes very simple complex expressions such as *red ball*, *green apple*, and *daddy away*, each paired with a representation of a stereotypical exemplar or situation (Pinker [139]). For this reason, it is safe to hold that the meanings of tokens of complex expressions included in E_1 can be generated using a compositional SR, such as (4.32) above. Now, compositional SRs such as (4.32) are compatible with CO, LOC and GEN. But since LOC and GEN also generate other SRs, e.g., (2.22) and (2.23) respectively, $P(E_1|CO) > P(E_1|LOC)$ and $P(E_1|CO) > P(E_1|GEN)$.

It follows that:

(2.24) $P(CO|E_1) > P(LOC|E_1)$ and $P(CO|E_1) > P(GEN|E_1)$

such that:

(2.25) $P(CO|E_1)/P(LOC|E_1) > P(CO)/P(LOC)$ and
$P(CO|E_1)/P(GEN|E_1) > P(CO)/P(GEN)$

Informally, (2.25) tells us that, although HBM_L is initially biased to favor the simpler CO over the non-compositional MDCs, HBM_L favors CO even more strongly after processing E_1. Furthermore, once there is a strong initial bias for CO, this selection would likely remain stable through the rest of the process of acquiring Ψ-English, essentially for the reasons given in response to Objection 4. So even if we hold that MDCs are acquired in early language acquisition, it seems quite likely—given that HBMs are currently our best idea of how this process could work—that the MDC selected by language learners would be more like CO than like any of the non-compositional MDCs.

2.6.7 Open Issue: Implications for Constructionist Approaches

Constructionist approaches to language have been gaining popularity (Smith [182], Hoffman and Trousdale [74]). A full discussion of the implications of our argument for CO for Constructionist approaches is outside the scope of this essay. Still, the issue merits some preliminary discussion, in part because it might seem that Constructionist approaches are, on the one hand, in tension with some of the basic assumptions we made about FL, and, on the other, undermined by the arguments against non-compositional MDCs. However, I will briefly explain why the implications are more nuanced and interesting.

What sets apart Constructionist approaches from mainstream Generative approaches is their emphasis on phrasal constructions in language acquisition. Like traditional lexical items, phrasal constructions are *learned* pairings of form and function (for an overview, see Goldberg [59]). Different types of phrasal constructions are associated with different types of functions. In our terminology, this means that phrasal constructions are associated with particular construction-specific SRs. According to Constructionist views, for speakers to determine the meaning of a complex expression of a certain phrasal type, it is not enough that they know its form and the meanings of its parts; they must also know which construction-specific SR is associated with the phrasal type. This entails that Constructionists implicitly reject the strong version of CO according to which SRs must be general.

We should not conclude from this that our approach and Constructionist approaches are incommensurable.We can straightforwardly frame a version of the compositionality debate in a Constructionist framework. But to do that we have to make an important modification: the compositional MDC would have to be a weak version of CO which allows construction-specific SRs, but only compositional ones. The competing non-compositinal MDCs could still be formulated basically like LOC, GEN, and its variants, since these allow learned pairings of phrasal types with particular SRs. What distinguishes the non-compositional MDCs from weak CO is that only the former allow non-compositional SRs.[24]

Once framed in this way we can see that the criticism based on unattested

[24]Some of the literature that contrasts Constructionist with mainstream Generative approaches tend to characterize the former as non-compositional (see e.g. Smith [182]: p. 380). What I think they usually mean when they make those remarks is that Constructionists cannot accept that FL represents only general SRs, hence they cannot say that, in all cases, the meaning of a complex expressions is determined by the meaning of the parts and their structure. But clearly Constructionists can in principle accept a compositional view, as long as this view allows construction-specific compositional SRs. In this case the position can be expressed by saying that the meaning of a complex expressions is determined by the parts, their structure, and the function associated with that structure.

patterns of mistakes such as (11)-(13) and (18)-(20) would apply straightforwardly to a Constructionist approach that assumes one of the non-compositional MDCs, but not to one that assumes a weak version of CO. For although weak CO allows construction-specific SRs, it does not allow non-compositional construction-specific SRs, which blocks cases like (11)-(13) and (18)-(20).[25] Non-compositionalist Constructionists cannot respond that the phrasal form-function pairings are innate, since they are committed to these pairings being learned (Goldberg [58, 59], Tomasello [200]). In addition, our argument didn't make any strong assumptions about the learning mechanism responsible for acquiring SRs (e.g., they can be domain general), so there is no obvious assumption there to reject (for discussion see footnote 8). Finally, although most Constructionist are anti-nativists (Goldberg [58,60], Tomasello [200]), this does not force them to deny that, unlike particular SRs, MDCs are plausibly innate; but if they do deny that, then the remarks made in Objection 5 above—regarding why Bayesian learners would tend to acquire a compositional over-hypothesis—directly apply to this case.

In short, our argument for CO is not in tension with all Constructionist approaches. Accepting a Constructionist approach does require replacing a strong version of CO with a weaker version. Aside from that, none of the further assumptions we made about FL or language acquisition to defend compositional MDCs are inconsistent with Constructionist approaches per se. But once properly modified, the argument for CO supports the view that we should adopt compositional over non-compositional Constructionists view.

[25]This is not to deny that our approach is more in tune with mainstream Generative approaches. For example, some of the considerations we presented in favor of compositionality tend to support a compositional MDC which only allows general SRs. What I am arguing here is that, if for other reasons we favor a Constructionist approach, we can frame a version of the compositionality debate within this approach. Once we do that, we can see that our previous argument favors weak compositional over non-compositional Constructionist approaches.

2.7 Conclusion

Debates about the non/compositionality of FL seem to reach a standstill when we acknowledge, with recent critics, that there are non-compositional MDCs that can account for $P\&S$. To resolve this standstill, we first argued that we should frame these debates as debates about which MDC is the most plausible functional constraint on the semantics of FL, specifically, on the allowed forms of SRs which FL can represent. We then saw that each non-compositional MDC involves a weakening of CO, the point of which is to make FL compatible with some adequate non-compositional (construction-specific) SRs. However, theorists generally fail to notice that each weakening also makes FL compatible with many other incorrect SRs. As a result, if FL was constrained by these non-compositional MDCs, speakers would, in the course of early linguistic development, have to test and reject at least some reasonable but incorrect construction-specific non-compositional SRs. This predicts patters of early linguistic development which actual speakers never seem to go through. In contrast and more consistent with actual linguistic development, CO predicts patterns of development that do not involve any testing of reasonable but incorrect construction-specific SRs. We also considered some seemingly plausible additional constraints on the non-compositional MDCs which seek to constrain the search space of construction-specific SRs during acquisition. The two main proposals are that construction-specific SRs are innate and that non-compositional MDCs are the non-default options of complex MDCs with compositional defaults. We showed that none of these moves saved the non-compositional MDCs without giving rise to other unacceptable problems. This strongly suggests that the MDC of FL is closer to CO than to any of the non-compositional MDCs currently on offer. Admittedly, this argument for CO is not a wholly general argument that ranges over all conceivable non-compositional MDCs combined with all conceivable additional constraints. This is why it is important that the particular non-compositional MDCs and additional constraints which we examined are currently the most plausible, motivated and popular. Perhaps more importantly, this overall approach—which focuses the implications of MDCs on patterns of language acquisition via their constraints on the SRs which FL can represent—can be used to evaluate future proposals for MDCs.

Chapter 3

The Productivity and Systematicity of *FL*

As I argued in Chapter 2, most theorists who debate whether FL is compositional agree that FL is a recursive computational system part of whose function is to compute not only the phonetic and syntactic structure of expressions but also their meaning. The reason why parties to this debate agree on these general properties of FL is that they also agree that FL satisfies PRODUCTIVITY and SYSTEMATICITY ($P\&S$), in the sense specified at the beginning of Chapter 2. However, as I also mentioned there, some prominent theorists question whether our linguistic competence really satisfies $P\&S$. In their most radical form, these skeptical arguments threaten to undermine the basic assumption of our approach: that we should model FL as a recursive computational system that generates not only the phonetic and syntactic structure of expressions but also their meanings. The task of this chapter is to address the most important objections which have been raised against P&S.

3.1 Introduction

The arguments against the view that FL is compositional addressed in Chapter 2 all assume that FL satisfies $P\&S$. For most theorists, this is what makes those arguments against compositionality especially persuasive: non-compositional MDCs such as LOC and GEN can account for $P\&S$. In what follows, we will consider more skeptical challenges that, if successful, would seriously question the assumption that FL computes the meaning of expressions. These skeptical arguments question whether our linguistic competence satisfies $P\&S$ in the way assumed by theorists who debate whether or not FL is compositional. To address these challenges, it is useful to briefly consider an outline of the previous argument for the compositionally of FL. 'FL_S' stands for the FL of an arbitrary competent speaker S_1 of English. As we argued in Chapter 2, the argument doesn't appeal to any particular properties of S_1 or English, so it can be generalized to other speakers of English and other natural languages.

(PRODUCTIVITY) FL_S can generate *correct* interpretations (relative to English) for complex expressions which S_1 has never encountered before. FL_S has this capacity for indefinitely many distinct complex expressions, generating a distinct meaning for an indefinite number of these expressions.

(SYSTEMATICITY) The generative capacity of FL_S is structured in the following way: if it can generate *correct* interpretations (relative to English) for complex expressions $e_1, \ldots, e_n$, it can generate correct interpretations for all other complex expressions constructed from: (i) constituents of $e_1, \ldots, e_n$ and (ii) syntactic structures employed in any of the complex expressions $e_1, \ldots, e_n$.

(A-1) There must be some cognitive computational state and capacity of FL_S

on the basis of which it can generate such (indefinitely many) correct (relative to English) novel $< LF,\ \text{meaning} >$ pairs.

(A-2) FL_S could not have acquired the information required to generate each of the novel $< LF,\ \text{meaning} >$ pairs by memorizing the meaning of each LF indepedently—i.e., in the list-like manner in which we memorize a finite lexicon.

(WEAK CONCLUSION) What explains why FL_S satisfies $P\&S$ is that: (i) FL_S has access to information about the structure of complex LFs and the meaning of their constituents; (ii) FL_S has the computational capacity to use that information to generate the meanings for the complex LFs; (iii) conditions (i) and (ii) constrain the way in which FL_S can generate meanings for complex LFs; (iv) conditions (i), (ii) and (iii) are sufficient to generate the *correct* (relative to English) meanings for complex LFs.

(STRONG CONCLUSION) There are other explanations of $P\&S$ that assume that the MDC of FL_S is non-compositional, such as LOC and GEN. However, unlike WEAK CONCLUSION (i.e., unlike assuming compositionally), the non-compositional MDCs predict incorrect patterns of early linguistic development. So compositionality, in the sense specified in WEAK CONCLUSION, is the best explanation of $P\&S$.

This argument for compositionality is a straightforward inference to the best explanation. The starting point of most recent debates is that the traditional argument only warrants WEAK CONCLUSION. In Chapter 2 we presented a detailed defense of STRONG CONCLUSION. However, all the premises previous to WEAK CONCLUSION have also been questioned, especially $P\&S$. We now turn to a detailed defense of those premises. Still, from the outset, it is important to note that we formulated $P\&S$ differently from the way they are usually formulated, mainly because they explicitly appeal to the notion of 'correct interpretations relative to English'. That something like this latter notion is part of the datum is obvious. What needs to be explained is *not* that FL_S can generate *some* interpretation for an unbounded number of complex expressions of English, for example that it can map *brown cat* to BROWN CAT or to GRAY CAT or to YELLOW FELINE; what needs to be explained is that FL_S generates *the* English interpretations.

3.2 PRODUCTIVITY

Most theorists who attack the view that FL satisfies P&S focus on SYSTEMATICITY ('S'). It is not hard to see why. PRODUCTIVITY ('P') only states that FL_S can generate the correct meaning for an unbounded number of novel expressions, whereas S imposes a particular structure on this unbounded capacity. In other words, S is a stronger and more interesting claim since it specifies the main (structural) relation involved in productive capacities between the expressions we understand and the unbounded number of novel ones that we have the capacity to understand. We will address objections against S in §3.3. Here we focus on 'P'. To begin this discussion, it is useful to contrast two formulations of productivity, the first, OP, is closer to observational evidence, but the second, P, is the notion we need to defend:

(OP) S_1 can correctly use complex expressions of English which S_1 has never encountered before, including indefinitely many distinct complex expressions each with a distinct use.

(P) FL_S can generate *correct* interpretations (relative to English) for complex expressions which S_1 has never encountered before. FL_S has this capacity for indefinitely many distinct complex expressions, generating a distinct meaning for an indefinite number of these expressions.

Clearly, no one would doubt OP, but some theorists might still be skeptical or agnostic towards P. Still, we need something like P. For OP says something about S_1's linguistic use, whereas P says something about S_1's FL. OP is an observational generalization, but is not directly about FL_S; P is more removed from observational evidence, but is directly about FL_S.

In what follows, I will argue that we should accept P, by defending two claims: (i) that OP is clearly correct, and (ii) that OP gives us good reason to accept P. Interestingly, we can defend these claims in a way that allows for a wide range of accounts of the nature of linguistic meaning.

OP says that S_1, an arbitrary competent speaker of English, can use indefinitely many complex expressions of English which S_1 has never encountered before.[1] This claim is obviously correct. Now, by saying that S_1 can 'use' a novel complex expression e of English, we do not mean that S_1 can use e appropriately in every single way in which it can be used, but only that S_1 can, minimally, use e in some canonical way, e.g., that S_1 could assert it in some canonical circumstances. When we use the concept of 'understanding' in a use-theoretic way, as we often do in ordinary discourse, this is often what we mean when we say that

[1] The sense of the unboundedness of our linguistic competence here assumed is basically the same as that assumed by Fodor and Pylyshyn [51] in their classic (1988) article:

> It goes without saying that no one does, or could, in fact utter or understand tokens of more than a finite number of sentence types; this is a trivial consequence of the fact that nobody can utter or understand more than a finite number of sentence tokens. But there are a number of considerations which suggest that, despite de facto constraints on performance, ones knowledge of ones language supports an unbounded productive capacity in much the same way that ones knowledge of addition supports an unbounded number of sums. Among these considerations are, for example, the fact that a speaker/hearers performance can often be improved by relaxing time constraints, increasing motivation, or supplying pencil and paper. It seems very natural to treat such manipulations as affecting the transient state of the speakers memory and attention rather than what he knows about—or how he represents—his language (Fodor and Pylyshyn [51]).

speakers can 'understand' an indefinite number of novel complex expressions.[2]

The important question then is whether we can infer P from OP. Although the validity of this inference might seem to depend on the account of 'meaning' paired with P, I will argue that the inference holds for most accounts of linguistic meaning. There is a general reason for this. Most theories of linguistic meaning that aim to be part of theories of linguistic competence accept the USE-MEANING CONSTRAINT ('UMC'):

(UMC) A theory of meaning is minimally adequate only if it can be wedded to an account of why S_1's generating the meaning of e is an essential part of the explanation for S_1's ability to use e.

To hold UMC is to hold that, in general, if S_1 can use e then S knows the meaning of e, among other things. In other words, UMC entails that to use an expression requires, in general, that we grasp that expression's meaning, even if full use might also require other capacities, including non-linguistic ones. Most theorists accept UMC for a good reason. Prima facie, a theory of meaning which rejects UMC undermines the central justification for postulating the theoretical level of linguistic meaning: i.e., that what we are ultimately trying to explain is our capacity to *use* natural languages.[3]

It is not hard to see why, for all theories of meaning which respect UMC, the inference from OP to P seems safe. For theories which accept UMC, to appropriately use an expression involves, minimally and in general, grasping

[2]There are many examples of this use of the word 'understanding'. Suppose that S_1 asserts *is an apple* of apples, except in one case, where S_1 says of an apple looking pear *that is an apple*. In this case, most competent speakers would still be willing to say that S_1 understands *is an apple*, although perhaps they would not make that attribution if S_1 refuses to accept that prototypical looking apples are *apples*, unless there is some reasonable explanation for such refusal. Such examples suggest that the ordinary notion of understanding distinguishes between uses of e which are central for attributing understanding of e and non-central uses.

[3]For a forceful defense of these points see Dummett [39]: 1-202. As is well known, Dummett is also very sympathetic to the idea that use-theoretic accounts of meaning are more successful than truth-theoretic accounts of meaning in satisfying the demands of the UMC. Given our purposes, we can remain neutral on this controversial issue.

its meaning, and possibly more. So to accept that S_1 knows how to use novel expressions, including indefinitely many novel expressions, each with a distinct use, is to accept that S_1 has the capacity to determine the meaning of most of these indefinitely many novel expressions. In other words, assuming UMC and OP, we can, a fortiori, infer P.[4]

Unfortunately, there are some complications to this story. Accounts of linguistic meaning also have to meet other theoretical constraints. This might force theorists to enrich the notion of meaning such that knowing the meaning of e becomes, in certain crucial respects, more demanding than using or understanding e, with the result that we cannot infer from an instance of the former an instance of the latter. This point is pressed by Szabo [187].

Following a venerable tradition in philosophy of language, Szabo [187] suggests that theories of meaning are also constrained by what he calls 'the second dogma about meaning' ('SD'):

(SD) Whatever meaning is, it must have the property that the meaning of an expression, given the way the world is, determines its referent.

To properly discuss the relation between SD and UMC, let us introduce some terminology. 'Cognitive content' stands for what one has to grasp in order to

[4] That accepting UMC ensures the validity of the inference from OP to P is easy to see in the case of use-theoretic accounts of meaning. For use-theoretic accounts construe 'S_1 knows the meaning of e' in terms of some type of basic use of ee.g., to know under what conditions e can be asserted, or what you can infer from asserting e, or a combination of both. In any of these cases, the uses of e which constitute 'knowing the meaning of e' are a subset of the uses which speakers can perform with e. So if OP holds, then, a fortiori, P also holds.

Although less directly, this same conclusion holds for theories which, unlike use-theoretic accounts of meaning, are based on a principled distinction between linguistic performance and linguistic knowledge/competence, like most truth-conditional and computational theories. For if such theories accept UMC, then the reason why they posit a level of linguistic competence is to explain linguistic performance. The way in which we characterize, from this perspective, what is involved in knowing how to use e is by saying that it involves knowing the meaning of e, where meaning is construed in a truth-theoretic way, although it might also involve other performance capacities, some not specifically linguistic. So if we have evidence that S_1 knows how to use novel expressions, including indefinitely many novel expressions, each with a distinct use, then we have evidence that S_1 has the capacity to determine the meaning of these novel expressions, including that of indefinitely many novel expressions, each with a distinct meaning. This entails that if we can attribute OP, we can, a fortiori, attribute P.

use or understand an expression; and 'referential content' for what determines, given the way the world is, an expression's referent. Thus far we have bridged the gap between use or understanding and meaning by holding the following view, which is a consequence of UMC: to understand e, S has to generate, at least, the meaning of e. If we combine this view with SD, we get: to understand e, S has to generate, at least, the referential content of e.

Szabo [187] raises some doubts about this last claim. He argues that it is not true that, in general, whatever one has to grasp in order to understand an expression determines that expression's referential content. In other words, understanding sometimes underdetermines referential content, and referential content is part of meaning; so understanding sometimes underdetermines meaning.

Szabo's argument is based on a certain reading of Putnam's twin-Earth thought experiments. Suppose S_1 and twin-S_1 are doppelgangers, specified in the usual way, and that the only difference between Earth and twin-Earth is that water on Earth is H_2O and on twin-Earth is XYZ. S_1 and twin-S_1 have the same linguistic competence, individualistically described, and, let us assume for now, the same psychological states. What each grasps or generates in understanding *water* is determined by their equivalent psychological states. Since S_1 and twin-S_1 understand *water* in the same way, they associate the same cognitive content with *water*. At this point, it seems that we can still assume that cognitive content determines referential content. Given earth, this cognitive content determines that *water* for S_1 refers to H_2O. Given twin-earth, this cognitive content determines that *water* for twin-S_1 refers to XYZ. So far so good: a different world was given in each case, and the cognitive content for *water* determined a different referent for each.

But here's the problem. Suppose twin-S_1 comes to Earth and asserts, pointing to a typical example of water, *that is water*. The intuition, according to

Szabo, is that twin-S_1 is then saying something false, although in the exact same context S_1 would be saying something true. This intuition holds despite the observation that, by assumption, S_1 and twin-S_1 associate the same cognitive content with *water*.[5] So cognitive content (the grasping of which is all that is required for understanding), which is the same in this case, does not determine referential content. Since referential content is part of meaning (according to SD), it follows that S_1 and twin-S_1 can use and understand *water* without thereby grasping or generating its full meaning.

Szabo's interpretation of the thought experiment forces us to reject UMC, and so threatens the inference from OP to P. However, his interpretation of the thought experiment is inconsistent. To see this, consider two key assumptions made by Szabo's account: (1) S_1 and twin-S_1 have the same psychological states, and (2) SD is non-negotiable. Szabo assumes that we can consistently hold both (1) and (2). But that is an unacceptable assumption. As we will now see, the two most general (and plausible) philosophical views on the nature of mental content and linguistic meaning—externalism and internalism—are compatible only with either (1) or (2), but not both.

Consider first externalism. According to externalism, relations to the world are essentially involved in the individuation of mental/psychological states. Since S_1 is related in the relevant way to water and twin-S_1 to XYZ, S_1 and twin-S_1 are not in the same psychological states. So externalists are forced to reject (1) (to be fair, Szabo also points this out in his discussion of possible responses to the thought experiment). Furthermore, since psychological states determine cognitive content, externalists have to deny that S_1 and twin-S_1 asso-

[5] As is well known, some theorists reject the basic intuition of this thought experiment. Other theorists accept the intuition, but argue that variations of this thought experiment result in conflicting intuitions, so that we should not take the intuitions too seriously in theorizing (see e.g., Chomsky [25]). I don't disagree with these positions. Here I accept the basic intuition of the thought experiment for the sake of argument, esp., because many philosophers accept it. What I try to show is that doing so doesn't undermine UMC.

ciate the same cognitive content with *water*. So if Szabo's thought experiment is analyzed within an externalist framework, it does not threaten the idea that cognitive content determines referential content, which makes it compatible with UMC.[6]

Consider next internalism. Given the set up of the thought experiment, internalists are forced to accept (1); however, they should and usually do reject (2), i.e., they abandon the idea that cognitive content always determines referential content. Now, this position entails neither that cognitive content has nothing to do with referential content, nor that cognitive content never determines referential content. In at least some cases, meaning, or cognitive content, can constrain referential content, but, even given the way the world is, it doesn't fully determine it (see Chomsky [25]; Pietroski [134]).

Consider a possible internalist account relevant to the thought experiment. Assume that the meaning of *water* for S_1 includes a prototype and some rule which says that things fall in its extension if they are related in way R to things that clearly fall under the prototype in some setting M. Suppose also that the specific relation R and the setting M are context-sensitive, within certain limits. One way to specify the relation R, which would be attractive to many theorists (for the particular case of natural kind terms), is this: only things made of the same stuff as things that fall under the prototype in setting M are in the extension of *water*. In this case, the way in which we fill the meaning of *water* in Earth and twin-Earth is different because the setting M is different for each case: (simplifying) in Earth H_2O and not XYZ falls under the prototype, and in twin-Earth only XYZ and not H_2O falls under the prototype. Once we specify values

[6] As I argued in Chapter 2, externalism about mental content and linguistic meaning is compatible with the idiolectical conception of natural languages adopted here. FL is a cognitive computational structure in the minds of individual speakers. This position is compatible with the view that these structures, or certain aspects of them, are externally individuated. For an illuminating defense of the view that it is possible and desirable to reconcile an idiolectical conception of language with externalism about mental and linguistic content, see Higginbotham [72].

for R and M we specify a referential content for *water*, but the point is that such specification is not required to grasp the cognitive content of *water*. Although S_1 and twin-S_1 associate the same cognitive content with *water*, and hence the same meaning, their tokens of *water* have a different referential content, because the parameter M was given a different value in each case. According to this view, meaning, which is cognitive content, plus fixing context-sensitive parameters, and given the world, determines extension. So meaning plus fixing context-sensitive parameters determines referential content.

Szabo [187] thinks that this internalist position is unavailable because he insists that we cannot give up on SD, i.e., on the idea that meaning includes referential content:

> Notice that what forces the second dogma on us is that we regard any two genuine names that refer to different objects, any two one-place predicates that have different extensions, and in general, any two expressions that have different referents as non-synonymous. What this means is that given the way the world is, two meaningful expressions that are distinguishable in terms of their referents must be distinguishable in terms of their meanings. This is equivalent to the second dogma, if one reads it as a strong supervenience claim (Szabo [187]: 43).

The internalism is in tension with the view of synonymy espoused by Szabo in that passage. For according to internalism, two expressions, relative to the same world, might have different extensions and still have the same meaning, if meaning is just cognitive content. This is because the difference in extension, given the same world, might be accounted for by a difference in the ways in which the context-sensitive parameter(s) of each expression are filled in, just like we saw in the case of *water* as used by S_1 and *water* as used by twin-S_1. So we can't deny that this account issues in a view of sameness of meaning that is technical, and at least in some cases, one that doesn't exactly correspond to our pre-theoretical notion of sameness of meaning. This consequence of the view might be intolerable, but Szabo [187] doesn't give any reason to think that it is, and I can't think of one.

On the other hand, internalism has the resources to explain why we have such pre-theoretical intuitions. Here is a brief sketch of how such an explanation might go. When we are asked in a pre-theoretical setting whether two expressions are synonymous we often implicitly assume their use in some general or usual context, with their context-sensitive parameters filled in the way that is common in that context. Our intuitions are about content after these saturations (which we informally call 'meaning'), so they are about referential content. This is why we assume that if two expressions, given the same world, have a different referent, then their meaning (broadly construed in informal settings to include referential content) cannot be the same.

To conclude, it seems clear that if we adopt either externalism or an internalism about linguistic meaning, Szabo's thought experiment does not undermine UMC, hence does not threaten the inference from OP to P. Now, it might seem that we can accept this result and still doubt whether this covers all the relevant cases needed to properly establish the inference from OP to P. But recall what we have shown: the inference from OP to P is correct, whether you are an internalist or externalist about linguistic meaning. So what about its status if we assume other positions? We have covered all the plausible positions on the metaphysics of linguistic meaning currently on the market, which is good enough for our purposes.[7]

[7]Of course, we could be anti-realists about linguistic meaning, but in this case when we say either that our grasp of meaning if productive or systematic or compositional, we are saying something that is strictly speaking either false or empty.

3.3 SYSTEMATICITY

We now turn SYSTEMATICTY ('S'), which is the more interesting and controversial premise. As should be clear from the outline of the argument, S attributes a certain structure to P:

(S) The generative capacity of FL_S is structured in the following way: if it can generate *correct* interpretations (relative to English) for complex expressions $e_1, \ldots, e_n$, it can generate correct interpretations for all other complex expressions constructed from: (i) constituents of $e_1, \ldots, e_n$ and (ii) syntactic structures employed in any of the complex expressions $e_1, \ldots, e_n$.

Suppose you have two sets, A and B, of complex LF expressions such that B can be obtained from set A in the following way:

- If e belongs to A then e belongs to B.
- If e is an expression with constituents which are also constituents of expressions of A, and e is structured using only syntactic structures used in expressions of A, then e belongs to B
- No other expressions belong to B

When a set B can be obtained from a set A in this way, we say that B is a 'systematic variant' of A. S entails that if FL_S can generate interpretations for the complex LF expressions of A, then FL_S has the competence to generate interpretations for the 'systematic variants' of A, which include B.

Note that many sets of complex expressions (of natural languages) have the property that competent speakers who can generate their interpretations can also generate the interpretations of the expressions in their systematic variant sets.[8] To my knowledge, no one denies this. In addition, we can allow a finite

[8]In case this point is not prima facie obvious, consider some examples. If S_1 can generate interpretations for the set which is constituted by *John likes brown dogs* and *John likes black*

number of exceptions to S—as long as there is no systematic way to generate them. Overall, S seems at least roughly correct.

However, some prominent theorists recently argue that, once we look beyond some often cited examples, there are many sets of common types of complex expressions such that competent speakers can interpret them but do not seem to have the linguistic competence to interpret some of the expressions in their systematic variant sets. Despite initial appearances, these theorists argue, FL_S does not satisfy S. In what follows, we will consider two objections against S which try to establish that point—one based on certain adjectival modifiers and another one based on certain verb-phrases. The objections we will consider are representative; by responding to them we will achieve two crucial things. Firstly, we will have a better understanding of S. Secondly, we will have a general understanding of why similar types of objections to S are unpersuasive.

Consider first the objection based on certain adjectival modifiers. Assume S_1 can understand (3.1)-(3.2):

(3.1) John has a red car

(3.2) John has a tall building

The question is whether S_1 can then understand (3.3)-(3.4):

(3.3) John has a tall car

(3.4) John has a red building

In a passage where he raises some doubts about S, Szabo [193] suggests that from understanding (3.1)-(3.2) it doesn't seem to follow that S_1 can understand (3.3)-(3.4). At least, defenders of S have the burden of argument:

---

cats, then S_1 has the linguistic competence to generate interpretations for the expressions in its systematic variant set which is constituted by *John likes black dogs* and *John likes brown cats*. If S_1 can generate an interpretation for *Mary loves John*, then S_1 has the linguistic competence to generate an interpretation for its systematic variant set which contains only *John loves Mary*.

> One might argue that to understand *red building* one has to know which parts of a red building are supposed to be red, and to understand *tall car* one would need to know how to compare cars in terms of height. Neither of these is something one must obviously know in order to understand *red car* and *tall building* (Szabo [193]: 18).

Szabo's point is that to understand *red car* one has to know which parts of a car have to be red for a car to count as a *red car*, but to understand *red building* one has to know which parts of a building have to be red for a building to count as a *red building*. Likewise, to understand *tall building* one has to know how to compare buildings in terms of height for a building to count as a *tall building*, but to understand *tall car* one has to know how to compare cars in terms of height for a car to count as a *tall car*.

According to this suggestion, if we assume that S_1 understands (3.1)-(3.2) it follows that S_1 knows: (i) which parts of a car have to be red in order for it to count as a *red car*, and (ii) how to compare heights in order to determine whether a building counts as a *tall building*. But this knowledge is not enough to determine the meanings of (3.3)-(3.4), which contain the expressions *red building* and *tall car*. It follows that from understanding the set which consists of (3.1)-(3.2), S_1 doesn't have the capacity to fully understand its systematic variant set which contains (3.3)-(3.4). If Szabo's suggestion is correct about these cases, we have a general type of complex expression—instances of which are ubiquitous in all natural languages—that does not conform to S.

Szabo's argument only threatens S if its assumptions about what it takes to 'understand' expressions like *red car*, *red building*, *tall building*, and *tall car* are taken as assumptions about what it takes to understand their literal meaning. In other words, this argument assumes a particular stance on the semantics/pragmatics divide, which is based on the following two main assumptions[9]:

(I) The notion of 'understanding the literal meaning of e' should be construed as 'knowing the content of e'. For example, if e is a sentence, to understand its literal meaning is to know its truth-conditions or the proposition it expresses.

(II) Sentences such as (3.1)-(3.4) have truth-conditions that are specific in ways such as the following: (3.1) is true only if John has a car and the car is painted red in some particular way (e.g., on the outside), (3.2) is true only if John has a building and the building is above a particular height (presumably one that exceeds the average height of buildings), etc.[10]

The notion of (literal meaning) understanding expressed in assumption (I) is somewhat controversial. Also controversial are the truth-conditional assignments (literal content assignments) to expression like (3.1)-(3.4) expressed in assumption (II). That both claims are controversial is not, of course, an objection to Szabo's argument. Assigning literal meanings to expressions requires that we take a stance on the semantics/pragmatics divide. No stance on that divide is currently uncontroversial.

However, there is an interesting problem with this argument against S. Theorists who accept assumption (I) have to reject assumption (II), and vice-versa. In other words, there seems to be no plausible position on the semantics/pragmatics divide which would allow theorists to consistently hold both (I) and (II).[11].

[9]Strictly speaking, Szabo only needs to hold that these assumptions apply at the level of NPs.

[10]Different truth-conditions for (3.1)-(3.4) are compatible with Szabo's suggestion as long as they are determinate and specific enough such that the following holds: a car's way of being red is not the same as a building's way of being red; a car's way of being tall is not the same as a building's way of being tall; etc.

[11]I should note that, from other writings (see e.g. Szabo [190]), it seems that Szabo doesn't think that this argument against S is too persuasive. So although I present the argument following his presentation in Szabo [193], we should not assume that he doesn't agree with something like the line I take in response (see e.g. Szabo [190]: 31-2).

A theorist's stance towards assumption (I) and (II) is largely determined by her particular take on the semantics/pragmatics divide. Most theorists defend some version of either Contextualism or Minimalism. There are various ways of distinguishing these positions and their variants. For our purposes, there is a simple way of dividing the camps. We will classify as contextualist all accounts which allow that non-obvious indexical terms like *cut* and *red* can be context-sensitive, and we will classify as minimalists all accounts which allow context-sensitivity only to obvious indexicals such as *I*, *there*, etc.

Consider first Contextualism. Whether moderate or radical, one general characteristic of contextualist positions is that they hold that terms like *tall* and *red* are context-sensitive (see Heim [70]; Recanati [163]; Kennedy [88]; Segal and Rothschild [173]). Different contextualist positions sometimes propose different context-sensitive parameters for each type of expression. But independently of such particular proposals, a general consequence of contextualist positions is that they have to reject (I).

This is easy to see. To grasp the literal meaning of context-sensitive terms is to grasp something like their characters.[12] The full content of a complex expression e which has a context-sensitive constituent can only be derived after the relevant contextual information is supplied. But grasping this full content is not required to understand e's literal meaning. e's meaning can be computed

[12]Some contextualists argue that the context-sensitivity of expressions like *tall* and *red* is different, and more radical, than the context-sensitivity of ordinary indexicals like *I*. In the former cases, literal meaning is semantically more under-determined than in the case of ordinary indexicals. The idea is this. For an ordinary indexical, it is plausible (though open to debate in some cases) to suggest that its meaning is such that, given a context, it determines its content in a determinate way; so its literal meaning can be represented as a character in the Kaplan sense. In contrast, the literal meanings of other context-sensitive expressions have free-variables which have to be contextually assigned values, and those values are not completely determined by their literal meanings given a context. When a complex-expression has context-sensitive terms of this more radical variety, the semantics delivers, not really a character in the Kaplan sense, but something closer to a semantic schemata in Recanati's sense (see Recanati [161]: 56-7). It seems to me that this is an attractive contextualist position. Indeed, all the points I make above apply if we replace 'characters' with 'semantic schemata' (understood in Recanati's sense), so they should be interpreted as using the term 'character' in a very loose sense.

prior to the delivery of the relevant contextual information demanded by its context-sensitive constituents, in which case the complex expression e is also a character (of course, the contextual information can also be provided prior to the computation of e's meaning). Still, S_1 grasps the literal meaning of e if S_1 grasps e's character. For example, S_1 understands the literal meaning of *tall* x if, roughly, S_1 understands that x falls under it if its height is above some height h, where h is the average height of some (still to be specified) contextually provided set of x's. So, for the contextualist, the full-content account of literal understanding presented in (I) does not apply to sentences with context-sensitive constituents such as (3.1)-(3.4).

Relative to this account, expressions like (3.1)-(3.4) accord with S. S_1 understands the literal meaning of *tall building* if S_1 understands that something falls under it if it is a building and its height is above some height h_1, where h_1 is the average height of some contextually specified set of buildings. So assume that S_1 understands *tall building* and *car*, it seems to follow that S_1 has the linguistic competence to understand the literal meaning of *tall car*. For to understand the literal meaning of *tall car*, S_1 only needs to know that something falls under it only if it is a car and its height is above some height h_2, where h_2 is the average height of some contextually specified set of cars. Similar points apply to *red car* and *red building*, if *red* is treated as a context-sensitive term, except that the context-sensitive parameters are different.[13]

[13] As I made clear above, the point of my response to Szabo's challenge to S based on examples like (3.1)-(3.4) is not to defend a particular stance on the pragmatics/semantics divide, but only to show that there is no plausible stance under which we can jointly hold assumption (I) and (II), which are both required for the criticism to go through.

However, I should note that it seems to me that the contextualist account described here has some interesting intuitive support (though I do not mean to say that, by itself, this is a strong enough reason to accept it). For note that Szabo's suggestion that some competent speakers might understand (3.1)-(3.2) and not understand (3.3)-(3.4) seems too strong, too categorical a claim in a case where a notion of 'stages' or 'degrees' of understanding seems more suitable (I do not think Szabo would disagree with this). There is a clear rationale behind this intuition, one which relies on the notion of 'understanding' we employ when we are talking about expressions involving paradigmatic indexicals. So e.g. if S_1 understands *I like chocolate* when uttered by John, and *Mary likes candy* when uttered by John, we generally assume that

Now, once the contextual-values for *tall building*, *tall car*, *red car*, and *red building* are supplied (and this can be done either before or after the full sentential composition), you can get something like the specific contents for (3.1)-(3.4) suggested in (II). However, for the contextualist, grasping these contents or full truth-conditions is not required for literal understanding of (3.1)-(3.4). In this sense, contextualists can accept something like (II), but this is no threat to S, since they reject (I). Of course, there are different contextualist proposals, but the reason why they relieve the tension between examples like (3.1)-(3.4) and S is insensitive to the particular proposals.

Consider next minimalist positions. Minimalists, whether radical or moderate, reject context-sensitive accounts of non-obvious indexical expressions (see Borg [14], [16]; Cappelen and Lepore [20]). In addition, and unlike contextualists, most minimalists can consistently accept (I). In fact, many theorists think that one of the defining features of minimalism is the assumption that indexical-free declarative sentences express complete propositions or truth-conditions (Borg [16]). In any case, the important thing to note here is that all plausible minimalist positions have to (and in fact do) reject (II).

According to minimalism, we should not treat expressions like *red* or *tall* as context-sensitive. We should assume instead that sentential expressions which contain them—expressions like (1)-(4)—have 'minimal' truth-conditions. How to determine the 'minimal' truth-conditions of sentences is a matter of ongoing debate (see e.g., Borg [14]: 209-258; Stanley [185]; Recanati [161]: 51-67). To get a sense of the current suggestions, consider two popular proposals:

S_1 can understand *I like candy*, even when S doesn't know who uttered that expression. This is because we take it that S_1 tacitly understands that the (context-insensitive) character of *I* is a function from a token of *I* to the utterer of that token. In other words, we take it that to understand *I* is, roughly, to understand that, in general, tokens of *I* refer to whoever uttered them. Something like this framework is what we intuitively apply in the case of (3.1)-(3.4). This is why it is natural to hold that if S_1 understands (3.1)-(3.2), then S_1 understands (3.3)-(3.4), in the sense that S_1 can understand, at least, their characters. This is why it seems strange to suggest that we could find a competent speaker who, despite understanding (3.1)-(3.2), doesnt understand the character of (3.3)-(3.4).

(T_1) The minimal proposition of sentence p is the proposition that all utterances of p share. It is the proposition that all utterances of p express no matter how different their contexts of utterance (Cappelen and Lepore [20])

(T_2) The minimal proposition of sentence p in context C is what would be asserted and conveyed in every normal context $C*$ in which the reference of all indexicals in p is the same as their reference in C. (Soames 2002)

The idea that rough tests like T_1 or T_2 generally succeed in isolating minimal propositions or contents for sentences has been extensively (and justly) criticized (see Recanati [161]: 58-61; Stanley [185]). But in this discussion we can assume that tests like T_1 or T_2 serve to isolate minimal propositions or truth-conditions for sentences. For even if the details are incorrect, at least the general idea behind these tests is clear: the minimal content expressed by a sentence p is something like a common content denominator shared by different assertions of p, so that p's minimal content is generally logically weaker than each full content expressed by a particular assertion of p.

The question here is whether, following T_1 or T_2, we would assign (something like) the minimal truth-conditions to sentences (3.1)-(3.4) suggested in (II). And the answer is clearly 'no'. Based on tests like T_1 or T_2, the minimal contents for (3.1)-(3.4) are something closer to (3.1*)-(3.4*):

(3.1*) If p is an utterance of *John has a red car* in a context C then p is true iff, in C, John has a car and the car is red in some way.

(3.2*) If p is an utterance of *John has a tall building* in a context C then p is true iff, in C, John owns a building and the building has a greater height than the average height of some set of buildings.

(3.3*) If p is an utterance of *John has a tall car* in a context C then p is true iff, in C, John has a car and the car has a greater height than the average height of some set of cars.

(3.4*) If p is an utterance of *John has a red building* in a context C then p is true iff, in C, John has a building and the building is red in some way.

These minimalist truth-conditions for (3.1)-(3.4) are radically different—since they are logically much weaker—than the literal truth-conditions suggested in (II).[14] Take (3.1) as an example. (II) says that its literal truth-conditions could be something like this:

(3.1**) If p is an utterance of *John has a red car* in context C then p is true iff, in C, John has a car and the car is painted red on the outside.

Clearly (3.1**) does not specify minimal truth-conditions, at least according to conditions like T_1 or T_2. The reason is simple. We can easily imagine a context C_1 in which S asserts *John has a red car*, uses that expression literally, and means that John has a car that is completely red on the inside. So the content specified in (3.1**) is not asserted in context C_1. There are of course other contexts where something like the content specified in (3.1**) is asserted. But if some content is shared between the assertion made in these other contexts and the assertion made in C_1, it is clearly not the content specified in (3.1**), but something much weaker, something closer to the content specified in (3.1*).

[14] Some minimalists might suggest that the minimal truth-conditions of sentences (3.2)-(3.3) are not something like (3.2*)-(3.3*) but something like (3.2**)-(3.3**):

(3.2**) If p is an utterance of *John has a tall building* in a context C then p is true iff, in C, John owns a building and the building has a greater height than the average height of the set of buildings.

(3.3**) If p is an utterance of *John has a tall car* in a context C then p is true iff, in C, John has a car and the car has a greater height than the average height of the set of cars.

The difference between truth-conditions (3.2**)-(3.3**) and (3.2*)-(3.3*) is that the comparison class set for determining tallness in (3.2**)-(3.3**) is much wider—the set of buildings (in (3.2**)) and the set of cars (in (3.3**)). In other words, (3.2**) entails (3.2*) and (3.3**) entails (3.3*), but not vice-versa. The problem with truth-conditions (3.2**)-(3.3**) is that there are many normal assertoric contexts where (3.2)-(3.3) are used without asserting the strong content specified in (3.2**)-(3.3**). For example, any assertion of (3.2) where the intended comparison class is some very restricted set of buildings does not include as part of its assertoric content the content specified in (3.2**), but it does include the content specified in (3.2*), and this is generally the case. So I think it is clear that, if (3.2)-(3.3) express minimal contents, then those contents are more like (3.2*)-(3.3*) than like (3.2**)-(3.3**).

As formulated, (II) is consistent with other content assignments for (3.1), not just (3.1**). The same is true for the other examples. However, (II) does place an important restriction on the range of possible content assignments, namely, that the contents for (3.1)-(3.4) be specific enough such that the following holds: a car's way of being red is not the same as a building's way of being red, and a car's way of being tall is not the same as a building's way of being tall. However, reasoning analogous to the above shows that any such specific content will not specify a minimal content. For example, suppose we change (3.1**) so that the condition is that the car be painted red on the inside, or in any other specific way. We can easily imagine a context C_2 in which S asserts *John has a red car*, uses that expression literally, and means that John has a car that is completely red on the outside. So the content specified in revised (3.1**) is not asserted at all in context C_2, etc. Clearly, only something as weak as (3.1*) will work as a specification of the minimal content of (3.1). Analogous points hold for (3.2)-(3.4). So if we assume that the literal meanings of (3.1)-(3.4) are the minimal contents specified in (3.1*)-(3.4*), or something close to that, then there seems to be no doubt that competent speakers who know the meanings of (3.1)-(3.2) have the linguistic competence to generate the meanings of (3.3)-(3.4).

The 'common content denominator' idea behind the view that there are minimal contents/propositions has, as I said, been extensively criticized in the literature. However, to determine the plausibility of S we do not have to side with Minimalists or Contextualists. The only point we need and have defended here is that there seems to be no principled position on the semantics/pragmatics divide from which we can, consistently, accept (I) and (II). Since this is required to turn Szabo's argument into an argument against S, examples such as (3.1)-(3.4) not only give us no reason to doubt S, but when consistently interpreted, actually support it. Again, this holds whether one prefers a Minimalist or a

Contextualist position on the semantics/pragmatics divide.

Johnson [81] presents an objection against S which is superficially similar to the one just addressed. Johnson is even more convinced than Szabo [193] that if we move beyond some often cited simple examples, we begin to notice the pervasive non-systematicity of our linguistic competence.

Johnson's paradigmatic type of example of non-systematicity consists of certain cases of VP recombination. For example, assume S_1 can understand both:

(3.5) John put his gear down.

(3.6) John stowed his gear.

It doesn't seem to follow that S_1 can understand (through pure linguistic means[15]) either of:

(3.7) *John put his gear.

(3.8) *John stowed his gear down

Likewise, assume S_1 can understand both:

(3.9) John drenched the flowers with water.

(3.10) John threw water unto the flowers.

It doesn't seem to follow that S_1 can understand (through pure linguistic means) either of:

(3.11) *John drenched water unto the flowers.

(3.12) *John threw the flowers with water.

[15] We qualify this notion of 'understanding'—i.e., to 'understanding via an I-language'—because these kinds of ungrammatical expressions can often be assign interpretations (and sometimes seemingly effortlessly) via other non-linguistic means—e.g., by using pragmatic/general reasoning.

So understanding (3.5)-(3.6) and (3.9)-(3.10), doesn't guarantee that S_1 can understand (3.7)-(3.8) or (3.11)-(3.12) because these latter expressions are not even grammatical. Still, these ungrammatical expressions are all obtained by replacing, in complex expressions S_1 understands, one verb S_1 understands, with another verb that S_1 also understands. So it *seems* that S falsely predicts that if S_1 understands (3.5)-(3.6) then S_1 should understand (3.7)-(3.8), and if S understands (3.9)-(3.10) then S should understand (3.11)-(3.12).

However, it is easy to see that S doesn't make those false predictions; it only seems to do so if you consider an informal formulation of systematicity, one that is formulated at the level of surface expressions, instead of LFs.

A crucial feature of Johnson's examples is that no competent speaker of English can understand (by means of his/her FL) expressions (3.7)-(3.8) and (3.11)-(3.12). In other words, expressions (3.7)-(3.8) and (3.11)-(3.12) are not expressions that belong to the set of English expressions. So what do these examples suggest? Prima facie, all that these examples seem to show is that, in describing the systematicity data, we need to appropriately restrict the systematic variant sets of sets of expressions. A plausible and principled way of doing this is suggested by Johnson's own examples: formulate S so that it excludes, from the systematic variant sets, expressions that are not expressions of English.

Indeed, it is easy to see that this restriction is implicit in S. Still, given Johnson's objection, perhaps it would be useful to explicitly state this restriction. Consider *rephrasing* S as S*:

(S*) The generative capacity of FL_S is structured in the following way: if it can generate *correct* interpretations (relative to English) for complex LF expressions of English $e_1, \ldots, e_n$, it can generate correct interpretations (relative to English) for all other complex LF expressions of English constructed from: (i) constituents of $e_1, \ldots, e_n$ and (ii) syntactic structures employed in any of the LF expressions $e_1, \ldots, e_n$.

S* does not predict that if S_1 can understand (3.5)-(3.6) and (3.9)-(3.10) then S_1 should also be able to understand both (3.7)-(3.8) and (3.11)-(3.12), for the latter expressions are not expressions of English (note the italicized condition on S*). Of course, this formulation of S doesn't amount to much unless we can determine whether an expression is an expression of English independently of the fact that the FL of some competent speaker of English (in this case FL_S) doesnt generate an interpretation for it.

There is a straightforward way of doing this. To do so we have to invoke the somewhat controversial notion of a 'community' of competent speakers of English. How do we employ this notion in solving our problem? Here is an initial, rough default rule: a complex expression e is not an expression of English if the FL of competent speakers of English don't generate an interpretation for e, even when they have access to the meanings of e's constituents.[16] Put informally: a complex expression e which superficially seems like an expression of English—and which some arbitrary competent speaker of English doesn't understand—is not an expression of English if other competent speakers of English find it ungrammatical or somehow blatantly deviant, when they understand the meanings of its constituent words. Of course, this is only a default rule. In special circumstances we might have theoretical reasons to count an expression as an expression of English despite the fact that (perhaps all) speakers of English find it deviant or incomprehensible. Some examples include expressions that have three or more center embeddings and expressions that are too long. In these special cases, we can hold that these expressions are expression of English partly

[16] As it stands, this default rule is incorrect: it overshoots in some cases and undershoots in others. It seems to exclude: *colorless green ideas sleep furiously.* This sentence arguably doesn't have a semantic interpretation (but see Camp [19]; Higginbotham [72]), but it surely is a well-formed expression of English, and most English speakers would recognize it as such. It also seems to exclude expressions which are ungrammatical but which apparently have semantic interpretations (see Higginbotham [72]). Do we want to exclude these expressions from systematic variant sets?

because we have a clear non-linguistic explanation for why speakers find them deviant or incomprehensible. But expressions like (3.5)-(3.6) and (3.9)-(3.10) don't seem to be special in any of these ways, so the fact that all speakers of English find them deviant or incomprehensible suggests that they are not expressions of English.[17]

We have, then, an independent way of reliably determining whether a complex expression e, which the FL of some speaker doesn't generate or interpret, is an expression of English. When this te st is applied to examples (3.7)-(3.8) and (3.11)-(3.12), the result is that these are not expressions of English. So S does not predict that if S_1 understands (3.5)-(3.6) and (3.9)-(3.10), then S_1 has to understand (3.7)-(3.8) and (3.11)-(3.12). In short, Johnson's examples do not undermine S in any way.

Let us briefly recap our defense of S. We considered two arguments which try to undermine the prima facie plausibility of S. The first argument, due to

[17]Here is further consideration which reinforces the claim that (3.7)-(3.8) and (3.11)-(3.12) are not expressions of English. Note that we have a plausible explanation (compatible with compositional theories) for why expressions like (3.7)-(3.8) and (3.11)-(3.12) are not expressions of English. For brevity, consider only (3.7)-(3.8), which do not differ in relevant respects from (3.11)-(3.12). One hypothesis is that *put*, unlike *stowed*, is a transitive verb one of whose arguments is an adjective or adverb; semantically we can capture this by saying that it is a function of type $<< e, t >, <<< e, t >, < e, t >>, < e, t >>>$, whereas *stowed* is a function of type $<< e, t >, < e, t >>$. So in (3.5) *put* lacks an argument, and it therefore comes out as deviant.

The case of (3.8) is more complicated. Initially, we might think that it consists of a type mismatch; but this is not easy to show, at least if we use the semantic types and type assignments commonly used in formal semantics. Here's why. Consider its combinatorial structure [[John] [[stowed [his gear]] [down]]], and let us determine whether there is a type mismatch at any stage of the com-position process. We might think that it occurs when combining [stowed his gear] with [down], because the former expression, of type $< e, t >$, requires an argument of type e, which is a type that we cannot plausibly assign to *down*. However, *down* is plausibly an expression of type $<< e, t >, < e, t >>$, so it can take $stowedhisgear$ as an argument—hence there is no type mismatch. Now, this doesn't completely rule out a type-mismatch explanation for (3.8). For some theorists think that we should adopt a finer type-system than the one commonly used in Montague semantics (see e.g. Pustejovsky [148]; Pinker [140]). If we do this, we can then place more specific restrictions on the type of predicate that *down* can take as an argument so that the type of *stowed his gear* is not in its domain. A finer type system also allows us to explain other examples of this kind: e.g., why *John likes ice-cream* is a sentence of E-English, but *ice-cream likes John* arguably is not. The type of explanation is just as before: the type of function assigned to *likes* is such that its second argument has to be an animate being, a condition violated in *ice-cream likes John* but respected in *John likes ice-cream.*

Szabo, focuses on NPs, and the second, due to Johnson, focuses on VPs. Although different in important ways, both objections had the following general form. Each considered a particular set of expressions A and one of its systematic variant sets B, and argued that, contrary to what S seems to predict, assuming FL_S can generate interpretations for the expressions in A, it does not seem to follow that this is sufficient for FL_S to generate interpretations for the expressions in B.

In response, we pointed out two types of mistakes that theorists often make when presenting these sorts of objections against S.

The first type of mistake involves assigning literal meanings to the systematically related expressions in A and B in a way that betrays an inconsistent position on the semantics/pragmatics divide. To show this, we did not take a particular position on the semantics/pragmatics divide. So we did not make the illegitimate move of assuming an ad hoc position merely because it insulates S from counter-examples. Instead, we saw that as long as we consistently adopt any version of one of the two most general positions—Minimalism or Contextualism—S holds for the supposedly problematic examples.

The second type of mistake involves claiming that a set B, which is a systematic variant of a set A, contains certain expressions which S, when properly interpreted, does not predict should belong to B. To show this, we did not adopt an ad hoc reformulation of S merely to exclude these purported counter-examples. Instead, we saw that, properly understood, S does not in fact predict that the problematic expressions belong to the systematic variant set B.

To conclude the defense of S, I want to consider a more general objection which claims that S assumes an unjustified idealization of the data. According to this objection, the only thing we can infer from the competence patterns exhibited by S_1 is that FL_S can determine the meaning of a novel expression *e* from the meaning of its constituents, its structure, and possibly the meaning of other expressions. Szabo [193] makes essentially this point in the following passage:

> There is also the problem that the argument shows less than it claims to. If we run the argument for the pair of sentences *apples are red* and *bananas are yellow* we can conclude that the meanings of *apples*, *bananas*, *are red*, and *are yellow* plus predication determine the meaning of *bananas are red.* It does not follow that the meanings of *bananas* and *are red* plus predication do that (Szabo [193]: 18).

S is somewhat different from the premises of the traditional systematicity argument which Szabo is directly attacking in this passage. Still, Szabo's observation, if correct, strongly suggests that S is an unjustified simplification of the data. Szabo's objection has been approvingly repeated by other theorists (see e.g. Baggio, Van Lambalgen and Hagoort [2]: 657)

However, Szabo's objection does not work as an objection to S, although it might work as an objection to other formulations of systematicity found in the literature. Consider Szabo's example. S entails that if S_1 understands *apples are red* and *bananas are yellow*, S_1 has the linguistic competence to understand *bananas are red.* But S also entails, for another competent speaker S_2, that if S_2 understands *trees are red* and *bananas are small*, S_2 has the linguistic competence to understand *bananas are red.* This suggests that *apples* and *are yellow* do not have an essential role in determining the meaning of *bananas are red.*

One might object that this is compatible with the possibility that the meanings of *apples*, *bananas*, *are red*, *are yellow*, *trees*, and *are small*, plus predication, determine the meanings of *bananas are red.* However, other facts predicted by S block this conclusion. For example, consider another competent speaker S_3, and assume that S_3 does not understand *trees are small* or *bananas are yellow*, but understands *bananas are sweet* and *cherries are red*. S entails that S_3 has the linguistic competence to understand *bananas are red*. So S entails that the meaning of *bananas are red* does not essentially depend on the meanings of expressions other than *bananas* and *are red*.[18]

Note that of we accept CO or any of the competing non-compositional MDC we predict those systematic data patterns: if what is essential to understand the meaning of *bananas are red* is just to understand *bananas*, *are red*, plus their mode of combination (which might be non-compositional), then we allow the possibility for speakers who understand *bananas are red* and do not understand *are yellow*, *apples*, *trees*, *are small*, etc. (e.g., this possibility might obtain for some speakers at certain early stages in the acquisition of English; we could easily replicate an analogous condition in an experimental learning setting).

3.4 Conclusion

To conclude this defense of the premises of the argument for CO, I want to briefly address some objections which have been raised against both A1 and A2. I don't think these are serious objections, but addressing them does allow us to further clarify this argument. Consider first A2:

(A2) S_1 could not have acquired the information required to generate each of the novel $<$ LF, meaning$>$ pairs by memorizing the meaning of each novel LF indepedently—i.e., in the list-like manner in which we memorize a finite lexicon.

Some theorists justify A2 by appeal to the memory limitations of human cognition, a fortiori to limitations of the knowledge or information that FL_S can draw

[18]This reading of S does not rule out the possibility of local (or non-local) holisms.

on to generate interpretations. This justification is rejected by some theorists on the grounds that (i) we do not know enough about the limitations of long term memory to impose on it any particular limit, and (ii) maybe the fact that there is such a limit actually explains the performance limits of our productive capacities.

However, the correct justification of A2 doesn't depend on any assumptions about our memory limitations. The crucial fact which underlies A2 is this: S_1 has the capacity to understand an open-ended set of *novel* expressions; hence, S_1 has not encountered these expressions in a learning context; a fortiori, S_1 has not learned their meanings by memorizing a list of these $< LF$, meaning$>$ pairs. So even if we assume that human memory has no limitations, A2 is still a correct premise.

Of course, we can imagine other ways of rejecting A2; but, as far as I can see, they all depend on adopting very implausible assumptions. For example, we could undercut this argument for A2 if we hold that S_1 has innate knowledge of the novel $< LF$, meaning $>$ pairs. I doubt that any theorist would pay this price to reject A2; such a theorist would have to hold that S_1 has innate knowledge of the meaning of sentences like *all purple cars are hidden behind chairs.*

Consider next A1:

(A1) There must be some cognitive computational state and capacity of FL_S on the basis of which it can generate such (indefinitely many) correct (relative to English) novel $< LF,$ meaning $>$ pairs.

It is hard to directly justify A1 because it is both challenged and defended by appeal to fundamentally different theoretical approaches to the study of mind/language. For most theorists, their degree of sympathy towards A1 seems to depend on their general views about the metaphysics of mental processes and unconscious mental states. On one side of the spectrum you find theorists who accept computational explanations of certain complex cognitive capacities. These theorists tend to accept A1 (e.g., Chomsky, Egan, Pietroski). A subset of these theorists—those who also accept a Representational Theory of Mind—tend to be very sympathetic to A1 (e.g., Rey, Fodor, Pylyshyn). On the other side of the spectrum, you find theorists who reject computational or intentional-level explanations of cognitive capacities (e.g., behaviorists), or who accept them only in an instrumentalist sense, with varying degrees of approval (e.g., Dennett and the Churchlands). These theorists tend to reject A1. Specifically, what this group of theorists rejects is roughly this: that what determines the understanding of novel complex expressions is a computational/representational cognitive state or system, one that is tacitly known or has been internalized (e.g., a recursive generative grammar with a compositional semantics).

Now, there is a standard justification for this skepticism, aimed specifically at something close to A1, which goes roughly as follows (Fodor 1998; Schiffer 1991). If a finite creature has an infinite or unbounded representational capacity—e.g., knowledge of English—then the capacity must be finitely determined. But it doesn't follow that to have finite determination the creature has to represent the compositional structure (or any structure along the lines of one of the competing MDCs) of its unbounded representational repertoire, or even that the unbounded representational repertoire has a compositional structure. For example, it may be that sufficient conditions for having an infinite or unbounded representational capacity can be finitely specified in and only in the language of neurology.

This objection—which I don't think is very compelling—forces us to be very clear about something: we are not presenting a demonstrative argument. There is no reason why we cannot accept that the claim made in A1—that what determines the unbounded computational/representational capacity is some finite computational/representational state—might turn out to be empirically false. At the same time, we should not forget that, currently, there is no non-computational/non-representational theory that can compete with current computational/representational theories in their power to explain how an unbounded computational/representational capacity can be determined from a finite state.

So although (it seems like) A1 should be rejected if presented as part of a demonstrative argument from $P\&S$ to CO, it is a very plausible assumption when presented as part of a non-demonstrative inference to the best explanation from $P\&S$ to CO. In this context, the argument presented above against A1 is no better than writing a blank check on neurological explanations of unbounded computational/representational capacities. It is possible that such explanations of $P\&S$ might one day be available, but until they are there is no reason why admitting their possibility should substantially reduce our confidence in our computational/representational level explanations.

Part II

On the Semantic Representations of *FL*

Chapter 4

Dual Content Semantics, Privatives, and Dynamic Compositionality

This chapter defends the view that common nouns have a dual semantic structure that includes extension and non-extension determining components. I argue that the non-extension determining components are part of linguistic meaning because they play a key compositional role in certain constructions, esp., in privative noun phrases such as *fake gun* and *counterfeit document.* Furthermore, I show that if we modify the compositional interpretation rules in certain simple ways, this dual content account of noun phrase modification can be implemented in a type-driven formal semantic framework. In addition, I also argue against traditional accounts of privative noun phrases which can be paired with the assumption that nouns do not have a dual semantic structure. At the most general level, this paper presents a proposal for how we can begin to integrate a psychologically realistic account of lexical semantics with a linguistically plausible compositional semantic framework.

4.1 Introduction

Are lexical items the atoms of meaning or do they have internal semantic structure? If the latter, what kind of structure—definitions, prototypes, some sort of holistic structure? These questions have received a lot of attention in philosophy, linguistics and psychology, in part because of their intrinsic interest, but mostly because they are intimately connected to questions about nativism, the structure of lexical concepts, the nature of language acquisition, and the relation between language and thought.

We will explore questions about lexical decomposition by focusing on noun phrases (NPs).[1] We will try to uncover the semantic structure of nouns (Ns) by examining their compositional behavior in 'privative NPs' such as *fake gun* and *counterfeit document*. This is a class of systematic and productive NPs which have privative adjectives (Adjs) as main modifiers and artifact terms as head Ns. Privatives have puzzled many theorists because of their unique compositional behavior. For example, privatives behave differently from the intersectives and the subsectives. (4.1) displays the characteristic inference pattern of intersectives:

$$(4.1)\quad \text{a.}\ \frac{x \text{ is a red gun}}{x \text{ is a gun}} \qquad \text{b.}\ \frac{x \text{ is a red gun}}{x \text{ is red}}$$

(4.2) displays the characteristic inference pattern of subsectives, where '#' marks incorrect inferences:

$$(4.2)\quad \text{a.}\ \frac{x \text{ is a good gun}}{x \text{ is a gun}} \qquad \text{b.}\ \frac{x \text{ is a good gun}}{\#\ x \text{ is good}}$$

[1] Recent discussions of lexical decomposition tend to focus on verbs. Although some theorists defend atomism (Fodor [42]; Fodor and Lepore [49]), the emerging consensus is that most verbs have internal semantic structure, with much disagreement about the specifics of this structure (see Wunderlich [211]; Harley [69]; Hinzen [73]; Pustejovsky [150]).

Privatives have inference patterns like (4.1b) but not like (4.1a) or (4.2a):

(4.3) a. $\dfrac{x \text{ is a fake gun}}{\#\ x \text{ is a gun}}$ b. $\dfrac{x \text{ is a fake gun}}{x \text{ is a fake}}$

Privatives also differ from two important classes of non-subsective Adjs, the plain non-subsectives (mainly constituted by certain epistemic modals) and the temporal and alethic modals. (4.4) displays the characteristic inference pattern of plain non-subsectives:

(4.4) a. $\dfrac{x \text{ is an alleged criminal}}{\#\ x \text{ is not criminal}}$ b. $\dfrac{x \text{ is an alleged criminal}}{\#\ x \text{ is a criminal}}$

(4.5) displays the characteristic inference pattern of temporal modals:

(4.5) a. $\dfrac{x \text{ is a former/future criminal}}{\#\ x \text{ is not a criminal}}$ b. $\dfrac{x \text{ is a former/future criminal}}{x \text{ was/will be a criminal}}$

Unlike plain non-subsectives and temporal modals, privatives allow inferences like (4.4a)-(4.5a).[2] In addition, they do not allow inferences like (4.5b):

(4.6) a. $\dfrac{x \text{ is a fake gun}}{x \text{ is not a gun}}$ b. $\dfrac{x \text{ is a fake gun}}{\#\ x \text{ was/will be a gun}}$

Note, finally, that from *x is a former/future criminal* you cannot infer anything about what x does or seems like; but from *x is a fake gun* you can infer something about what x seems like or about the intention for which x was created: a fake gun either seems like a gun or was created to seem like a gun.[3]

The inferential behavior of privative NPs presents substantial challenges for compositional semantic theories (see e.g., Franks [53]; Partee [126], [128]; Recanati [163]: ch. 2); some theorists even argue that they force us to abandon that framework (Lakoff [98]; Coulson and Fauconnier [32]). In contrast, I will

[2]One might question (4.5a): it might seem that from *x is a former/future criminal* you can infer that, at the moment the assertion is made, *x is not a criminal*. However, this is not the case. It is easy to think of examples in which we assert that x is a former/future criminal and know that x is also currently a criminal.

[3]There have been various attempts to reduce the number classes of Adjs, and especially to try to lump as many Adjs as possible into the intersectives (Heim and Kratzer [70]; Partee [130], [126], [128]; Recanati [163]). Here we are mainly concerned with privatives, so we will discuss attempts to reduce these classes only when they bear on the privatives, esp., when they attempt to reduce them to the subsectives or intersectives.

argue that privative NPs are not a threat to compositional formal semantics. What they do suggest, however, is that we abandon the assumption that most common Ns are semantically atomic.

In what follows, I defend a compositional account of privative NPs which can be implemented in a type-driven formal semantic framework, once this framework is modified in ways that I will present in detail in §4.3. This account is based on two ideas about the semantic structure of Ns. Assume 'e' stands for an arbitrary common N (e.g., an artifact or natural kind term).

(I) e has a complex semantic structure. Part of this structure determines e's extension (call this its 'E-structure'), and part of it does not determine e's extension (call this its 'C-structure').

(II) e's C-structure includes, at least, representations of perceptual features, functional features and genealogical features related to e, and it determines e's compositional contribution to certain complex constructions.

According to this 'DUAL CONTENT' account, to fully grasp the literal meaning of artifact and other types of lexical terms such as natural kinds, speakers need to grasp both their E-structure and their C-structure. The E-structure plays a role in, e.g., predicative uses of Ns. The role of C-structure is less straightforward, but no less important. I will argue that to explain why speakers have a productive and systematic competence with privative NPs we need to assume that Adjs such as $fake$ operate on the C-structure of the Ns they modify. I will also point out other types of complex expressions in which C-structure seems to play a crucial compositional role.

DUAL CONTENT is based on a non-atomistic theory of lexical meaning, but it should be distinguished from the two most famous non-atomic theories—prototype and definitional theories. Unlike pure prototype theories, according to DUAL CONTENT the meaning of lexical terms includes E-structure, i.e., an extension determining atomic component. In addition, the C-structure of terms is, as we will see, richer than the structures assumed by most prototype theories. Unlike definitional theories, according to DUAL CONTENT C-structure often determines neither the E-structure nor the extension of terms. C-structure is part of the meaning of terms because it is a necessary part of what language-users need to grasp to have full (literal) linguistic competence.

It might seem that I'm using the notion of 'linguistic meaning' in an idiosyncratic way; but my usage is appropriate in the context of debates about lexical decomposition. As I use the term, the 'meaning' of a lexical item e includes that which e contributes to the determination of the meaning of complex expressions of which e is a constituent. More generally, meaning is that which or part of what speakers grasp when they learn how to 'use' e, which includes the systematic and productive use of e in complex expressions such as privative NPs. Some prominent theorists also assume that the meaning of e completely coincides with the extension-determining semantic features of e. I take this claim as a hypothesis about linguistic meaning and not as part of the meaning of 'linguistic meaning', and part of the task of this paper is to undermine it.

4.2 DUAL CONTENT and privative NPs: Philosophical motivations

According to DUAL CONTENT, we can account for the inference pattern of privative NPs if we assume that (i) the head Ns have a complex semantic structure which includes E-structure and non-extension determining C-structure, and (ii) that privative Adjs operate in certain ways on the C-structure of the modified Ns. Before developing this account in §4.3, I want to discuss its two basic ideas, in part because some of the details of the implementation can be somewhat intricate and only if we are clear about the sort of general account we are trying to develop will some of the technical decisions make sense.

The idea that some types of Ns have a complex semantic structure which includes E-structure (which determines their range of correct application) and C-structure (which includes associated beliefs), and that the reason why we posit C-structure as part of the meaning of Ns is to explain certain aspects of our linguistic competence, was originally defended by Putnam [152].

When Putnam first defended this idea, versions of the definitional theory of lexical terms and concepts were quite popular. According to definitional theories, it usually holds that if e is a N the meaning of e can be given by specifying a conjunction of properties $P_1 \ldots P_n$ that (mature speakers know) define or determine e's extension. For each P_i, the statement *e has the property* P_i is an analytic truth. Putnam argued that these theories are clearly incorrect for a rather simple reason: for any property that supposedly 'defines' an artifact or natural kind for a mature speaker, one can (as a representative mature speaker) imagine a counterfactual situation in which something does not have that property and yet still correctly falls under the kind. At around the same time, Kripke was developing a similar objection to definitional theories of natural kinds [95].

Putnam's criticism of definitional theories is widely accepted by philosophers and linguists, but it is sometimes ignored or too lightly discussed by psychologists. Among theorists who take Putnam's view seriously, most have abandoned the idea that the meaning of terms can be represented as non-atomic structures with semantically significant parts. One famous example is Fodor [43], who argues that the failure of definitional theories and their descendants—essentially for the reasons Putnam originally presented—forces us to accept atomistic theories of lexical terms and concepts. Theorists like Fodor, who on the strength of Putnam's criticisms of definitional theories see atomism as the only route, remain attached to the idea that there can be nothing more to the meaning of an expression than its E-structure.

However, there is another way to proceed from Putnam's criticisms of definitional theories, somewhat lost in subsequent discussions, but actually suggested by Putnam himself: hold that most Ns have a complex semantic structure, but deny that the role of that structure is only to determine their extension. Putnam suggests that the meaning of an arbitrary natural-kind term e is a complex involving two representational dimensions: a representation that e stands for a natural-kind, and a representation of the 'stereotype' associated with e, i.e., a structured set of properties which speakers associate with normal members of the kind e. According to Putnam, e's stereotype does not determine e's extension, but it is part of its core meaning and not just a bit of associated encyclopedic knowledge because it is part of what speakers need to know in order to have full linguistic competence with e.

Putnam thinks that the view that e's stereotype is part of its meaning becomes clear when we consider questions about language acquisition—about what we learn when we learn the meaning of e such that knowing that enables us to use e. According to Putnam, most language theorists fail to be impressed by the fact that, in many cases, the use of Ns can be taught by presenting speakers with the word's stereotype. More precisely:

> The hypothesis is that there are, in connection with almost any word (not just 'natural kind' words), certain core facts such that (1) one cannot convey the normal use of the word (to the satisfaction of native speakers) without conveying those core facts, and (2) in the case of many words and many speakers, conveying those core facts is sufficient to convey at least an approximation to the normal use. In the case of a natural kind word, the core facts are that a normal member of the kind has certain characteristics, or that this idea is at least the stereotype associated with

the word [152].

Putnam's basic idea is that when you show a speaker the stereotype—the set of core facts—associated with a term you enable the speaker to competently use the term, and you cannot achieve this in any other way. In other words, Putnam's claim is that part of what we learn when we learn the meaning of a kind term is the associated stereotype, and this knowledge is what enables us to competently use the term. Still, as revealed by their response to various counterfactual scenarios, speakers know that the associated stereotypes and core facts do not, in general, determine the range of correct application of kind terms. This is why we must also assume that lexical terms include an extension determining E-structure, which C-structure in some ways traces, but does not fully determine.[4]

Putnam's idea that the semantic structure of lexical terms can be divided into E-structure and C-structure, and that C-structure is part of the literal meaning of terms because it is an essential part of our linguistic competence, is a move in the right direction, fully embraced by DUAL CONTENT. However, Putnam's account is incomplete in two important respects, which DUAL CONTENT tries to rectify. First, we need a detailed example of a *type* of linguistic competence that essentially involves C-structure. Otherwise one can agree with Putnam that we learn the meaning of terms *via* their associated stereotype (Putnam's C-structure), but deny that these stereotypes are part of their meaning.[5] Second, we need a more detailed account of the C-structure of terms, since it is unclear what information Putnam would include under his 'core facts'.

[4]For a detailed discussion of E-structure, C-structure, and their relation, see §4.7 below. For now, the basic idea is just that E-structure is a semantically atomic component of the meaning of lexical terms which, via some primitive relation, determines their extension or range of correct application.

[5]For example, Margolis [107] and Fodor [46] acknowledge that stereotypes might be 'sustaining mechanisms' involved in the acquisition and even application of concepts, but still deny that stereotypes are constituents of concepts. This same view was also defended in a very influential paper by Rey [165].

This is where privative NPs come into the picture. They provide us with a concrete example of a kind of (literal, non-pragmatic) general 'linguistic competence'—the capacity to systematically and productively use private NPs—that seems to essentially involve the C-structure of terms. In addition, privative NPs yield important insights about the details of C-structure, including that it seems to include more information than just a representation of the associated stereotype.

The idea that we can uncover the semantic structure of common Ns by investigating compositions involving privative Adjs is not new [33, 54]. One of the first and most insightful presentations of this idea is found in Layoff and Johnson [99]. They agree that the compositional behavior of privative NPs such as *fake gun* is adequately captured by the inference patterns presented in §1. So any adequate treatment of privative Adjs such as *fake* has to explain why from knowing that, e.g., x is a fake gun we can infer (i) that x is not a gun and (ii) that x was created with the intention that it only seem like a gun, but not that it can be used to shoot. Lakoff and Johnson argue that to explain (i) and (ii)—generalized to include other artifact kind terms—we need a non-atomistic account of the lexical semantics of the modified Ns.

Lakoff and Johnson's account of privative NPs is based on the idea that privative Adjs operate over the complex semantic structure of artifact terms. In the case of *fake gun*, they propose that *gun* has a complex semantic structure which includes at least three dimensions: perceptual, functional, and geneological representations. They then take the semantic potential of *fake* to be that of an operation over this complex semantic structure: the operation negates both the function and the geneological representations of *gun* (since fake guns are not supposed and were not made to function like real guns, i.e., to shoot), keeps the perceptual representation unmodified (since fake guns are supposed to look like guns), and takes this perceptual representation as the value of the genealogical dimension (since fake guns are made to look like guns). This last point is important, since something made to be a gun which doesn't function like a gun (because it is old or doesn't work) is a malfunctioning gun but not a fake gun. Fake guns are artifacts made with the intention that they be fake guns.

There is an important difference between DUAL CONTENT and the account of lexical decomposition defended by Lakoff and Johnson [99]. Both accounts agree that, in general, lexical terms do *not* decompose into sets of necessary and sufficient conditions. For example, we may categorize something as a gun even if it doesn't function like a gun or doesn't look like a typical gun. In other words, we accept that there can be guns that don't shoot, or guns that don't look like typical guns. Indeed, we can imagine cases in which something that was not made to be a gun nevertheless is used and functions so well for shooting, that we would categorize it as a gun. So the perceptual, functional, and geneological features associated with *gun* are independent of its E-structure.[6] At the same time—and as we will argue in detail below—there are constructions in which functional or genealogical information is introduced into the E-structure for complex expressions. Although Lakoff and Johnson do not explicitly argue against this possibility, it cannot be captured in a framework, such as theirs, which does not preserve the distinction between E-structure and C-structure.

DUAL CONTENT, which we will now present in detail, takes Putnam's insights about the basic division of internal lexical semantic structure, and Lakoff and Johnson's insights about the operations performed by privative Adjs, and develops and implements them in a way that should be attractive to formal semantic theorists.

[6]For further discussion, see §4.7.2-4.7.3 below.

4.3 DUAL CONTENT and privative NPs: Implementing the account

The proposal that artifact terms have a semantic structure which includes E-structure and C-structure, and that privative Adjs are operators that modify that complex structure, can be implemented in a number of ways without affecting the basic idea behind DUAL CONTENT. The implementation I will present is attractive because it is an extension of the type-driven formal semantic framework (Heim and Kratzer [71]), widely used by linguists and philosophers. I suspect that the main reason why many theorists don't accept lexical decomposition, especially into E-structure and C-structure, is simply that our traditional formal semantic theories are not sensitive to the C-structure of lexical terms. However, only after working-out particular accounts can we really determine what complications would result from the required changes. The following account illustrates how a type-driven compositional theory can access the C-structure of terms, and shows that the required changes do not substantially complicate the general framework.

To begin, let us introduce the format we will use to represent lexical entries for Ns, which according to DUAL CONTENT include E-structure and C-structure. The C-structure of a lexical item e represents some common properties of e's extension which, as we argued above following Putnam, Lakoff and Johnson, go beyond what is captured by the stereotype associated with e. We will assume that this rich C-structure is organized around four semantic dimensions or 'qualia roles' (Moravcsik [113], Pustejovsky [149]):

- CONSTITUTIVE: the relation between the entity and its parts, what it is part of, or what it is made of.
- FORMAL: that which we use to distinguish or pick out the entity, e.g.,

shape, orientation, taste, magnitude, and color.

- TELIC: purpose or function of the entity.

- AGENTIVE: factors connected with the origin or way in which the entity came into being.

Qualia roles provide us with a powerful and flexible template for representing the meaning of Ns.[7] Not all lexical terms need to have a value for each quale, and distinct types of terms (e.g., artifact vs. natural kind vs. perceptual terms) give different weight or relative importance to each quale. For example, perceptual terms have a value for the FORMAL and, in some cases, for the CONSTITUTIVE quale, but do not have a value for the TELIC or AGENTIVE quale. Artifact terms usually have a value for the FORMAL, and assign high importance to the AGENTIVE and TELIC quale, but in many cases do not have a value for the CONSTITUTIVE quale. Natural kind terms have a value for the FORMAL, and assign high importance to the CONSTITUTIVE and AGENTIVE quale, but in many cases do not need to have a value for the TELIC quale.[8]

The idea that the meaning of lexical terms is organized around qualia roles is extensively defended by Pustejovsky [149]. However, Pustejovsky sometimes represents qualia roles as if they were 'essential' or 'definitional' properties of

[7] I should note at outset that it is not essential to DUAL CONTENT to commit to only (or precisely these) dimensions of C-structure. For the types of expressions we will consider here, these dimensions seem particularly important. However, some persuasive recent accounts on the family of DUAL CONTENT propose additional dimensions. For example, McNally [112] argues that gradable adjectives might have a scale structure dimension, and Knobe et al. [90] argue that certain concepts seem to have a normative dimension which is distinct from the TELIC dimension. I take these interesting proposals to be compatible with DUAL CONTENT.

[8] Two clarifications are in order. First, these suggestions are intended to reflect recent research on categorization and language acquisition, which suggest that what is given most weight when categorizing an artifact is the purpose for which it was created and how this relates to its function, whereas what is given most weight when categorizing a natural kind is what it is made of and/or what its natural origin is. For overviews of the data see Bloom [12], Keleman [86], Keil et al. [85], and Margolis and Laurence [108]. To properly account for categorization patterns, it is likely that we have to add weights to the dimensions or qualia roles of C-structure. For further discussion of these issues, see §4.7. Secondly, the way in which qualia roles are generalized to represent the C-structure of lexical terms that do not denote kinds of objects, such as adjectives and mass nouns, will become clear below once we deal with expressions such as *counterfeit money*, *fake lawyer*, and *fake red gun*.

the extension of terms. This aspect of Pustejovsky's theory is at odds with Putnam's lessons against definitional theories summarized above, and has been for similar reasons criticized by Asher [1], following Fodor and Lepore [50]. For these reasons, it should be clear from the outset that, by incorporating them into C-structure, DUAL CONTENT entails that qualia roles are *not*, in general, part of the extension determining aspect (E-structure) of lexical terms. In addition, as will become clear below, the specific compositional operations which, according to DUAL CONTENT, have access to qualia roles are also entirely different from the combinatorial operations used in Pustejovsky's Generative Lexicon theory. The latter operations have been persuasively criticized by Asher [1].

To illustrate DUAL CONTENT, consider the following simplified lexical entries to which we will often refer:

(4.7) *lion*

E-structure: $\lambda x.\ \textsc{lion}(x)$

C-structure:

CONSTITUTIVE: $\lambda x.\ \textsc{substance\_lion}(x)$.

FORMAL: $\lambda x.\ \textsc{perceptual\_lion}(x)$

TELIC:

AGENTIVE: $\lambda x.\ \exists e_1[\textsc{biological\_birth\_lion}(e_1, x)]$

(4.8) *gun*

E-structure: $\lambda x.\ \textsc{gun}(x)$

C-structure:

CONSTITUTIVE: $\lambda x.\textsc{parts\_gun}(x)$

FORMAL: $\lambda x.\ \textsc{perceptual\_gun}(x)]$

$\textsc{telic}$: $\lambda x.\ \textsc{gen}\ e[\textsc{shooting}(e) \wedge \textsc{instrument}(e, x)]$[9]

$\textsc{agentive}$: $\lambda x.\ \exists e_1[\textsc{making}(e_1) \wedge \textsc{goal}(e_1, \textsc{gen}\ e(\textsc{shooting}(e) \wedge \textsc{instrument}(e, x))]$

The E-structure of a lexical term e has two main roles. First, it specifies the number and types of arguments which e takes. Secondly, it stands for the (atomic) part of the meaning that determines e's extension. For now, we take the relation between the atomic E-structure and the extension of lexical terms as primitive (for further discussion, see §4.7 below). The C-structure of e, as illustrated by (4.7) and (4.8), is richer than e's associated prototype or stereotype, at least as the latter are traditionally understood. The C-structure of e can be thought of as a restricted set of general beliefs associated with e's extension, which includes, roughly, information about how entities that fall under e tend to look, feel or taste, what materials or parts they are made of, how or for what purpose they came into being or were created, and what their intended and typical function is.

The key question that any formal implementation of DUAL CONTENT needs to answer is this: how do the compositional operations interact with semantically complex lexical entries such as (4.7) and (4.8)? Type-driven theories use general compositional rules, i.e., rules that are not restricted to particular types of syntactic constructions. For our purposes, the most important are Functional Application (FA) and Predicate Modification (PM), here presented in their simplest form:

(FA) If α is a branching node, $\{\beta, \gamma\}$ is the set of α's daughters, and $[\![\beta]\!]$ is a function whose domain contains $[\![\gamma]\!]$, then $[\![\alpha]\!] = [\![\beta]\!]([\![\gamma]\!])$

[9]Note that we formulate the TELIC as involving generic or typical events in which the associated objects can be used. This is what is captured by the generic quantifier over events 'GEN'. Following Krifka et al. [94], we assume that the generic quantifier involves a contextually supplied restriction of the set of relevant events.

(PM) If α is a branching node, $\{\beta, \gamma\}$ is the set of α's daughters, and $[\![\beta]\!]$ and $[\![\gamma]\!]$ are both in $D_{<e,t>}$, then $[\![\alpha]\!] = \lambda x \in D_e.[\![\beta]\!](x) \wedge [\![\gamma]\!](x)$

To extend the type-driven framework so that it can properly interact with lexical entries which have a dual semantic structure such as (4.7) and (4.8), we have to modify FA and PM. The first obvious modification is that we need two interpretation functions, one which determines the E-structure of complex terms, which we will designate '$[\![\ ]\!]_E$', and one which determines their C-structure, which we will designate '$[\![\ ]\!]_C$'. In addition, we will assume that the original '$[\![\ ]\!]$' retrieves the entire meaning of expressions as an ordered set of the E-structure and C-structure. For example, $[\![e]\!]_E$ is just the E-structure of e, $[\![e]\!]_C$ is an ordered set of the C-structure of e, and $[\![e]\!]$ is an ordered set of the E-structure and the CONSTITUTIVE, FORMAL, TELIC and AGENTIVE of e.[10]

For clarity, we first discuss how the E-structure of complex NPs is determined by $[\![\ ]\!]_E$. Here is the first version of the new rules, specified only for E-structure, and focusing only on FA:

(FA^{DC}) If α is a branching node, $\{\beta, \gamma\}$ is the set of α's daughters, and $[\![\beta]\!]_E$ is a function whose domain contains $[\![\gamma]\!]$, then $[\![\alpha]\!]_E = [\![\beta]\!]_E([\![\gamma]\!])$ and $[\![\alpha]\!]_C = \ldots$

This minor modification of FA is required to specify what aspect of the semantic structure of lexical terms the interpretation function is retrieving and calculating. As we will see, for some types of complex NPs the E-compositional operations work just like the traditional compositional operations. In other cases, the E-compositional operations are more complex, which is why they need to take the entire meaning of expressions as arguments.

[10]For now, we will assume that terms that do not have a value for one or more of these dimensions—that are, in this sense, unconstrained—represent this via the identity function. If the terms are predicate modifiers instead of predicates, we can represent the empty value as a function maps properties unto themselves.

How do we bring the C-structure of lexical terms into the E-compositional operations in cases when we need to do so, such as in certain types of adjectival modifications? To do this, we introduce two tools:[11]

1. 'Qualia functions': partial functions from the meaning of terms into their respective C-structure denotations, namely, CONSTITUTIVE, FORMAL, TELIC, and AGENTIVE. The qualia functions are Q_C, Q_F, Q_T, Q_A. Consider the following examples, based on lexical entries (4.7) and (4.8):

   (4.9) $Q_C([\![lion]\!]) = \lambda x.\ \textsc{substance\_lion}(x)$

   (4.10) $Q_T([\![gun]\!]) = \lambda x.\ \textsc{gen}\ e[\textsc{shooting}(e) \wedge \textsc{instrument}(e, x)]$

   In addition, we will include a function Q_E which takes the meaning of terms and returns their E-structure denotations.

2. 'Semantic re-structuring operators': partial functions from the meaning of terms into certain combinations of their E-structure and C-structure. The operators are C, T, A, and F. Consider the following examples:

   (4.11) $A([\![lion]\!]) = \lambda x.\ Q_E([\![lion]\!])(x) \wedge Q_A([\![lion]\!])(x)$
   $= \lambda x.\ \textsc{lion}(x) \wedge \exists e_1[\textsc{biological\_birth\_lion}(e_1, x)]$

   (4.12) $T([\![gun]\!]) = \lambda x.\ Q_E([\![gun]\!])(x) \wedge Q_T([\![gun]\!])(x)$
   $= \lambda x.\ \textsc{gun}(x) \wedge \textsc{gen}\ e[\textsc{shooting}(e) \wedge \textsc{instrument}(e, x)]$

Using these tools, let us consider the case of privative NPs such as *fake gun*. As we saw above, the E-compositional effect of *fake* is more complex than the E-compositional effect of intersective and subsective Adjs. The main difference is that, according to DUAL CONTENT, *fake* interacts with various aspects of the E and C-structure of the term it modifies. For this reason, I propose that we

[11] Jensen and Vikner [207] also use *similar* tools in their account of genitives. Although their account and DUAL CONTENT are roughly in the same family, there are some important difference between their version of qualia functions and semantic restructuring operators and the versions presented below.

treat *fake* as a semantic re-structuring operator along the lines of C, T, A, and F. But unlike those operators, *fake* and other privatives are *lexicalized* semantic restructuring operators. To allow lexicalized semantic restructuring operators into the composition process, we need to incorporate Qualia functions into the specification of their E-structure.

Following this idea, we can take $[\![fake]\!]_E$ as an operator which takes the meaning of the noun it modifies and restructures it by using—via Qualia functions—the noun's C-structure. Here's a first try:

(4.13) $[\![fake]\!]_E = \lambda D_C.\ [\lambda x.\ \neg Q_E(D_C)(x) \wedge \neg Q_A(D_C)(x) \wedge \exists e_2[\textsc{making}(e_2) \wedge \textsc{goal}(e_2, Q_F(D_C)(x))]]]$

Note that 'D_C' is a variable that ranges over ordered sets of the E-structure and C-structure of common Ns. So according to (4.13), $fake$ takes the entire meaning, as an ordered set of the E-structure and all the elements of the C-structure, and outputs a new predicate of type $< e, t >$ that is not a $Q_E(D_C)$, was not created with the same goal specified in $Q_T(D_C)$, but was created to have the perceptual features specified in $Q_F(D_C)$. Consider then $[\![fake\ gun]\!]_E$, which by FA^{DC}, is equal to $[\![fake]\!]_E([\![gun]\!])$:

(4.14) $[\![fake]\!]_E([\![gun]\!])$

$= \lambda x.\ \neg Q_E([\![gun]\!])(x) \wedge \neg Q_A([\![gun]\!])(x) \wedge \exists e_2[\textsc{making}(e_2) \wedge \textsc{goal}(e_2, Q_F([\![gun]\!])(x))]]$

$= \lambda x.\ \neg\textsc{gun}(x) \wedge \neg\exists e_1(\textsc{making}(e) \wedge \textsc{goal}(e_1, Q_T([\![gun]\!])(x))) \wedge \exists e_2[\textsc{making}(e_2) \wedge \textsc{goal}(e_2, \textsc{perceptual\_gun}(x))]$

According to (4.14), $[\![fake\ gun]\!]_E$ is a predicate of type $< e, t >$, which is satisfied by objects that are not guns, were not made with the goal of being shooting instruments, and were made to have the perceptual features specified in $Q_F([\![gun]\!])$, i.e., to look like guns. This is exactly the result we wanted to get. Given appropriate lexical entries, this will get the intuitively correct results for *fake knife*, *fake headphones*, *fake pen*, and so on.

The entry presented in (4.13) captures several uses of *fake* that might initially seem quite different. For example, there is a sense of *fake* that is close to that of *counterfeit*. This use is illustrated in expressions such as [*fake* [*Chanel handbag*]] and [*fake* [*Polo shirt*]]. To account for these cases, the key thing to note is that (4.13) does *not* entail that a fake ϕ cannot have the function of a ϕ: in (4.13) the TELIC of the modified N is not negated. What (4.13) entails, however, is that a fake ϕ cannot have the same source as ϕ, i.e., cannot have the same AGENTIVE as ϕ. In the case of brand name artifacts, certain aspects of their institutional source are crucial, hence are likely included in their AGENTIVE. So (4.13) gives the intuitively correct results: e.g., fake Polo shirts and fake Chanel bags do not come from the required authorized sources, although they can be used as if they did. Now, phrases headed by artifact Ns modified by brand name terms are productive: we can understand *Chanel motorcycle* and *Polo cell-phone* even if the relevant products do not yet exist. This suggests that, in general, brand-name modifiers add a condition to the AGENTIVE of the modified N. Since *fake* negates that AGENTIVE, this entails that [*fake* [*Chanel handbag*]] can also mean a fake handbag made by Chanel. That this reading should be available is perhaps more obvious in the case of novel products. We can easily imagine that Chanel made a fake motorcycle for an exhibition, or that Polo made a fake cell-phone for some runway show.[12]

In addition, (4.13) also works well for expressions that denote fake gestures and fake social roles, e.g., *fake smile* and *fake lawyer*. A fake smile is intentionally produced to look like a smile, just like the role of a fake lawyer is usually to behave or look like a lawyer. Both cases, however, clearly have a different

[12]To be clear, this second reading is not possible under the structure [[*fake Chanel*] *handbag*], which is also available. We deal in more detail with recursive NPs below—including modifications of complex privative NPs and different syntactic structures.

source or AGENTIVE than their real counterparts. The source of a real smile involves a certain sensation that is lacking in the fake version, although the intention is to exhibit the same outward features. The source of a real lawyer—i.e., how one becomes a lawyer—includes certain socially prescribed milestones (e.g., finishing law school, passing the bar, etc.). A fake lawyer has the outward appearance and behavior of a lawyer (e.g., 'looks like a lawyer'), and can even carry out some of the same or similar functions of a lawyer. However, a fake lawyer lacks the AGENTIVE of a real lawyer—and since that is what gives lawyers their authority, fake lawyers do not have the relevant legitimacy.

These examples deserve more discussion, but for now what I want to highlight is that DUAL CONTENT, combined with an entry like (4.13), can account for the sense that *fake* is doing something at least slightly different in, e.g., *fake gun*, *fake smile*, and *fake Chanel handbag*. This is explained on the basis of the important differences in the AGENTIVE of *gun*, *smile*, and *Chanel handbag*. In each case, *fake* itself is carrying out essentially the same operation, which includes negating the AGENTIVE of the modified N. However, different types of terms, including subtypes of artifacts, have different sorts of information in their AGENTIVE, which accounts for the intuitive differences.[13]

Privative Adjs in general can be treated as lexicalized semantic restructuring operators. Indeed, this approach has special advantages when their differences are subtle. Take the case of *counterfeit* and *artificial*, which contrast in interesting ways with *fake*. An important difference between many paradigmatic uses of *counterfeit* and *fake* is that, unlike a fake, a counterfeit ϕ is usually made to look and function like a ϕ. For example, a counterfeit Rolex is made both to look like and function like a Rolex. *Counterfeit*, in its most usual sense,

[13]Even if we accept DUAL CONTENT, there are cases that crucially involve some degree of context-induced meaning modulation. For discussion of cases when the meaning of *fake* itself is modulated, and more generally on the relation between DUAL CONTENT and contextualism, see §4.7.1 below.

can be applied to terms for artifacts whose origin gives them an added value, with the result that this appropriate origin is lacking in the counterfeit versions. Consider the following entry:

(4.15) $[\![counterfeit]\!]_E = \lambda D_C.\ [\lambda x.\ \neg Q_E(D_C)(x) \wedge \neg Q_A(D_C)(x) \wedge$
$\exists e_2[\textsc{making}(e_2) \wedge \textsc{goal}(e_2, Q_F(D_C)(x) \wedge Q_T(D_C)(x))]]$

If we apply (4.16) to a plausible lexical entry for *Rolex*, we get that a *counterfeit Rolex* is not a Rolex, does not come into being in the manner in which a Rolex comes into being, as specified in $Q_A([\![Rolex]\!])$, and is an artifact that was made to look *and* function like a Rolex, as specified in $Q_F([\![Rolex]\!])$ and $Q_T([\![Rolex]\!])$ respectively.[14] Consider, on the other hand, the case of *artificial*, as used in expressions such as *artificial leg* and *artificial heart*. Unlike a fake heart, an artificial heart is made with the intention that it function like a heart. In addition, unlike fakes and counterfeits, an artificial heart need not be made to look like a heart, as long as it was made to function like one. The following entry captures this important distinction between *artificial*, on the one hand, and *fake/counterfeit* on the other:

(4.16) $[\![artificial]\!]_E = \lambda D_C.\ [\lambda x.\ \neg Q_E(D_C)(x) \wedge \neg Q_A(D_C)(x) \wedge$
$\exists e_2[\textsc{making}(e_2) \wedge \textsc{goal}(e_2, Q_T(D_C)(x))]]$

Unlike the case of *fake* and *counterfeit*, in this case we might want to eliminate '$\neg Q_E(D_C)(x)$' from (4.16). Indeed, some people I have informally queried about whether, say, artificial hands and legs are really arms and legs report mixed and unstable intuitions. This is of course an empirical question, and both options are compatible with DUAL CONTENT. We reserve further discussion of the descriptive accuracy of these proposals, including a discussion of the ways in which DUAL CONTENT and contextualism are compatible, for §4.7.1 below.

[14] The difference between *fake* and *counterfeit*, on this account, is quite subtle: counterfeit artifacts are explicitly made to function like the artifacts they are counterfeits of. This is not required in the case of fakes, although in some cases it is possible to so use them.

In addition to privatives, other types of Adjs can also be treated as lexicalized semantic restructuring operators. Consider subsective Adjs such as *typical*, *representative*, or *normal*, as used in expressions such as *typical lion/gun*, *representative French/Harvard Alumni*, *normal lion/gun*, etc. We can treat these Adjs as operators which take the full meaning of Ns and return a predicate of type $< e, t >$ which includes in its E-structure a part of its C-structure. Consider the case of *typical*:

(4.17) $[\![typical]\!]_E = \lambda D_C.\ [\lambda x.\ Q_E(D_C)(x) \wedge Q_C(D_C)(x) \wedge Q_F(D_C)(x) \wedge Q_T(D_C)(x) \wedge Q_A(D_C)(x)]$

According to (4.17), *typical* takes D_C and outputs a new predicate that is satisfied by objects that fall under $Q_E(D_C)$, $Q_F(D_C)$, $Q_T(D_C)$, and so on. So $[\![typical\ gun]\!]_E$, which by FA is equal to $[\![typical]\!]_E([\![gun]\!])$, would be:

(4.18) $[\![typical]\!]_E([\![gun]\!])$

$= \lambda x.\ Q_E(D_C)(x) \wedge Q_C(D_C)(x) \wedge Q_F(D_C)(x) \wedge Q_T(D_C)(x) \wedge Q_A(D_C)(x)]$

$= \lambda x.\ [\text{GUN}(x) \wedge \text{PARTS\_GUN}(x) \wedge \text{PERCEPTUAL\_GUN}(x) \wedge$

$\text{GEN}\ e[\text{SHOOTING}(e) \wedge \text{INSTRUMENT}(e, x)] \wedge \exists e_1[\text{MAKING}(e_1) \wedge$

$\text{GOAL}(e_1, \text{GEN}\ e(\text{SHOOTING}(e) \wedge \text{INSTRUMENT}(e, x))]]$

According to (4.18), a *typical gun* is a gun that has all the common features of a gun, as represented in the C-structure of *gun*: it looks like a gun, can be used for shooting, was made to be used for shooting, and is made out of whatever materials and parts guns are typically made out of (which, for simplicity, we represented as PARTS_GUN). Now, we said before that some default lexical entries might lack a quale value in their C-structure. In addition, *typical* does not seem to have any restrictions with the type of common Ns it can combine with. However, in (4.17) *typical* is defined over all the values of the qualia. There are two simple ways to deal with this. One is to assume that *typical* is ambiguous: one of its senses operates over the qualia roles defined for artifacts, the other over the qualia roles defined for natural kinds, and so on. Another is to assume that *typical* has the meaning proposed in (4.17) but that it coerces the meaning of the lexical terms it modifies to specify a value for each of the qualia roles, and that we use our background (encyclopedic) knowledge to do this online.

The discussion of *typical* raises two further questions about *fake*. The first is how to deal with cases when *fake* modifies a N that is not an artifact term, or more precisely, one that is undefined with respect to its TELIC or AGENTIVE. Just as in the case of *typical*, there are at least two ways to deal with this, which are not mutually exclusive and could sometimes work together. The first is to assume that the meaning of the modified head N is modulated online: using general encyclopedic information, it is enriched to include a value for the TELIC and AGENTIVE. The second is to assume that the meaning of *fake* is modulated so that it can apply to the head N. This could be proposed to deal with modifications of mass terms such as *gold* and *titanium*, which, some would argue, don't seem to have a value for the AGENTIVE. The modulated meaning of *fake* in these cases could be something like:

(4.19) $[\![fake_m]\!]_E = \lambda D_C.\ [\lambda x.\ \neg Q_E(D_C)(x) \wedge \neg Q_C(D_C)(x) \wedge$
$\exists e_2[\textsc{making}(e_2, x) \wedge \textsc{goal}(e_2, Q_F(D_C)(x))]]$

One difference between *fake* and *fake*$_m$ is that the former negates the AGENTIVE of the term it modifies and the latter negates the CONSTITUTIVE. This would explain why a fake gun can be made of steel but could not have been made to shoot perfectly, whereas fake gold cannot be made of gold but could be used, in some cases, to pass for gold. However, we need not assume this kind of meaning modulation of *fake*. We can still explain cases such as *fake gold* and *fake titanium* by insisting that the meaning of *fake* is as in (4.13). To do so we need to assume that, either in the context of privative NPs or in general, substance terms such as *gold* and *titanium* have an AGENTIVE. This is a reasonable move: for we could represent in their AGENTIVE the information that gold and titanium 'come from' particular molecular compositions. In this case, (4.13) would give the intuitively correct results for *fake gold* and *fake titanium*, namely, that they have some of the perceptual features of gold and titanium, but are not made up of the same stuff as gold and titanium.[15]

The other question raised by the discussion of *typical* is how to deal with modifications of complex privative NPs. We must account for the meaning of expressions such as *typical fake gun.* We said that *typical* takes the full meaning of a lexical term, which involves operations on its C-structure. If this is correct, it follows that *fake gun* must itself have a C-structure. To account for this, we first have to fully define FA^{DC}, since so far we have only defined it for the E-structure of terms.

(FA^{DC}) If α is a branching node, $\{\beta, \gamma\}$ is the set of α's daughters, and $[\![\beta]\!]_E$ is a function whose domain contains $[\![\gamma]\!]$, then $[\![\alpha]\!]_E = [\![\beta]\!]_E([\![\gamma]\!])$ and $[\![\alpha]\!]_C =< Q_C([\![\beta]\!])([\![\gamma]\!]), Q_F([\![\beta]\!])([\![\gamma]\!]), Q_T([\![\beta]\!])([\![\gamma]\!]), Q_A([\![\beta]\!])([\![\gamma]\!]) >$

FA^{DC} captures a rather simple idea about C-structure combinatorics, namely, that the relevant quale of the modifier applies via functional application to the corresponding quale of the modified expression.[16] To do this, we assume that

[15] I would argue that this second route is the better bet. If we follow this route, what would be stored in the CONSTITUTIVE is information about the 'identifiable' part-whole structure; in this case, that mass terms do not have identifiable parts. Treating the CONSTITUTIVE in this way would also allow us to incorporate into DUAL CONTENT some of the suggestions for color and degree modifications presented by McNally [112]. Then the information that we we intuitively regard as 'what substances are made of'—e.g., their molecular composition—would, when available, go in the AGENTIVE. Furthermore, it is worth noting that this option interacts nicely with some recent views on artifactual substances. For example, Bloom [13] argues that some substances can be used in an artifactual sense. One way to represent this in DUAL CONTENT is to hold that their AGENTIVE has two conditions: the molecular composition of the substance and the production source. In this case, DUAL CONTENT would predict that some substance could be *fake Coke* or *fake Sprite* either because because of differences in composition with the originals, or because it does not come from, say, the companies that are authorized to make Coke or Sprite. If we follow Bloom, this result seems along the right lines.

[16] A reviewer suggests that this point-wise formulation of FA^{DC} might overcommit DUAL

each quale in the C-structure of semantic-restructuring operators can take as argument the full meaning of the N they modify. Consider then the full lexical entry for $fake$:

(4.20) *fake*

E-structure: $\lambda D_C.\ [\lambda x.\ \neg Q_E(D_C)(x) \wedge \neg Q_A(D_C)(x) \wedge$
$\exists e_2[\textsc{making}(e_2) \wedge \textsc{goal}(e_2, Q_F(D_C)(x))]]$

C-structure:

CONSTITUTIVE: $\lambda D_C.\ Q_C(D_C)$

FORMAL: $\lambda D_C.\ Q_F(D_C)$

TELIC: $\lambda D_C.\ \neg Q_T(D_C)$

AGENTIVE: $\lambda D_C.\ [\lambda x.\ \exists e_2[\textsc{making}(e_2) \wedge \textsc{goal}(e_2, Q_F(D_C)(x)]]$

Given (4.20) and (4.8), by FA^{DC} the full entry for *fake gun* is:

(4.21) *fake gun*

E-structure: $\lambda x.\ \neg Q_E([\![gun]\!])(x) \wedge \neg Q_A([\![gun]\!])(x) \wedge$
$\exists e_2[\textsc{making}(e_2) \wedge \textsc{goal}(e_2, Q_F([\![gun]\!])(x))]$

C-structure:

CONSTITUTIVE: $Q_C([\![gun]\!])$

FORMAL: $Q_F([\![gun]\!])$

TELIC: $\neg Q_T([\![gun]\!])$

CONTENT to a particularly strong view of how the C-structures of complex NPs in general are determined. For the cases we are considering, this version of FA^{DC} works well. In addition, in DUAL CONTENT modifiers take the full meanings of heads and can use qualia functions to shuffle the internal semantics, so this gives some extra flexibility to this account even if we stick to the point-wise combinatorics. However, it is undeniable that other types of NPs might involve different combinatorial dynamics. If this turns out to be the case, we can replace the current pointwise formulation of FA^{DC} with a more general formulation, which just says that $[\![\alpha]\!]_C$ is a function f of $[\![\beta]\!]_C$ and $[\![\gamma]\!]_C$. We might then further specify that for the types of expressions we are considering (i.e., subsectives and privatives), what f does is to perform the previously specified function application along each dimension of C-structure.

$$\text{AGENTIVE: } \lambda x.\ \exists e_2[\text{MAKING}(e_2) \wedge \text{GOAL}(e_2, Q_F([\![gun]\!])(x)]$$

We can now compute $[\![typical\ fake\ gun]\!]_E$, which has the structure [*typical* [*fake gun*]], hence, by FA:

(4.22) $[\![typical]\!]_E([\![fake\ gun]\!]) =$

$$\lambda x.\ Q_E([\![fake\ gun]\!])(x) \wedge Q_C([\![fake\ gun]\!])(x) \wedge Q_F([\![fake\ gun]\!])(x) \wedge Q_T([\![fake\ gun]\!])(x) \wedge Q_A([\![fake\ gun]\!])(x)$$

Comparing (4.14) with (4.22), we can see that there are important differences between *fake gun* and *typical fake gun.* According to (4.14), a *fake gun* is not a gun, was not made to function for shooting, and was made to seem like a gun. However, there are malfunctioning fake guns and badly made fake guns, which is part of the reason why we did not say anything about whether or not fake guns can actually shoot or not look like real guns. We can imagine a badly made fake gun that can shoot or that can be easily identified as a fake gun. However, a *typical* fake gun, as correctly captured in (4.22), is an artifact that looks like a gun and that cannot be used to shoot.

Just like *typical*, *fake* can modify complex NPs of the same type as *gun* such as *red gun* and *heavy gun.* So just like privative NPs, intersective and subsective NPs must also have an E-structure and a C-structure. There are various options for how to determine the C-structure of intersective, subsective and other types of complex NPs. Indeed, there is plenty of research on how to determine the meaning, including the C-structure, of complex expressions when their constituents also have complex meanings [68, 112, 131, 144, 149, 151]. Although some look very promising, most of the proposals are preliminary. Hence we cannot yet provide a full account of how to determine, for each type of complex NP, their C-structure from the meanings of their parts. However, we can show that, at least for paradigmatic intersective and subsective NPs, this is not an in-principle problem. Furthermore, some of the most plausible suggestions can be implemented in the DUAL CONTENT framework.

Consider the following lexical entry for *plastic* (for brevity, we simplify the representations of the values of C-structure):

(4.23) *plastic*

E-structure: $\lambda x.$ PLASTIC(x)

C-structure:

CONSTITUTIVE: $\lambda x.$ PLASTIC_MASS(x)

FORMAL: $\lambda x.$ PERCEPTUAL_PLASTIC(x)

TELIC:

AGENTIVE: $\lambda x.$ PLASTIC_COMP(X)[17]

We can easily come up with a modified rule of PM that can be used to combine (4.23) with *gun* and get the appropriate C-structure for *red gun*. However, so that we do not have to further modify the general compositional operations we have introduced so far to account for the E-structure and C-structure of complex privative NPs, we can assume that when used attributively, *plastic* and other intersective Adjs are type-shifted into semantic restructuring operators, similar to the case of privative Adjs, except that they are simpler.

(4.24) *plastic*

[17] This entry follows the suggestion—made above in the discussion of *gold*, *fake gold*, etc.—that the CONSTITUTIVE of mass terms only specifies that they do not have identifiable parts. The information about their molecular composition (e.g., that they come from a certain combination of elements), when available, goes into the AGENTIVE. Now, it might seem strange to distinguish, as we do in this entry, between the E-structure and the AGENTIVE of *plastic*. Various theorists argue that the essence of substances is just the stuff they are made of. Although we could implement this view within DUAL CONTENT, it is worth noting that it is not uncontroversial (see e.g. Bloom [13]), and it is arguably implausible for the case of *plastic*. In addition, there are cases of intersective Adjs that seem to have an E-structure that cannot be identified with any particular dimension of their C-structure. For example, color Adjs might have E-structures with functions of type $< e, t >$ that are uncommitted to the particular quantity/quality of the way in which entities that fall under them have to be colored. Still, these quantity/quality dimensions are arguably part of the C-structure of colors, i.e., part of the typical way in which we conceive of objects being colored.

E-structure: $\lambda D_C.\ [\lambda x.\ \textsc{plastic}(x) \wedge Q_E(D_C)(x)]$

C-structure:

constitutive: $\lambda D_C.\ [\lambda x.\ \textsc{plastic\_mass}(x) \wedge Q_C(D_C)(x)]$

formal: $\lambda D_C.\ [\lambda x.\ \textsc{perceptual\_plastic}(x) \wedge Q_F(D_C)(x)]$

telic: $\lambda D_C.\ Q_T(D_C)$

agentive: $\lambda D_C.\ [\lambda x.\ \textsc{plastic\_comp}(x) \wedge Q_A(D_C)(x)]$

By FA^{DC}, if (4.24) is combined with *gun*, as specified in (4.8), we get the following result:

(4.25) *plastic gun*

E-structure: $\lambda x.\ \textsc{plastic}(x) \wedge \textsc{gun}(x)$

C-structure:

constitutive: $\lambda x.\textsc{plastic\_mass}(x) \wedge \textsc{parts\_gun}(x)$

formal: $\lambda x.\ \textsc{perceptual\_plastic}(x) \wedge \textsc{perceptual\_gun}(x)$

telic: $\lambda x.\ \textsc{gen}\ e[\textsc{shooting}(e) \wedge \textsc{instrument}(e, x)]$

agentive: $\lambda x.\ \textsc{plastic\_comp}(x) \wedge \exists e_1[\textsc{making}(e_1) \wedge$
$\textsc{goal}(e_1, \textsc{gen}\ e(\textsc{shooting}(e) \wedge \textsc{instrument}(e, x))]$

It is easy to see that *fake* can apply to *plastic gun*, following the procedure used in the previous cases: a fake plastic gun is, in the most salient reading, a plastic artifact that is a fake gun, in the sense explained above. Also available, on this view, is the reading according to which a *fake plastic gun* is a real gun that is made of plastic. Now, note that sometimes *fake* combines first with another N used as a modifier, and only then with the head N. In this case, a *fake plastic gun* or *fake steel gun* can only mean a real gun that is made of fake plastic or fake steel. This is often the intended meaning in cases such as *fake gold ring* and *fake diamond necklace*. Given the correct underlying structure—e.g., [[*fake* *plastic*] *gun*], [[*fake gold*] *ring*], etc.—and lexical entries, DUAL CONTENT easily accounts for these cases.

This example is only meant to illustrate a simple case of intersective composition, but it does suggest a general strategy for calculating the E-structure and C-structure of intersective NPs. Namely, to assume that the lexical entries of modifiers provide, in their C-structure, information about how and which qualia of the N they combine with should be modified, and which qualia should be directly inherited from the modified N. Of course, within this strategy, different types of modifiers demand slightly different treatments. Some modifiers might raise special challenges for the composition of the C-structure of their corresponding complex NPs.[18] Constructing a full compositional account to determine the C-structure of all types of complex NPs is an ongoing project—but one which is constantly advancing and, despite some early challenges, most notably by Fodor [43] and Fodor and Lepore [50], does not seem to face any deep in-principle problems.[19]

To further develop and defend DUAL CONTENT, I will compare it with another account of privative NPs, one which can be combined with atomistic lexical semantic theories (§4.4-§4.6). I then conclude by addressing some objections to DUAL CONTENT (§4.7).

[18]For example, in some cases the modifier might contribute a quale value which is inconsistent with the value of the head N. This might happen in cases such as *purple dollar-bill*, since it is plausible to assume that the colors specified in the FORMAL for *dollar-bill* are inconsistent with *purple*. In these cases, there must be a mechanism which resolves these inconsistencies, favoring the value of the modifier. This kind of solution is compatible with DUAL CONTENT.

[19]For an interesting response to the compositionality challenge to prototype theory, see Chapter 5 below and Prinz [144]. See also Part V of *The Oxford Handbook of Compositionality*, which has various essays that address issues related to the combinatorial properties of lexical items with complex semantic structures.

4.4 Contextualist account of privative NPs

Partee [127,129] presents a contextualist account of privative NPs which can also be implemented in a standard type-driven semantic framework. In addition, her account, which has recently been gaining approval (Recanati [164], Morzycki [114]), can be paired with an atomic account of the internal semantic structure of head Ns. So before we can confidently hold that privative NPs motivate DUAL CONTENT's account of lexical decomposition, we have to evaluate the prospects of Partee's account.

Partee's account rests on two simple proposals: (i) eliminate the class of privative Adjs, and (ii) treat privatives as subsective Adjs that coerce their N arguments into systematic meaning expansions. Take the following examples:

(4.26) Is that gun real or fake?

(4.27) A fake gun is not a gun

Partee argues that when interpreting (4.26) we expand the meaning of *gun* to include both fake guns and real guns. In (4.27), the same is true of the first occurrence of *gun*, modified by *fake*, but not of the second. In short, Partee's account amounts to the following readings of (4.26) and (4.27):

(4.26) Is that gun_2 real or fake?

(4.27) A fake gun_2 is not a gun_1

Subscript '1' marks 'literal' interpretations and '2' marks 'loose' interpretations. For example, the extension of gun_2 includes not only real guns but also things that look like guns, such as toy guns and fake guns.

Cases like (4.26) are neutral between DUAL CONTENT and Partee's account. In many ordinary contexts we use artifact Ns such as *gun* in the loose sense of gun_2—even when we use them in NPs that don't have modifiers such as *fake* or *real*. Some examples are when *gun* appears in demonstrative expressions such as (4.26), and phrases like *that red gun*, *that shiny gun*, etc. In general, when such demonstratives are used to pick out an object that loosely looks like a gun, they are taken as properly used, whether or not the object is a real gun. In these and many other cases, artifact Ns are interpreted in the loose sense. DUAL CONTENT is compatible with this fact.[20]

Consider next examples like (4.27). According to Partee, we should treat privative Adjs as subsectives that coerce certain meaning-shifts in the Ns which they modify, so we need an account of what determines the direction of such meaning-shifts. Partee appeals to two plausible principles of interpretation [131]:

(NON-VACUITY) In any given context, try to interpret any predicate so that both its positive and negative extension are non-empty.

(HEAD-PRIMACY) In a modifier-head structure, the head is interpreted relative to the context of the whole constituent, and the modifier is interpreted relative to the local context created from the former context by the interpretation of the head.

NON-VACUITY applies not only to simple predicates, but also to complex predicates, e.g., to predicates obtained via a Adj+N combination. HEAD-PRIMACY ensures that the interpretation of a head N in a given context restricts the local domain for the interpretation of the Adj to the positive extension of the head N in that context.

Partee is surely right that NON-VACUITY and HEAD-PRIMACY help explain certain patterns of context-induced meaning-shifts. Evidence for HEAD-PRIMACY comes from examples like these:

[20] As we will discuss in detail in §4.7, DUAL CONTENT is compatible with contextualist compositional accounts—such as that defended in Recanati [164]—which hold that the literal meaning of the constituents of complex expressions is usually affected, to some degree, by meaning modulations. However, the issue in this discussion is about the correct account of the default interpretation of privative NPs, and in this particular case DUAL CONTENT, unlike SHIFTING HEADS, does not invoke meaning modulations.

(4.28) (i) giant midget

(ii) midget giant

In most contexts, we take (4.28i) to mean a midget, but a very large one, and (4.28ii) to mean a giant, but a very small one. Such examples suggest that in Adj+N constructions, one first interprets the head N against the general context of the utterance, and then shifts the meaning of the modifying Adj so that—to respect NON-VACUITY—it subsect the extension of the head N.

There is also plenty of evidence for NON-VACUITY. NON-VACUITY is a particular manifestation of the rational pursuit of the basic goals of communication, as conceived in the Gricean tradition. In some contexts, NON-VACUITY can be taken as a consequence of general principles of cognitive processing, e.g., the Relevance Principle [183]. Overall, NON-VACUITY is a descriptively adequate information maximization principle of interpretation, which might follow from certain general cognitive biases and/or basic goals of rational communication. We will follow Partee and assume that something close to NON-VACUITY is a pragmatic principle of interpretation.

The next step in Partee's account of the direction of the meaning-shifts of the heads of privative NPs is a proposal about how NON-VACUITY and HEAD-PRIMACY interact. NON-VACUITY and HEAD-PRIMACY usually work together to partly determine the pragmatic interpretation of complex expressions, but sometimes satisfying one principle requires violating the other. In cases of conflict, Partee argues that NON-VACUITY ranks higher than HEAD-PRIMACY.

There is a productive class of complex expressions that supports Partee's proposed ranking, called the 'constitutive material NPs', e.g., *stone lion*, *wooden horse*, and *rubber duck*. When interpreting these expressions, we automatically and effortlessly loosen the meaning of the head Ns. For example, we generally interpret *stone lion* not as a real lion made of stone, but as a lion representation which is made of stone (e.g., a lion stone-sculpture). This suggests that in *stone lion* we interpret *lion* as $lion_2$—i.e., as including lions and things that only look like lions. In these cases, the direction of the meaning-shift of the head Ns seems to be determined by NON-VACUITY. For most mature speakers strongly believe that, e.g., (in most context) the set $\{x : x$ is a lion and x is made of stone$\}$ is empty, whereas the set $\{x : x$ is a lion or representation of a lion and x is made of stone$\}$ includes some objects and excludes others.

In short, Partee proposes that the default interpretation of certain constitutive material NPs involves the following processing sequence: (i) triggered by a violation of NON-VACUITY, the interpretation proceeds to (ii) coerce the meaning of the head N such that when composed with the literally taken modifying A, (iii) the resulting expression is believed to respect NON-VACUITY, and (iv) this meaning-shift occurs despite the fact that HEAD-PRIMACY is thereby violated. This is a persuasive account of constitutive material NPs; we will assume in what follows that it is essentially correct.

Partee's key suggestion is that the principles involved in the default interpretation of privative NPs are the same as those involved in the default interpretation of constitutive material NPs. To illustrate, consider the process which, according to Partee, leads to the default interpretation of *fake gun*. Assume that $[\![gun_1]\!]$ is a function of type $< e, t >$, $[\![fake]\!]$ is a function of type $<< e, t >, < e, t >>$, and that the extension of *fake* gun_1 is $\{x : x$ is a fake (thing) and x is a gun$\}$. According to Partee, in most contexts mature speakers (firmly) believe that this set is empty, i.e., that there are no objects which are both real guns and fake things. If so, *fake* gun_1—the literal interpretation of *fake gun*—violates NON-VACUITY, which triggers a pragmatic meaning-shifting process. Partee implicitly assumes that there is no 'easy' way of respecting both NON-VACUITY and HEAD-PRIMACY by shifting the meaning of $fake$.[21] Since

[21] It is not clear why Partee does not consider this route (i.e., loosening the meaning of the

NON-VACUITY ranks higher than HEAD PRIMACY, the next option is to shift the meaning of *gun* to the expanded gun_2. Most speakers believe that the set $\{x : x$ is a fake (thing) and x is a gun or looks like a gun$\}$ includes some objects and excludes others. So by interpreting $fake\ gun$ as $fake\ gun_2$ we satisfy NON-VACUITY.

Call Partee's account of privative NPs, 'SHIFTING HEADS'. We will now examine some cases in which DUAL CONTENT and SHIFTING HEADS make different predictions about the interpretation of privative NPs, and argue that DUAL CONTENT makes the correct predictions.

4.5 DUAL CONTENT vs SHIFTING HEADS

The main difference between DUAL CONTENT and SHIFTING HEADS is this:

- According to SHIFTING HEADS, pragmatically-induced meaning-shifts of the head Ns, guided by principles such as NON-VACUITY, are an essential part of the process leading to the default interpretation of privative NPs. These are the default interpretations because in most contexts we (explicitly or implicitly) follow NON-VACUITY and believe that *fake* gun_1 and the like have an empty extension.

- According to DUAL CONTENT, the default interpretation of privative NPs does not involve a violation of NON-VACUITY: e.g., $[\![fake]\!]$ takes gun_1 and returns the predicate specified in (4.14); a predicate which most mature speakers (correctly) believe has a positive (e.g., fake guns) and a negative extension. According to DUAL CONTENT, the default interpretation of *fake*

privative Adj instead of that of the head N), esp., since this would satisfy both NON-VACUITY and HEAD-PRIMACY. In many Adj+N constructions, we modify the meaning of the Adj rather than that of the head N. Maybe Partee does not consider this option because she thinks that—unlike the case of, say, color Adjs—it is not easy to come up with a 'loose' interpretation of $fake$ and other privative Adjs. However, given certain plausible proposals for lexical entries for $fake$, this is a very problematic assumption, as we will see in §4.6.

gun is based on the literal interpretation of the parts and rests only on linguistic competence.

This contrast suggests a way of comparing these accounts. NON-VACUITY is a pragmatic principle of interpretation. Like other such principles, we can relax or suspend it in certain contexts. In such contexts, SHIFTING HEADS predicts that the interpretation of privative NPs should differ from their default interpretation. In contrast, DUAL CONTENT doesn't predict this, for according to it the default interpretation of privative NPs is roughly the literal interpretation. Furthermore, according to SHIFTING HEADS the process that determines the default interpretation of privative NPs is analogous to the process that determines the default interpretation of constitutive material NPs. So SHIFTING HEADS predicts that, when NON-VACUITY is relaxed, the interpretations of privative NPs and of constitutive material NPs should be affected in similar ways.[22]

This prediction, however, is incorrect. When NON-VACUITY is suspended or relaxed, the interpretation of privative NPs is affected in a completely different way from the interpretation of constitutive material NPs, as predicted by DUAL CONTENT.

Consider first the interpretation of constitutive material NPs. The default interpretation of expressions like *stone lion* and *plastic rabbit* is *stone* $lion_2$ and *plastic* $rabbit_2$. The standard explanation of this is that, in general, (i)

[22]Whether we want to say that NON-VACUITY can be 'suspended' or just 'relaxed' depends on the way in which the principle is fleshed out. Either option is consistent with the position we will present. The basic point can be illustrated with a simple example. In most contexts we would interpret an utterance of *many people believe that politicians are snakes* metaphorically. But there are cases in which, based on certain cues, we can interpret it as meaning that many people really believe that politicians are, literally, snakes. This is so even if we in fact believe that this assertion is false, i.e., that it cannot possibly be the case that many people believe such a thing. It is in this sense that NON-VACUITY can be suspended or relaxed: namely, that in some contexts, we can interpret expressions in ways that, relative to our beliefs, result in highly implausible assertions. This suggests another way one might flesh out the notion of 'relaxing' NON-VACUITY, which some theorists might find more congenial: we always follow NON-VACUITY but in certain special contexts we can adopt specially unconstrained (or different, e.g., fictional discourse) beliefs sets relative to which assertions are evaluated. For our purposes, any of these notions of suspending or relaxing NON-VACUITY will do.

speakers believe that there are no stone lions_1 and plastic rabbits_1, (ii) they accept something like NON-VACUITY, and (iii) to satisfy NON-VACUITY they loosen the meaning of *lion* and *rabbit*. However, although accepting NON-VACUITY is the default stance, interpreters can and sometimes in fact do suspend or relax NON-VACUITY. They do this when they are instructed by the context—including hedges used by the speaker—to take the expressions literally, even if they believe that this would result in empty predicates. Consider the following examples:

(4.29) Something unbelievable happened at MIT, scientists discovered a way of making, literally, stone lions and rubber rabbits.

(4.30) Bio-technology is advancing at an astonishing pace. I am convinced that, in the future, we will be able to make, literally, silicon cows.

In cases like (4.29)-(4.30), the use of *literally* together with the reinforcing context have a unified effect: to instruct listeners that the head Ns should be taken literally, even if they believe that it is unlikely that the resulting expressions have a positive extension, in this or any likely future world. Part of the effect of contexts such as these, is to instruct interpreters to withhold NON-VACUITY-guided meaning-modulations. This is why (4.29)-(4.30) are interpreted as:

(4.29) ...stone lions_1 and rubber rabbits_1.

(4.30) ...silicon cows_1

As (4.29)-(4.30) illustrates, relaxing NON-VACUITY directly affects the interpretation of constitutive material NPs in the expected ways, since there is no doubt that NON-VACUITY determines the direction of the meaning-shifts of the head Ns in the default interpretations of constitutive material NPs.[23]

[23]To be sure, one could claim instead that this suggests that NON-VACUITY is not really responsible for the interpretation of constitutive material NPs. For don't (4.29) and (4.30)

At this point, we should present some new terminology and brief clarifications regarding 'literal' contexts. Let 'L-constitutive material NPs' stand for constitutive material NPs in contexts that call for strict literal interpretations. Similarly, let 'L-privative NPs' stand for privative NPs in contexts that call for strict literal interpretations. We can induce literal interpretations by general contextual cues. So, slightly reinforced, examples like (4.29)-(4.30) call for a literal reading even if we delete the tokens of *literally*. Still, the use of *literally* in these cases falls squarely in what Israel [78] argues is its orthodox usage, namely, as a metalinguistic regulator or instruction to take expressions literally, to add nothing or change nothing to the meaning of the words. According to Israel, this orthodox use of literally has four important characteristics: (i) it is used with expressions that lend themselves, in context, to multiple interpretations, (ii) at least one of the interpretations is pragmatic, (iii) the pragmatic interpretation is likely to be seen as more plausible than the literal interpretation, and (iv) the speaker is committed to the truth of the literal interpretation, even if it is remarkable or might seem highly implausible. This account of the orthodox metalinguistic use of *literally* fits perfectly with the observed patterns in the case of L-constitutive material modifiers.[24]

show that, in some sense, the extension of constitutive material NPs are not really empty, at least relative to some possible worlds? However, the crucial point is that in most contexts the relevant possible worlds are such that only the loose interpretations of constitutive material NPs have a positive and negative extension. This is all that is presupposed by Partee's account of constitutive material NPs (for an elaboration of Partee's account, see Hogeweg [75].

[24] A reviewer suggested that although, as stated, both DUAL CONTENT and SHIFTING HEADS assume that the standing or default meaning of, say, *lion* and *gun* is $lion_1$ and gun_1, this assumption might be incorrect. The idea seems to be that we might give a simpler account of privatives if we just assume that the default meaning Ns is the loose meaning, namely, $lion_2$, gun_2, etc. This suggestion seems intuitively incorrect, but more importantly, it fails to account for the compositional behavior of L-constitutive material NPs, under the plausible assumption that *literally* is a metalinguistic operator of the sort proposed by Israel [78]. These cases show that, when instructed to take things literally, speakers interpret *stone lion* and *silicon rabbit*, etc. as *stone* $lion_1$ and *silicon* $rabbit_1$, and not as *stone* $lion_2$ and *silicon* $rabbit_2$. This strongly suggest that the literal, default meaning of common Ns is not the loose meaning. This same conclusion is also supported by expressions which use *strictly speaking.* For example, *Strictly speaking, penguins are birds* and *Strictly speaking, tomatoes are fruits* are clearly true. However, *Strictly speaking, lion sculptures are lions* and *Strictly speaking, rubber toy rabbits are rabbits* seem clearly false, although they should come out as obvious truths under the assumption that the default meanings of *lion* and *rabbit* are $lion_2$ and $rabbit_2$.

Recall the key move in Partee's account: in the relevant processing respects, treat the default interpretation of privative NPs just like the interpretation of constitutive material NPs. It follows, prima facie, that the account predicts that, when NON-VACUITY is suspended or relaxed, the interpretation of constitutive material NPs and of privative NPs should be affected in analogous ways. In other words, this account predicts that, in L-privative NPs, we interpret *fake gun* as *fake* gun_1, not as *fake* gun_2.

These predictions are incorrect. In most contexts, the constituent privative NPs in L-privative NPs are interpreted in essentially the same way as the original, unmodified privative NPs, unlike the cases of L-constitutive material NPs. To see this, compare (4.29)-(4.30) with the following examples of L-privative NPs:

(4.31) Listen to this unbelievable story, some immoral toy store owner was, literally, selling fake guns at his store.

(4.32) Something amazing happened at MIT, some engineer managed to make, literally, a fake gun.

(4.31)-(4.32), just like (4.29)-(4.30), instruct interpreters to take the expressions literally; however, this does not affect the interpretation of privative NPs in the way it affects the interpretation of constitutive material NPs. In (4.31)-(4.32) the interpretations of *fake gun* are just their corresponding default interpretations, namely, *fake gun*, which SHIFTING HEADS represents as *fake* gun_2. In other words, relaxing/suspending NON-VACUITY doesn't change the default interpretation of privative NPs, at least not in the way it changes the default interpretation of constitutive material NPs, which are paradigmatic cases in which the default interpretation is partly determined by NON-VACUITY.

This result is problematic if, following SHIFTING HEADS, you hold that NON-VACUITY is responsible for determining the default interpretations of privative

NPs. But it is expected if you accept DUAL CONTENT, according to which the default interpretation of privative NPs is neither triggered nor affected by NON-VACUITY. In this respect, privative NPs behave like complex NPs involving genuine modifiers, e.g., *dark red paint* and *perfect gun*, which, in most contexts that call for literal interpretations, can only be interpreted as referring to (good examples of) dark red paint and perfect guns. In short, in contexts that call for literal interpretations, privative NPs behave more like paradigmatic cases of NPs involving genuine modifiers than like paradigmatic cases of NPs involving pragmatic meaning-shifts of the head Ns. This is exactly what is predicted by DUAL CONTENT. Hence the case of L-privative NPs strongly supports DUAL CONTENT over SHIFTING HEADS.[25]

4.6 An attempt to save SHIFTING HEADS

To defend SHIFTING HEADS, one might suggest that the lack of analogy between the interpretation of L-privative NPs and L-constitutive material NPs springs from an important difference in the status of the common-ground beliefs relevant to each case.

Here is the idea. Most speakers believe that there are no stone lions$_1$ and fake guns$_1$. The first belief is dropped from the common ground when we are instructed to take things literally, even if the resulting assertion might seem unbelievable. This explains cases like (4.29)-(4.30). By parallel reasoning, it would seem that the belief that there are no fake guns$_1$ should be dropped from the common ground in cases such as (4.31)-(4.32). However, maybe the belief

[25] This criticism of SHIFTING HEADS extends, mutatis mutandis, to other 'contextualist' accounts of privative NPs. For example, one might propose replacing NON-VACUITY with other more traditional Gricean maxims of interpretation, such as BE TRUTHFUL. However, essentially the same criticism applies to Gricean reconstructions of SHIFTING HEADS, since they still depend on the idea that the default interpretation of privative NPs involves a pragmatically induced meaning shift of the head Ns. This assumption is precisely what is in conflict with the behavior of L-privative NPs, especially when compared with the behavior of L-constitutive material NPs.

that there are no fake guns$_1$ is a more entrenched belief than the belief that there are no stone lions$_1$. There are possible worlds in which there are stone lions$_1$; but there are no possible worlds in which there are fake guns$_1$. We can't believe that there are fake guns$_1$ in something like the sense that we can't believe that there are married bachelors. So whatever we do to make sense of expressions such as (4.31)-(4.32), what we cannot do is drop from the common ground the belief that there are no fake guns$_1$. This explains why expressions like (4.31)-(4.32) feel genuinely redundant: the contextual instruction to take the privative NP literally does not change its meaning relative to the default cases.

However, this response is based on a fundamental mistake: interpreted in the way suggested by SHIFTING HEADS, *fake gun*$_1$ is not an empty predicate, nor is there reason to think that mature speakers believe that it is.

To evaluate this response, we need to consider possible lexical entries for $fake$. Partee herself does not pronounce on this, but others who accept her account of privatives assume that they are first order predicates.[26] What could the meaning of $fake$, taken as a first-order predicate, be? Initially, lets assume a simple atomic account:

(4.33) $[\![fake]\!] = \lambda x.\ \textsc{fake}(x)$

The point of (4.33) is just to capture the idea, which seems correct, that when we use $fake$ predicatively, as in *everything that they sell at that store is a fake*, we mean something like a fake thing of some sort. As we said before, DUAL CONTENT is compatible with the fact that we sometimes use $fake$ in this way; but unlike the contextualist accounts, it does not assume that this is how we use it in the default interpretation of privative NPs. This proposed response in defense of SHIFTING HEADS assumes that when $fake$ is interpreted roughly as

[26] For example, this is how Recanati [164]: ch. 2 approvingly interprets Partee's account of privative NPs.

(4.33) the intersection between $fake$ and gun_1 is empty, or at least that most mature speakers strongly believe this, and are unwilling to drop this belief from the common ground.

Now, here's the crucial question: is it really the case that, so interpreted, there are no fake guns$_1$? In other words, is it really the case that no guns$_1$ are also fake things? Consider the following scenarios:

(EVIL PARK) Mark is an evil paintball gun park owner. He really hates John. Unaware of this secret hate, John goes to Mark's park to play. Mark comes up with the plan of secretly replacing John's paintball gun with a fake paintball gun that is actually a real gun, with the intention that John kill someone. Fortunately, when John handles the fake paintball gun he notices something suspiciously off with its weight and refuses to use it.

(EVIL PARTY GOER) Peter is a clever terrorist who wants to cause havoc at a halloween party. To evade the security measures, he designs a gun that is intended to look like a toy gun but is actually a very powerful gun. Fortunately the security personnel are suspicious of Peter and his fake toy gun, so he gets caught and his plot is averted.

EVIL PARK presents an object that we appropriately called a *fake paintball gun*, since it is an object that was made to look like a paintball gun and feel like a paintball gun but is not really one, since it was made for and can be used to kill by shooting. Similarly, EVIL PARTY presents an object that we appropriately called a *fake toy gun*, since it is an object that was made to look like a toy gun but is not really a toy gun, since it was made for and can be used to kill by shooting. So these are examples of guns$_1$ that are also fake objects. Furthermore, these scenarios are not far fetched. Indeed, they are closer to our world than extraordinary scenarios in which MIT scientists create real animals from non-organic materials. If under instruction to take things literally, we have no hesitation to widen the common ground so that it includes worlds in which there are stone lions_1 and rubber rabbits_1, then under the same instruction, we should have even less hesitation to widen the common ground so that it includes worlds, such as our own, in which there are guns_1 which are also fake things (of some sort).

But this prediction is incorrect: we do not interpret L-privative NPs in the literal sense (e.g., as *fake* $\textit{gun}_1$ taken in the sense it would have according to this reconstruction of SHIFTING HEADS) in which we interpret L-constitutive material NPs (e.g., as *stone* $\textit{lion}_1$). In other words, these accounts fail to explain why in L-privative NPs we never use, say, *fake gun* to denote things like the subset of fake toy guns or fake paintball guns that are real guns, even in scenarios that call for this reading (such as cases like (4.31)-(4.32)).[27] This, as we just saw, should be quite easy to do if $fake$ when used attributively were really a first-order predicate. Furthermore, assuming that $fake$ is a gappy or context-sensitive first-order predicate won't solve this problem:

(4.34) $[\![fake]\!] = \lambda x.\ \text{FAKE-AS-}\mathbf{p}(x)$, where the value of $\mathbf{p}$ has to be provided by the context.

In some cases, gappy expressions use the immediate linguistic context to determine the meaning of the open parameter. For example, *small elephant* usually means an elephant that is small for an elephant, but as is well known, in some contexts, *small elephant* can mean an elephant that is of a normal size for an elephant but is small compared with other kinds of giant creatures. If $fake$, when used attributively, was a gappy first-order predicate, then when combined with *gun* it would usually mean fake for a gun_1; but in certain contexts, such

[27] Note that this argument does *not* assume that x's being a fake paintball gun entails that it is a fake gun. What it assumes is that *some* fake paintball guns could be real guns. The reason is simple: a fake paintball gun is something that was made to look like a paintball gun but is not really a paintball gun. As in EVIL PARK, some objects could be deceivingly be made to look like paintball guns but be in fact real guns.

as some of the ones mentioned above, it could also mean fake for a toy gun, etc., even when combined with gun_1. So we should get the reading, which we do not, according to which *fake gun* means fake toy gun that is a real gun, fake paintball gun that is a real gun, and so on.[28]

4.7 DUAL CONTENT: Objections and clarifications

To conclude this discussion of DUAL CONTENT, I want to make some clarifications and address some important objections. We first discuss the relation between DUAL CONTENT and contextualism, and then the notions of E-structure, C-structure, and their relation.

4.7.1 Is DUAL CONTENT compatible with Contextualism?

Contemporary formal semantics has moved in the direction of contextualism. To mention a few examples, there are influential accounts of quantifiers (Szabo and Stanley [195]), modals (Kratzer [93]), predicative and attributive adjectives (Rothschild and Segal [174], Kennedy and McNally [89], Recanati [164]), and noun-noun compounds (Weiskopf [208]) that posit context-sensitive parameters

[28] To close this discussion of SHIFTING HEADS, I should add two brief but important points. First, I take this criticism of SHIFTING HEADS to undermine one way of trying to save lexical atomism. Now, although SHIFTING HEADS is one of the most plausible and influential accounts, a full reply to lexical atomism would require that we consider other options. This latter task is beyond the scope of the present work. Second, there is good reason to think that debates of theories of modification could soon use direct psycholinguistic and neurolinguistic techniques. This is because theories such as DUAL CONTENT and SHIFTING HEADS can be distinguished by the way they classify, in terms of their coercion patterns, privative NPs and constitutive material NPs (and other subclasses). According to DUAL CONTENT, the default interpretations of expressions like *fake gun* do not involve coercion, but those for *stone lion* and *fake lion* do (in the latter case, because a TELIC has to be artificially provided). According to SHIFTING HEADS, the default interpretations of *fake gun*, *fake lion*, and *stone lion* all involve coercion. Recent efforts to find the neural correlates and signatures of operations of coerced meaning modulation have been quite successful, e.g., Pylkkanen and McElree [155] and Pylkkanen [154]. This justifies some optimism that these accounts could soon be evaluated using direct neurolinguistic data.

in logical form. There are also various arguments for the view that the meaning of expressions in general, and not only those that have context-sensitive parameters, is usually modulated during online processing (Carston [21], Recanati [160, 164]). To accommodate these account, recent discussions of compositionality suggest that, to model our linguistic competence, we should think of compositionality as the principle that the meaning of token expressions is determined from the contextually saturated and modulated meaning of their constituents (Pagin [122], Szabo [192], Recanati [164]).

It is important to clarify the place of DUAL CONTENT in this general debate, especially since we criticized the main 'contextualist' account of privative NPs. To do so, it is useful to introduce some new terminology, essentially following Recanati [164]. First, since lexical items can have free parameters, we will say that, for any expression e:

- $[\![e]\!]_c = f_e(c)$

where f_e is the character of e and $f_e(c)$ is the occasion meaning of e in context c. If e has no free parameters, then for all c's, $f_e(c) = m$, where m is the standing meaning of e. Secondly, since the meaning of tokens is determined from the meanings of the modulated (instead of the standing) meanings of their immediate constituents, we can represent the modulated meanings of an expression e, $[\![e]\!]_{M,c}$, as follows:

- $[\![e]\!]_{M,c} = mod(e,c)([\![e]\!]_c)$

mod takes as an argument an expression e and context c in which e occurs and returns as value the modulation function $f_{M,e}$, which takes $[\![e]\!]_c$ and returns the meaning that is salient/relevant/appropriate for e in c.[29]

[29] Different contextualist accounts can further restrict *mod* in different ways. That some such restriction is required seems to me uncontroversial; otherwise this sort of account will over-predict 'reasonable' meaning-shifts that are actually not available. The over-generation problems faced by unrestricted contextualist accounts is illustrated with several interesting

For simplicity, assume that traditional FA is the main compositional operation. We can incorporate these two forms of context-sensitivity by reformulating FA as follows (assume that all non-branching nodes are terminal nodes):

(TN) If α is a terminal node, then $[\![\alpha]\!]_{M,c} = mod(\alpha, c)([\![\alpha]\!]_c)$, where $[\![\alpha]\!]_c$ is specified in the lexicon.

(FA) If α is a branching node, $\{\beta, \gamma\}$ is the set of α's daughters, and $[\![\beta]\!]_{M,c_1}$ is a function whose domain contains $[\![\gamma]\!]_{M,c_2}$, then $[\![\alpha]\!]_{M,c} = [\![\beta]\!]_{M,c_1}([\![\gamma]\!]_{M,c_2})$

According to this implementation, *mod* does not operate on the outputs of FA, but only on terminal nodes/lexical items. This allows a constrained form of meaning modulation.[30]

DUAL CONTENT is compatible with this general contextualist framework. The revised version of FA we introduced in §4.3—FA^{DC}—can be easily reformulated to allow for these two forms of context-sensitivity. But what is important to note, for our purposes, is that unlike other accounts of privative NPs, DUAL CONTENT does not assume that either privative Adjs or common Ns have context-sensitive free parameters. It also does not assume that, when tokens of privative NPs are interpreted in the default way, the standing and modulated meanings of the token Adjs and Ns differ significantly. In this sense, according to DUAL CONTENT, the default interpretation of privative NPs is not context-sensitive, or rather, does not involve significant meaning modulations of the head Ns.

examples in Asher [1] and, focusing on color words, in Kennedy and McNally [89]. For our purposes, we can ignore further details of how *mod* should be constrained.

[30] There are at least two ways of implementing the context-sensitive *mod* function to get compositional rules that determine the meanings of complex expressions in terms of the modulated meanings of their constituents. On the unconstrained implementation, *mod* is generalized to apply at every level of interpretation. On the constrained implementation—which is the one we adopt here—*mod* applies only to lexical items, i.e., terminal nodes. Recanati [164] and Pagin and Pelletier [122] favor the more unconstrained version. For our purposes, this choice is open, but for detailed reasons to favor the constrained implementation, see chapter 2 and chapter 5.

That does *not* mean, however, that there is *never* any meaning modulation involved in the interpretation of privative and other NPs involving modifiers that are treated as semantic restructuring operators. Consider the following example:

(BAD HUSBAND) It was recently revealed that Smith, a supposedly ideal husband, had an illicit love affair. Ann, a famous town gossip, tells Sam, 'this does not surprise me at all. Smith is a typical man'.

To properly understand Ann's claim that Smith is a *typical man* we have to modulate the C-structure of *man*—specifically the TELIC—to include cheating as one of the things that a man does. For most of us, this is a sufficiently salient cultural stereotype and the required modulation is 'easy'. Once the C-structure of *man* is enriched in this way, we can continue to treat *typical* as a semantic restructuring operator, along the lines of (4.17) above. This, in turn, would explain why in BAD HUSBAND Ann is asserting, about Smith, that he is a cheat. Examples like this are common. What is important to note, however, is that allowing the general possibility of meaning modulation does not affect the basic idea of DUAL CONTENT, namely, that some Adjs partly operate on the C-structure of Ns.

Indeed, any theory which admits something like C-structure, such as DUAL CONTENT, should be made compatible with some form of contextualism (for further discussion, see Chapter 5). As is well known, there is plenty of evidence showing that the prototypes associated with terms are modulated as a function of features such as preceding discourse and task demands (Roth and Shoben [172], Barsalou [7]). For example, when *animal* is processed in the context of *milk*, there is a prototype effect (e.g., ease of categorization) for *cow* and *goat* relative to *horse* and *mule*, but when *animal* is processed in the context of *riding* the effect goes in the other direction, favoring *horse* and *mule* [172].[31] This sort of finding has been so often replicated, that it is fair to say that the relative instability of lexical prototypes is by now a well established tenet in cognitive psychology Murphy [118]. Insofar as the C-structure posited by DUAL CONTENT incorporates something like enriched prototypes, we should also expect that it undergoes some modulation across contexts. These considerations should be kept in mind when judging the descriptive adequacy of proposed lexical entries for privative and subsective Adjs and common Ns.

4.7.2 Why not include functional and genealogical information in the E-structure of lexical artifact terms?

Consider the following objection to DUAL CONTENT. As we argued in §4.2, Putnam and Kripke gave us good reasons to reject accounts of natural kinds that assume that their extension is determined by associated descriptions, including definitions and stereotype information. So if we need something like enriched stereotype information to fully account for our linguistic competence, as we argued above, it follows that we need to make a distinction between two components of meaning, E-structure and C-structure. However, one could argue that the reasons to reject the idea that definitions determine the extension of lexical terms do not apply to the case of artifacts. Indeed, some linguists and philosophers have argued that, in general, artifact kinds can be defined in terms of their functional and/or genealogical properties (Thomasson [198]). If this is correct, it might seem reasonable to incorporate the TELIC or AGENTIVE of artifact terms into their E-structure. In turn, this would mean that we might

[31] Importantly, the effects of context on prototypes can be relatively permanent. For example, in a classic experiment by Barclay et al. [4], one group of subjects was asked to encode sentences such as *the man lifted the piano*, and another group was asked to encode *the man tuned the piano*. On a subsequent test, *something heavy* was a more effective cue for recalling *piano* than *something with a nice sound* for the first group of subjects, and the opposite pattern was observed for the second group.

be able to account for privative NPs without assuming that common Ns have a dual semantic content which includes a non-extension determining C-structure.

This is an interesting objection, but it is not hard to see that lexical artifact terms cannot, in general, be defined in the way proposed. For example, we know that there are malfunctioning guns and antique or display guns which cannot shoot. There are computers which have ceased to work and cars which do not even have an engine. Some erasers do not erase well, and some dinning tables are used as computer desks but do not thereby change their identity. Most chairs are used for sitting but not all items used for sitting are chairs, just as not all things that can be used to shoot are guns. In all these cases and countless more, we distinguish between artifacts and their typical function, so the TELIC cannot in general be part of E-structure. These intuitions are supported by some experimental results (Keleman [86]). In one representative example, subjects were shown teapots used to water plants, and lampshades used as umbrellas. Most subjects still choose *not* to name these objects based on their current use (Rips [166], [109]).

We can also easily distinguish between artifacts and the function for which they were created. Suppose that, in some remarkable situation, you witness the spontaneous emergence of a gun or a pencil, you handle both objects and they have their usual functions, to shoot and write. It is appropriate to call these objects *gun* and *pencil*, despite their lacking the typical AGENTIVE of guns and pencils. Indeed, talking about the spontaneous generation of artifacts does not seem to involve any kind of incoherence, unlike talking about, say, unperceived sensations. Furthermore, not all artifacts that are made for shooting are guns, just as not all artifacts that are made for sitting are chairs. That the typical AGENTIVE of simple artifact terms is not part of E-structure is also suggested by experimental studies. For example, Chaigneau [22] conducted an experiment in which subjects were given information about the intention for which an object was created. Subjects were also given information about the object's physical structure and current use. In some scenarios, the creator's intention and the actual use of the object differed. For many kinds of artifacts, most subjects choose to individuate them based on actual use. For example, objects created for a non-mopping purpose but which were successfully used for mopping, were categorized by most subjects on the basis of their actual, as opposed to the creator's intended use.

Now, to say that the TELIC and AGENTIVE are not components of the E-structure of lexical artifact terms is not to say that they are not part of the E-structure of *complex* artifact terms in certain constructions. That they are inherited into the E-structure of certain complex terms such as *perfect gun* and *counterfeit Rolex* is precisely the point of DUAL CONTENT.[32] In addition, we are not denying that in most cases the typical TELIC and AGENTIVE associated with lexical artifact terms plays a crucial role in categorization. Indeed, some prominent developmental psychologists argue that, in the case of artifacts, beliefs about the creator's functional intention is the most important categorization criterion (Bloom [10, 11], Keleman [86]). Other cognitive psychologists argue that the most important feature is instead the TELIC (Chaigneau and Barsalou [**?**]). Perhaps most persuasively, Malt and Sloman [105, 179] propose, partly on the basis of an extensive evaluation of the literature, that which features—amongst

[32] A reviewer notes that if we concede that, say, guns can be created ex-nihilo, then why not say the same thing about *fake guns*, and so deny that AGENTIVE information is represented in its E-structure. In response, note, first, that the point is not obviously generalizable. For example, assume that, as we argued, *gun* does not include TELIC information in its E-structure; it clearly does not follow that the same holds for *perfect gun*—as we said above, intuitively, a perfect gun is one that, among other things, shoots perfectly. Likewise, we can agree that a tiger might not have the usual stripped coat, but *typical tiger* arguably includes that information in its E-structure. According to DUAL CONTENT, certain complex expressions bring into their E-structure information that was only part of the C-structure of the constituents. So a testable prediction of this view is that, in the case of privatives such as *fake* and *counterfeit*, speakers will exhibit stronger and less flexible intuitions about the AGENTIVE requirements for something to qualify as a fake or counterfeit artifact than they would for artifacts in general.

form, function and genealogy—are most important for artifact categorization depends on aspects of the relevant context, including task demands. These proposals are compatible with DUAL CONTENT as long as they accept that features of lexical terms that are important for categorization are, in most cases, only part of C-structure, as most cognitive psychologists would accept in any case.[33]

4.7.3 What then is E-structure? Why not just drop that notion altogether?

At this point, one might wonder why we should maintain that lexical artifact terms have an E-structure at all. The reason for this, however, is basically the same as for why we should maintain the distinction between E-structure and associated descriptions in the case of natural kind terms. The standard arguments are well known, so I will only briefly rehearse them to show that they also apply to artifact terms. The aim here is only to show that the case for assuming an E-structure independent of associated beliefs is at least as strong for artifacts as it is for natural kind terms.[34]

The main reason for rejecting descriptive theories and keeping a directly referential component stem from 'the problem of ignorance' and 'the problem of error' [17, 91, 95, 152, 209]. The problem of ignorance is that speakers can refer to a kind even if they do not know any descriptions which uniquely pick out the

[33] To properly account for categorization patterns, the representations of the C-structure of lexical terms have to be supplanted with weights on each quale, reflecting their relative importance. This is compatible with DUAL CONTENT, but it would have introduced technical complications which are orthogonal to the main aim of this paper. We ignored information about the relative weights of the dimensions of C-structure because it does not seem to play a key compositional role in privative NPs. For example, to capture the difference between *fake*, *counterfeit*, and *artificial*, we had to refer, in their E-structure, to qualia functions that access directly different dimensions of the C-structure of the N they modify (and there was no need to further operate on their weights). One might think that the compositional behavior of privatives can be captured by operating just on the weights of the modified Ns (e.g., by inverting them), without direct reference to their quale. However, it seems impossible to capture the difference between e.g. *fake*, *counterfeit*, and *artificial* by operating just on the weights of the modified Ns.

[34] For a more detailed response along these lines, to which the following discussion is indebted, see Kornblith [92].

kind. The problem of error is that speakers can refer to a kind even if most of their beliefs about the kind are incorrect. We can talk about elms and beeches even if we do not know much about their differences, and we can discuss the basic properties of molybdenum even if we are radically mistaken about what they are. Reference to artifactual kinds exhibits the same patterns (Kornblith [92]). We can talk about amulets and nuclear reactors even if we only have a vague idea of their form and function, one which might not distinguish them from similar artifacts. We can talk about rheostats and spandrels even if we are ignorant of or entirely mistaken about their basic function. Just like in the case of natural kind terms, we may acquire these terms from other speakers and use them with the intention to co-refer. This division of linguistic labor explains why the ability to refer is compatible with such ignorance and error. In our terms, this is just to say that we need a notion of E-structure, independent of their C-structure, to account for our referential competence with lexical artifact terms.[35]

Now, this story faces the same challenge faced in the case of natural kind terms, namely, that we do not yet have an accepted account of what ultimately determines the extension of the kinds. Combined with the idea that descriptions or C-structure do not in general determine reference, this is part of the reason why we took the E-structure of common Ns as semantically atomic, and their relation to their extension as primitive. We can still ask, of course, how the

[35] One could insist that there is an important difference between natural kind and artifact terms which might bear on whether we need to posit something like E-structure in the latter case. Plausibly, the rationale for preserving an E-structure for natural kinds presupposes that they have an extension which is independent of our current beliefs. The parallel presupposition might not seem plausible in the case of artifacts. As one anonymous reviewer puts it, 'we may lack definitions for artifacts, but it seems that it cannot be that most of our beliefs about a given artifact are mistaken'. However, I would dispute that there is a relevant distinction in this respect between artifact and natural kinds. Consider, for example, historical artifacts. Archeologists might discover some objects and classify them as antique grain storages, only to then discover that they are underground temples. This situation is in fact quite common: we discover an artifact, refer to the class via samples, are massively confused about their TELIC and AGENTIVE, and try to discover new facts to improve and perhaps radically change our current beliefs about the kind.

extension of terms is determined. As we mentioned above, some authors suggest that, for artifacts, it might be something like the intentions of their creators or producers. But, again, it is crucial to note that whatever story we give along these lines about the role of intended function in reference determination, it is going to be much more complicated and abstract than the information most speakers actually include in the AGENTIVE of artifact terms, which is usually just the *typical* reason or way in which the artifact came into being. So even if something like the producers' intentions fixes the correctness conditions of artifact kinds, this is no reason for suggesting that, for simple lexical artifactual terms, we should include the AGENTIVE or TELIC as part of E-structure. This is so even in cases where radical ignorance and error are rare, as in the case of common artifacts such as guns and pencils.[36]

[36]For example, [198, 199] argues that the essence of artifacts lies in the intentions of their makers:

> Necessarily, for all x and all artifactual kinds K, x is a K, only if x is the product of a largely successful intention that Kx, where one intends Kx only if one has a substantive concept of the nature of Ks that largely matches that of some group of prior makers of Ks (if there are any) and intends to realize that concept by imposing K-relevant features on the object (Thomasson [198]).

A consequence of this constitutive condition is that 'reference to artifactual kinds ... cannot proceed without someone (namely, those responsible for the production and reproduction of these artifacts) having a substantive concept of the nature of the kind', where the notion of the 'nature of the kind' is fleshed out in terms of its intended function (Thomasson [199]).

We might grant that there is some essential or strong connection between producer's or conceiver's intentions and the extensions of artifact terms, but it seems clear that Thomasson's proposal is too strong. To illustrate, note that it entails, counter intuitively, that in the following scenarios the objects we are talking about are not 'really' guns.

(COPY) Little Mark is a precocious infant with remarkable craftsmanship skills. He finds some very detailed instructions to make an object he has never seen before, which might have some function that is unspecified in the instructions. He somehow gets the materials and follows the instructions perfectly. His intention is only to execute the instructions; he does not really care or even wonder much about what the function of the object might be. Little Mark's father then enters the room only to find, much to his dismay, that Mark constructed a gun.

(ACCIDENT) Strange accidents and remarkable things happen in the magical world of HH. One day, the toy factory production line somehow popped a series of guns. After a series of unusual accidents, everyone in HH was frightened and demanded an investigation. After arduous work by the top detectives of HH, it was discovered that no one intended that this be so. It was merely a very strange accident due to some unintended glitch in the production line.

In COPY and ACCIDENT we can talk about the relevant objects as being real guns, despite their failing to satisfy Thomasson's condition, since their proximal creators did not intend that they be objects of the gun kind. One might reformulate Thomasson's story to deal with these sorts

4.7.4 What then is the relation between E-structure and C-structure?

To talk about the relation between E-structure and C-structure, we should consider the extension of each component, were we can take the C-structure of a term as a complex first order predicate that is the conjunction of all its quale values.[37] There are five ways in which the extension of the E-structure (call this set 'E') and that of the C-structure ('C') of artifact and natural kind lexical terms could be related (cf. Weiskopf [209]):

(I) $E = C$

(II) $E \subset C$

(III) $C \subset E$

(IV) $[E \cap C \neq \emptyset] \wedge \neg[(E \subset C) \vee (C \subset E)]$

(V) $E \cap C = \emptyset$

The theories that first introduced a distinction between cognitive and referential content assumed that the former fully determines the latter. These theories conform to case (I). Prominent examples are the Fregean theory of sense and the Millian theory of names. Recent dual content theories, especially in cognitive psychology, tend to hold that cognitive content does not in general determine referential content (Murphy [118], Prinz [142,144], Weiskopf [209]). DUAL CONTENT belongs to this class. For many artifact terms, the set of objects that satisfies their C-structure is usually a subset of the set of objects that satisfies

of cases. But the only point I am pressing is that these reference fixing conditions are quite complicated and independent of speakers' beliefs. In this sense, they are different from the information typically incorporated in the AGENTIVE of the C-structure of artifact terms.

[37] We can determine the extension of C-structure in other ways, e.g., by considering disjunctions involving each quale, or by combining each quale with weights and appealing to a threshold function. The points I make below also apply to these alternative ways of determining the extension of the C-structure of terms.

the E-structure (case III). For example, the C-structure assigned to *gun* in (4.8) above picks out guns that look and function like typical guns, but, again, some guns are not typical in this way. C-structure can also sometimes overextend (case II). For example, many speakers might have a C-structure for *couch* that extends to sofas. In short, for most artifacts we can only hold that C and E-structure have some extensional overlap (case IV). However, we saw above that there are important exceptions, including cases in which we are entirely wrong about the properties of a particular artifact term (case V). Following Weiskopf [209], we can call terms whose C and E-structure satisfy (II), (III) or (IV), 'adequate'. 'Fully adequate' terms satisfy (I); and terms that only satisfy (V) are 'inadequate'.

For most artifact and natural kind terms, then, the relation between their E-structure and C-structure is relatively loose. This does not mean, however, that the two components are independent. There are important constrains that govern their relation. We cannot discuss this issue in detail here, but I will mention two plausible candidates that should be further explored: these are constraints regarding 'accurate' and 'efficient' representation (Weiskopf [209]). 'Accurate representation' captures the idea that one aim of our conceptual system is to achieve full adequacy of C-structures with respect to E-structures: i.e., we aim to acquire beliefs about the properties of kinds which are accurate. 'Efficient representation' captures the idea that this has to be done under certain limitations of time, information, working memory and processing power. The interplay between these two constraints means that C-structure is often allowed certain degrees of inaccuracy, especially in cases when the cost of the inaccuracies is low and the cost of achieving full adequacy is high.

Having emphasized the distinction and sometimes loose relations between E-structure and C-structure, it is important to recall that the reason why C-structure is part of the *linguistic* meaning of lexical items is that, if DUAL CONTENT is correct, it plays a crucial compositional role in certain common constructions, including privative NPs such as *fake gun* and *counterfeit document* and subsective modifications such as *powerful gun*, *perfect knife*, and *typical chair*. In all these cases, aspects of the C-structure of the constituents enter into the E-structure of the resulting complex NPs via purely compositional operations. It is because of this combinatorial dynamic that DUAL CONTENT can claim that C-structure is part of linguistic meaning.

This highlights an important contrast between DUAL CONTENT and its predecessors: most previous theories do not show in detail how the C-structure of lexical terms can affect the E-structure of complex terms of which they can be constituents. Indeed, dual content theorists tend to assume that the composition of E-structure and C-structure proceeds along parallel, but completely independent routes (see e.g. Prinz [142] and Weiskopf [209]):

> When concepts combine, they contribute both their cognitive and referential contents... The familiar semantic rules of composition in each case produce, in parallel, complex contents from the contents of the constituents. R-contents of concepts combine to form R-contents of judgments, and similarly for C-contents, although neither C-contents nor R-contents determine the other (Weiskopf [209]: 145).[38]

Weiskopf's use of 'referential' and 'cognitive' content corresponds, roughly, to our E-structure and C-structure. From the perspective of DUAL CONTENT, holding to the view of parallel and independent composition of E-structure and C-structure blocks one of the most useful tools available to dual content theories, namely, semantic restructuring operators and the corresponding dynamic view

[38] I should note that, aside from this claim about compositionality, there is much that I agree with in Weiskopf [209]'s excellent defense of dual content theories.

of compositionality. If DUAL CONTENT is correct, certain constructions use the C-structure of constituents to determine the E-structure of complexes. It follows that we cannot dismiss the notion of C-structure on the grounds that, although it might be needed to account for certain psychological facts such as typicality patterns, it is not needed to account for purely linguistic facts.

4.8 Conclusion

According to DUAL CONTENT, common Ns have internal semantic structure, which consists of E-structure and C-structure. We showed how we can use this structure to provide a satisfactory compositional account of privative NPs, a traditionally problematic class of constructions, as well as for other interesting subsective and intersective NPs. DUAL CONTENT is not the *only* plausible account of privative NPs. As we saw, Partee's SHIFTING HEADS is compatible with atomic lexical semantic theories. Theorists keen on rejecting the view that lexical terms have internal semantic structure, or that the compositional operations have access to this structure, might be tempted to adopt SHIFTING HEADS. We showed, however, that SHIFTING HEADS and similar contextualist accounts are fundamentally mistaken. Still, given the potential variety of accounts of privative NPs compatible with lexical atomism, the crucial question is this: Why insist on constructing accounts of complex NPs that don't require that Ns have a complex semantic structure to which the compositional operations are sensitive?

If the representations and operations introduced by DUAL CONTENT could only be used to deal with privative NPs, it might be wise to try to explain these NPs in a more traditional way. But this is clearly not the case. On the one hand, DUAL CONTENT employs representations of the meaning of common Ns which are very close to those used to represent lexical concepts in recent influential psychological theories. These representations play a crucial role in accounts of language acquisition and fast and slow categorization. On the other hand, powerful accounts in the family of DUAL CONTENT have been presented to deal with many types of problematic complex expressions (Bouillon [18], Jensen and Vikner [207], Leslie [102], Pustejovsky [149, 151], McNally [112], Pinker [141], van Tiel [205], Sasson [175], Knobe et al. [90]). All in all, I think it is time to expand our basic formal semantic framework to include compositional operations which are sensitive to the internal semantic structure of lexical terms, including their non-extension determining C-structure. In this way, we will begin to integrate compositional formal semantics with a psychologically motivated and realistic lexical semantics.

Chapter 5

Making Prototypes Strictly Compositional

The aim of this chapter is to reconcile two claims that have long been thought to be incompatible: (i) that we compositionally determine the meaning of complex expressions from the meaning of their parts, and (ii) that prototypes are components of the meaning of lexical terms such as *fish*, *red*, and *gun*. Hypotheses (i) and (ii) are independently plausible, but most researchers think that reconciling them is a difficult, if not hopeless task. In particular, most linguists and philosophers agree that (i) is not negotiable; so they tend to reject (ii). Recently, there have been some serious attempts to reconcile these claims (see e.g., Prinz [142,144], Hampton and Jonsson [68,84], and Schurz [178]), but they have two crucial shortcomings: they propose an implausibly weak account of compositionality and they assume a problematic externalist notion of prototype individuation. In contrast, I argue that we can reconcile (i) and (ii) if we adopt, instead, an *internalist* conception of prototypes and a context-sensitive but *strong* notion of compositionality. I also show that each of these proposals is more plausible than the competing previous suggestions.

5.1 Introduction

Theories of lexical concepts have to explain two of our most distinctive cognitive capacities. The first is categorization: we can classify objects and events in such a way that we can make useful inferences about their properties. Scorpions are dangerous. So it is important that, upon facing one, we quickly infer that we should not cuddle it. The second is compositionality: we can combine words into complex phrases to express an unbounded number of novel thoughts. Smith can tell John that *Hegel is a better philosopher than Hume*, and John can understand this sentence even if it expresses an utterly bizarre and completely novel thought.

One of the most important theories of lexical concepts is the 'prototype theory'. Roughly, prototypes are structured representations of weighted dimensions and features typically possessed by objects in the category they represent. For example, the prototype for APPLE includes dimensions such as color, shape and taste, with features for those dimensions such as +RED, +ROUND and +SWEET. The features represented by prototypes are usually typical, diagnostic, or some function of both, relative to their associated category.[1] Prototypes were originally—and most agree, successfully—introduced to explain a set of results about lexical concepts and categorization known as 'typicality effects' (Margolis [106], Murphy [118], Machery [104]):

- When categorizing members of a class, subjects can easily order them according to their degree of typicality (e.g., a sparrow is judged a more typical bird than a penguin).

[1] A feature x is typical, relative to category A, if the probability is high that an entity has x if it belongs to A, and x is diagnostic if the probability is high that an entity belongs to A if it has x.

- Subjects know that most of the features used to represent a category are not necessarily possessed by all members (e.g., there can be male lions that do not have manes).

- Typical members of a class are categorized faster and more reliably than borderline or less typical members (e.g., a cow is judged to be a mammal quicker and more reliably than a Beluga whale)

Most typicality effects were obtained or subsequently replicated in experiments which *only* used *linguistic* stimuli.[2] For these reasons, many researchers propose that lexical concepts either are or have prototypes as essential components.

The problem with this suggestion, however, is that prototypes do not seem to be compositional (Fodor [44,47], Murphy [118], Connolly et al [30,55]). Compositionality says (roughly) that the meaning of a complex expression should be completely determined from the meaning of its parts. This condition seems to be generally violated by the prototypes associated with complex expressions: e.g., the feature LIVE IN TANKS is arguably part of the prototype of *pet fish*, but it is not part of the prototypes of either *pet* or *fish*. One might dispute this example, but the 'emergent feature' problem it illustrates is general. Several experiments suggest that the processes which determine the prototypes of complex expressions regularly add or subtract features in ways that are clearly not compositional (Kunda et al. [96], Johnson and Keil [80], Murphy [118]). If correct, this means that we cannot take prototypes as conceptual components.

Prototype theorists have directly addressed this important objection (Hampton and Jonsson [68, 84], Prinz [144]). Their basic idea is that we can abandon the traditional notion of compositionality and still explain the productivity of

[2]For example, if subjects have to decide whether something is a bird, they are faster and more accurate for sparrows than penguins regardless of whether they are shown pictures or words for sparrows and penguins (see Murphy [118]: Ch. 11). This stimulus independence challenges those who deny that a theory of lexical concepts should do double duty as a theory of linguistic meaning. The same data that was used to argue that concepts have a prototype structure can be used to argue that linguistic meaning has a prototype structure.

language and thought. In place of strict compositionality, they propose a non-compositional constraint that allows for general beliefs to enter into the processes which determines the meaning of phrases. From this perspective, LIVE IN TANKS is part of the prototype of *pet fish* because most subjects believe this from previous encounters with pet fish and they use this belief to determine the meaning of the phase. We will examine this non-compositional proposal in detail in §5.4. I will argue that the proposal is too unconstrained to serve any useful role in modeling language processing.

In contrast to previous accounts, I will argue that prototypes are strictly compositional. The argument is based on two independently plausible but mutually reinforcing claims, one about compositionality and the other about lexical prototypes.

- Compositionality allows some degree of context-sensitivity. The effect of context-sensitivity is constrained by information available in lexical entries. Once some subset of information is selected to represent the lexical entry in a language processing task, there is no additional effect of context. In particular, general beliefs do not affect the combinatorial operations of language.

This account follows recent contextualist proposals in Philosophy of Language, but implements them in a way that is appropriately constrained (§5.5). On this view, the processes of phrasal meaning-determination are more computationally tractable than they would be if we assume the non-compositional accounts. I will argue that, compare to weaker accounts, this implementation is supported by the available empirical data.[3] The second claim I defend is the following:

[3]For readers who are more familiar with issues about compositionality and context-sentivity as they arise in recent discussions in Philosophy of Language, I should point out that most contextualist positions, such as Recanati's Truth-conditional pragmatics and Carston's Relevance theory, defend versions of compositionality that are *substantially weaker* than the principle I will defend. The non-compositional views defended in Philosophy of Mind by prototype

- Lexical prototypes are constituted by features which often encode statistical central tendencies of their corresponding categories (e.g., typicality). However, this is not an essential property of prototypes. Furthermore, prototypes encode information about the perpetual characteristics of categories, but they also encode other kinds of information, e.g., the typical function and behavior of certain classes of objects.

I think current prototype theories would, for the most part, agree with this view of lexical prototypes. The problem, as we shall see, is that *in discussions of prototypes and compositionality*, theorist tend to both overemphasize the degree to which prototypes encode central tendencies and to under-appreciate the diversity of they type of information that is encoded by prototypes (§5.6). Once these revised accounts of compositionality and prototype individuation are in place, we can address the infamous 'emergent features' objection against prototype theories of lexical concepts (§5.7-§5.8).

5.2 Compositional prototype combinatorics

We begin this discussion by clarifying two aspects of the notion of compositionality: (i) why was compositionality originally thought to be a property of our linguistic competence? and (ii) how does it constrain *prototype* models of phrasal interpretation?

Compositionality, from the perspective adopted here, is a property of the Faculty of Language (FL)—a specialized mental faculty which plays a central role in the acquisition and processing of natural languages. There is much controversy regarding the basic properties of FL, but what is not in doubt is that it is productive and systematic. FL is productive in the sense that it can

theorists such as Prinz and Hampton belong to the same family as the contextualist views in the Philosophy of Language.

assign meanings to indefinitely many novel expressions. FL is systematic in the sense that its productivity is structured such that if it can assign meanings to a set of expressions D, it can also do so to all other expressions constructed from atomic expressions of D combined using syntactic structures employed in any of the complex expressions of D.

Several prominent authors argued that productivity and systematicity (P&S) can best be explained if we assume that FL is (i) a recursive computational system with (ii) a compositional semantics (Fodor [44,52]). We will see later that some—esp. prototype theorists—have questioned whether compositionality is really the 'best' explanation of P&S; but we will also see that explaining P&S is *not* the only reason for assuming compositionality. For now, the important point is just that compositionality should be understood as a constraint on the ways in which FL can determine the meanings of complex expressions:

(C) FL determines the meanings of complex expressions from the meanings of their syntactically immediate constituents, the way they are combined, and nothing else besides.

C guarantees that if s can determine the meaning of *purple apple* and *red ball*, then s has the competence to determine the meaning of *red apple* and *purple ball*. Note that C doesn't tell us which particular procedure determines the meaning of each type of complex expression—VPs, NPs, and so on. For example, it allows that one type of procedure determines the meaning of Advs combining with Vs and another type of procedure determines the meaning of Adjs combining with Ns. In other words, C specifies a *general* constraint on the ways in which FL determines meanings, in particular, it limits the potential sources of information drawn upon. We will call general semantic constraints on FL such as C 'meaning-determination constraints' (MDCs).

As formulated, C does not say anything about what the meaning of words and phrases are. But we can spell-out what C entails for models of compositional prototype combinatorics in terms of two generals constraints:

(Semantic Locality) If α is a complex expression constituted by $\{\gamma, \beta\}$, only features taken from the prototype of γ or the prototype of β can go into the prototype of α.

(Uniform Modification) If α is a complex expression constituted by $\{\gamma, \beta\}$, and of the form $[_Z X\ Y]$, only operations defined for all expressions of type Z can determine the way in which (i) features of γ and β are inherited by α and (ii) the weights on the dimensions of the prototype of α are adjusted.

As before, this leaves open what *particular* operations determine the prototype of different types of expressions, as long as they satisfy Semantic Locality and Uniform Modification. For example, the operations which determine the meaning of [N N] constructions might be different from those that determine the meaning of [A N] constructions. In what follows, we will focus, following others in recent debates, on intersective [A N] expressions such as *purple apple*. The crucial property of these expressions is that if x is an [A N], then x is an N.[4]

As an example of a compositional prototype model of [A N], consider the Selective Modification Model (SMM) of Smith et al [180]. SMM follows a simple procedure: it selects the dimension of the head N under which the modifier A falls, replaces the feature/s in that dimension with the value of A, and increases

[4]There is much debate about the basic types of [A N] modifications (Partee [124], Morzycki [114]). Most linguists begin with a classification that includes intersective, subsective, modal, and privative modifications. In modal and privative modifications—such as *former president* and *fake gun*—that x is an [A N] does *not* entail that x is an N. The modal cases do not present any challenges for prototype theory in particular: e.g., they shift the time or world of evaluation, and do not affect the content of the head N. The privative cases are problematic for all theories, so critics of prototype compositionality justly do not appeal to them in their critiques. Still, chapter 4 above argues that privative modifications (e.g., *fake gun* and *counterfeit dollars*, and some subsective modifications (e.g., *fast car* and *bad gun*) actually support the idea that the meaning of head Ns includes a prototype component, on which the modifiers partly operate. So it is not unfair to say that just those modifications that are hard for classical theories are easy for prototype theories, and vice versa. This is why it is particularly important for prototype theories to deal with simple intersective modifications.

the relative weight of the modified dimension. For example, when *purple* modifies *apple*, the color dimension in the apple prototype is selected, its relative weight increased, and its values replaced with the feature +PURPLE. This model has two key features:

- Most of the features of the prototype associated with the head N are preserved in the resulting complex prototype for the [A N], and *no* features that do not come from either the A or the N are added to the resulting prototype for the [A N]. This accounts for the fact that it is generally the case that if x is an [A N] then x is an N, while at the same time it respects Semantic Locality.

- When constructing a prototype for *purple apple*, the weight and value of the color dimension of *apple* is changed, but the weights and values of the other dimensions are preserved. Furthermore, the way in which the weights of modified dimensions of the head N are changed can be spelled out in a general way that makes no reference to particular [A N] combinations. This procedure respects Uniform modification.

Following others, we will call this particular procedure for determining the prototype of [A N] expressions, 'default to the head stereotype' (DS). Let us clarify the notions introduced thus far. C is a general constrain on the ways in which FL can determine the meanings of complex expressions. Spelled out for a prototype theory of meaning, C amounts to two constraints: Semantic Locality and Uniform Modification. DS is a *particular* procedure for determining the meaning of [A N] expressions which satisfies Semantic Locality and Uniform Modification.[5]

[5] To be clear, most models of prototype combinatorics do not stop at the composition stage. This is certainly true of Hampton's Composite Prototype Model and Costello and Keane's C^3 Model. What is important for now is only that they include an initial compositional stage of prototype combination, even if they also try to model at least some of the most common post-linguistic pragmatic modifications.

Most models of prototype combinatorics for [A N] phrases follow DS for two reasons. First, as we said above, for most complex expressions of the form [A N], it holds that if x is an A N then x is an N. To account for this, we must assume that the prototype of the head N is relatively unaltered in the resulting [A N]. For example, whatever happens to the prototype of *apple* when it combines with *purple*, it must still be sufficiently unaltered to entail that if x is a *purple apple* then x is an *apple*. Secondly, by respecting C, DS ensures that our conceptual competence is productive and systematic. This is because DS ensures that the procedure for combining prototypes in [A N] structures is uniform (including the readjustment of dimension weights), and that all the features of the complex are derived from the constituents. It follows from these assumptions that if s can determine the prototypes for *purple apple* and *red ball*, s thereby has the competence to determine the prototypes for *purple ball* and *red apple*.

5.3 The problem: emergent features

At this point, we can clearly specify the objection to DS and other procedures for determining the prototypes of complex expressions that respect Semantic Locality and Uniform Modification. Just as in the *pet fish* case, there are many examples of complex expressions in which general beliefs seem to intrude into the associated prototype, resulting in emergent features. Consider some famous early results (Hampton [64], Johnson and Keil [80]). Subjects said that +SPIKED TIRES is a typical feature of *arctic bicycle*, but not of either *arctic* or *bicycle*. Subjects also said that +IDEALISTIC is a typical feature of *Harvard-educated carpenter* but not of either *Harvard-educated* or *carpenter*. Unlike *pet fish*, neither *arctic bicycle* nor *Harvard-educated carpenter* stand for concepts with whose instances we are likely acquainted. So in these cases it seems undeniable that background theories and general beliefs play a role in producing the emergent features.

Indeed, whenever we modify a N with an A that stands for a property that is unusual for the N—e.g., *neon-green carrot* and *overweight cheetah*—the prototype associated with the resulting phrase seems to include emergent features. For example, most of us would likely associate *neon-green carrot* not with a carrot that is neon-green but just like a carrot in all other respects, but with one that lacks the feature +EDIBLE or has the feature +POISONOUS, although these modifications/features are not directly inherited from *neon-green* (e.g., some neon-green candies are quite edible). In these sorts of cases, new features are added or removed from dimensions of the prototype of the head N which are not directly modified by the A. This suggests that DS is systematically violated by the apparently non-compositional combinatorial processes which construct prototypes for complex NPs.

For our purposes, the most interesting study of emergent features is Study 1 of Johnson and Keil [80]. They asked participants to generate features for novel NPs such as *arctic bicycle* and *hospital bicycle*. They also asked other participants to generate features for the modifiers and head nouns, e.g., for *arctic*, *hospital*, and *bicycle*. Most of the features which ranked as typical for the novel NPs were emergent, i.e., were not typical features of the constituents. In these examples, *arctic bicycle* generated the emergent feature SPIKED TIRES and *hospital bicycle* generated the emergent features METERS and DIALS. These studies seem to confirm the suspicion of philosophers such as Fodor, who argue that key features of the prototypes associated with NPs do not seem to be derived from the prototypes of the constituents, and hence violate compositionality (in particular, Semantic Locality).

5.4 The usual response: abandon C

In response to the emergent features objection, some prominent prototype theorists argue that we should replace C with a non-compositional meaning determination constraint (MDC) (Hampton and Jonsson [84], Prinz [144]. To account for emergent features, they present multi-step models of prototype combination that allow general beliefs to partly determine the meanings of complex expressions. Let us examine the viability of these proposals.

The move of this reply to the emergent features objection is to grant that phrasal interpretation with prototypes systematically violates C, and to then argue that it satisfies a MDC that is not compositional but is still compatible with P&S. The following formulation captures the essence of the proposed MDC:

(*Modal C*) As a default, FL determines the meaning of complex expressions from the meaning of their syntactically immediate constituents, the way they are combined, and general beliefs. As a fallback, FL determines the meanings of complex expressions following C.

In this formulation, 'general beliefs' is meant to encompass not only beliefs about particular members of some category (often called 'extensional beliefs'), but also explicit and implicit beliefs and theories about all sorts of things. On this view, the extensional belief that pets tend to live in houses can enter into the linguistic mechanism that determines the meaning of *pet fish*. Similarly, the general belief that most neon-green vegetables have been artificially altered, and that artificially altered vegetables could be dangerous, can affect the meaning of *neon-green carrot*.

As Prinz [144] notes, although *Modal C* is not compositional, it still allows that FL is always *capable* of combining prototypes compositionally, even if in most actual instances this is not the selected procedure. In other words, according to *Modal C* FL *can*—but need not—determine the meanings of complex expressions from the meanings of their syntactically immediate constituents, the way they are combined, and nothing else besides. Now, if compositional processes are only optional processes of phrasal interpretation, how does FL 'decide' when to interpret phrases compositionally? *C* is a fallback strategy: FL only interprets expressions compositionally when it lacks 'relevant' extensional and general beliefs. For example, subjects assign to the prototype of *neon-green carrot* the emergent feature +INEDIBLE because there are relevant beliefs relating unusual and strange neon colors with non-edible properties. However, subjects that have no such relevant beliefs can revert to a purely DS-based construction of the prototype for *neon-green carrot*. Since it allows compositionality as a fallback strategy, *Modal C* is compatible with P&S.[6]

Despite the relative consensus within the prototype camp, *Modal C* is deeply flawed. The underlying reason for this is that explaining P&S is *not*—and not by a long shot—the only important desiderata on accounts of MDCs. Other relevant properties of FL—often ignored by those who adopt *Modal C*—are the following:

(I) FL can interpret expressions in real-time and in such a way that the FLs of various speakers of the same language succeed in real-time communication (Pagin [120]). This adds additional complexity constraints on the way in which FL must satisfy P&S.

(II) Even if FL is not an informationally isolated system, it exhibits a sub-

[6] One reason why replacing *C* with *Modal C* might seem like an innocent move is that the original—and still widely cited—justification for *C* is that it explains P&S [44, 71]. However, we saw in chapter 2 that recent work shows that some non-compositional MDCs—including *Modal C*—can also explain P&S [38, 83, 194]. Often this point is simply ignored by critics of prototype theory, as can be clearly seen in the outdated discussions of the relation between P&S and compositionality in Connolly et al [30, 55]. At any rate, the compatibility of non-compositional MDCs such as *Modal C* with P&S opens space for the previous response to the emergent features challenge: drop *C* and adopt models which allow for the interference of top-down beliefs. Several authors who defend theories of lexical concepts that include something like prototypes as components endorse the *Modal C* solution to the problem of emergent features. See, for example, Weiskopf [209], Robbins [167], and Schurz [178].

stantial degree of modularity in the sense that its operations, including semantic ones, are automatic and show characteristic patterns of development and breakdown (Larson and Segal [101], Borg [15]).

(III) Semantic processes interface with syntactic processes, so MDCs should cohere with the basic assumptions of our best syntactic theories (Collins [29], Jacobson [79], Pietroski [136]).

(IV) MDCs should be compatible with patterns of early language acquisition and development (Chapter 2 above). This interacts with (III), since some of the basic assumptions of current syntactic theories respond to language learning considerations.

The real advantage of assuming C is that it coheres extremely well with these additional properties of FL [29, 120, 192]. Indeed, it is because of how well it fits in this wider network—and not just to account for P&S—that most linguists still accept, especially in practice, something like C. In contrast, non-compositional MDCs such as *Modal C* are arguably in tension with most of these additional properties of FL.

For our purposes, the most important tension is that between *Modal C* and (I) and (II), namely, real time processing, information access, and successful communication. *Modal C* does not say anything about *which* sorts of general beliefs can affect how FL determines the meaning of phrases. Unless we introduce some restriction, an FL so unconstrained would, before assigning a meaning to a phrase, have to search its entire long-term memory base to determine if there are any extensional or general beliefs that might be relevant for its interpretation. At the same time, it is completely unclear what sort of restriction on the set of beliefs might do the required work. For example, we cannot hold that only extensional beliefs enter into the combinatorial operations of FL, for then we could not account for many examples of emergent features, such as those observed in the case of *arctic bicycle* and *neon-green carrot*.[7]

Theorists who propose replacing *C* with *Modal C* present their examples in ways that hide this problem. For example, they say that based on our extensional beliefs we take the prototype for *pet fish* to include the emergent features +SMALL and +LIVE IN TANKS. However, these are not the only available relevant beliefs, so it is not at all clear why those emergent features would be the ones selected by a model which follows *Modal C*. Most mature speakers have other beliefs about pet fish, including highly idiosyncratic ones. In addition, they also have beliefs about pet animals, about the differences between pet mammals and pet non-mammals, about people who like pets, fish, pet fish, and so on.[8] Indeed, even more abstract beliefs about how NPs are or should be interpreted might, if FL was so unconstrained, affect phrasal interpretation.

In short, *Modal C* has no account of how FL performs the search for 'relevant' beliefs in real time, or of when it stops so that further linguistic processing (including in some cases the fallback compositional processes) can continue. In addition, *Modal C* can't be coupled—at least in any obvious way—with an explanation of how different FLs assign meanings to phrases that are sufficiently similar to ensure the successful communication of speakers of the same language but with different sets of beliefs. For these reasons, addressing the emergent features objection by adopting a non-compositional MDC along the lines of *Modal C* is a move that will not persuade skeptics of prototype theory, especially

[7]The proposal that only something like highly weighted beliefs enter into compositional operations is also problematic. Firstly, this would still result in a rather intractable set, since we have highly weighted beliefs not only about the direct extensions of noun-phrases but also about more abstract categories that could be relevant for the composition. Secondly, different speakers share some highly weighted beliefs but also differ in many others. In short, much work would have to be done to make this view compatible with the relative success of linguistic communication. For a critical discussion of various proposals for how to restrict the set of general beliefs, see chapter 2 above.

[8]Of course, in experimental settings these additional features are not as reliably assigned to *pet fish* as are features such as +SMALL and +LIVE IN TANKS. But the question here is whether the processing models which use an MDC such as *Modal C* successfully predict this *constrained* set of emergent features.

those in linguistic semantics whose central concern is to explain how the meaning of novel phrases is determined from that of their constituents.[9] For theorists engaged in the project of building a combinatorial theory of meaning—and especially those who see themselves as contributing to the project of building a computational psychology for higher-cognition—to say that our model of phrasal interpretation tells us that the meaning of complex expressions is determined from (i) the meaning of the parts and (ii) all other relevant beliefs, is basically to admit that we have no serious model at all.[10]

Still, I agree with Prinz, Hampton and Jonsson that we cannot just ignore the overwhelming evidence that prototypes do seem to be conceptual components of lexical terms.[11] In what follows, I will argue that we can address the problem of emergent features by pairing a strong but context-sensitive account of C with an 'internalist' conception of prototype representations. I first explain and defend these two notions (§5.5- §5.6). I then illustrate the advantages of this revised framework by revisiting some classic examples of emergent features (§5.7). Finally, we discuss how this framework fits with recent experimental results.

[9] We cannot discuss here the tension between *Modal C* and the other basic properties of FL, but I should note that there are also serious problems with the syntax interface (III) and language acquisition (IV). For further discussion, see Collins [29] and chapter 2.

[10] Interestingly, early critics of formal approaches to linguistic meaning, such as Travis [201, 203, 204], basically wanted to show that general beliefs were involved in (broadly speaking 'literal') phrasal interpretation in the unconstrained ways granted by prototype theorists. In their minds, to grant this much was basically to give up on the project of constructing formal and computational models of the semantic aspects of language processing. The criticism of *Modal C* presented here agrees with this perspective.

[11] Since this section focuses on some important disagreements with prototype theorists such as Prinz and Hampton, I should also emphasize that there is much in their enlightening work on prototypes that I agree with. Indeed, as will become clear in what follows, I think Prinz's RCA model and Hampton's Composite Prototype Model are proposals in the right direction. With some re-formulation, these models can adopt the framework I will propose. We come back to this point.

5.5 Compositionality revisited

Despite the previous criticism of weak meaning-determination constraints such as *Modal C*, it is clear that, if it is to work as a serious constraint on how FL process expressions, C must allow some forms of context-sensitivity. The task now is to present and defend a context-sensitive version of C. As will become clear, the meaning-determination constraint we defend is much more constrained than both *Modal C* and other radical contextualist proposals.

To begin, C must allow that some expressions have 'characters' as meanings. Let '$[\![e]\!]_c$' stand for the meaning of expression e in context c. Following traditional terminology, we then say that f_e is the character of e and $f_e(c)$ is the occasion meaning of e in context c. If e has no free parameters, then for all c's, $f_e(c) = m$, where m is the standing meaning of e. Examples of expressions with context-sensitive parameters are demonstratives (*that*), indexicals (*I*, *you*), and some Adjs (*small*, *heavy*). It is important to ask, when evaluating alleged counter-examples to prototype compositionality, whether the emergent features are due to the presence of an open-parameter, context-sensitive expression.

The real challenge to prototype theory comes from other kinds of context sensitivity, esp., general meaning modulations that cannot plausibly be attributed to the effect of context on open parameters. Recent contextualist proposals in Philosophy of Language and Linguistics suggest that we reformulate C to allow the processes which determine the meaning of complex expressions to take the 'modulated' (instead of the standing or occasion) meanings of their immediate constituents (Pagin and Pelletier [122], Recanati [164]). To allow this type of context-sensitivity, we have to say that FL determines the meanings of token complex expressions in terms of the modulated (instead of the standing or occasion meanings) of their constituents:

(C^{mod}) FL determines the meanings of tokens of complex expressions from the modulated meanings of their syntactically immediate constituents, the way they are combined, and nothing else besides.

Note that C^{mod} does *not* allow that the combinatorial *operations* that determine the meaning of complex phrases are themselves context-sensitive. The combinatorial operations are as fixed and compositional as before. For example, if we assume that DS is the compositional procedure for determining the meaning of complex NPs, we can keep that assumption, mutatis mutandis, in the context-sensitive framework here proposed. What is flexible is the relation between the standing meaning of words and the meaning of their tokens in particular uses.

However, to avoid the criticism presented against *Modal C*, and for other reasons we discuss shortly, we will add two substantial additional constraints to C^{mod}. The first constrain has to do with the levels of syntactic structure at which modulation can apply. As C^{mod} is formulated, modulation could apply at every level of phrasal construction or branching node. However, we will assume that modulation applies only to lexical items, i.e., to the terminal nodes of syntactic trees. So modulation does not operate on non-terminal nodes such as at the highest node of a complex NP of the [A [A N]] form. This allows a constrained form of meaning modulation which can be described as pre-compositional, in the sense that the effects of the context-sensitive modulation function enter only via the terminal nodes of expressions.

The second constraint relates to the types of operations that are available to modulate the meanings of expressions. On the radical contextualist view, there is a modulation function which, in addition to applying at every level of structure, shifts the meaning of expressions so as to make them 'salient, appropriate or relevant' in a given context (Recanati [164]). Call this 'smart' relevance seeking function *mod*. We will severely constrain the kinds of meaning modulation operations that the Faculty of Language can perform. The basic idea is this.

Under most views, but especially under prototype views of lexical meaning, each lexical item stores a larger set of information in long term memory than it extracts when it is tokened for language processing task in working memory. We discuss this point in more detail in the next section, but for now some illustration will do. Consider the lexical entry for *book*. This entry likely includes information about some representative exemplars, and it also includes information about the typical function of books, e.g., to carry information. When you are thinking how much it might cost to ship some books, you might only consider their physical features; when you are thinking whether to read some book, you might only consider its informational content; and when thinking whether to take some book on the plane, you might consider a trade-off between its physical characteristics (e.g., its weight and size) and the value of its information content (e.g., whether it is entertaining). The 'modulation* function' is an abstract description of the process that selects, *within the options that are lexically available*, the most relevant aspects of the meaning. Call this lexically constrained modulation function *mod**.

Following these proposals regarding the way in which compositionality and some level of contextual sensitivity should be reconciled, we reformulate C^{mod} as follows:

(C^{mod*}) FL determines the meanings of tokens of complex expressions from the meanings of their syntactically immediate constituents, the way they are combined, and nothing else besides. If the syntactically immediate constituents are terminal/non-branching nodes, these meanings may be modulated*.

There are two reasons why we should adopt C^{mod*} over C^{mod}. First, most linguists now hold that syntactic/semantic computations work in phases that are sent off for pragmatic interpretation *before* full sentences or clauses are processed by FL. In Minimalist theories, the main phases are vPs and CPs, but due to the 'left edge condition', the phases that are sent out for pragmatic processing are more fine grained (Chomsky [26, 27], Cook [31], Redford [159]). Theorists who adopt Categorical Grammars also hold that the outputs to pragmatics are sub-sentential phrases such as DPs and VPs (Jacobson [79]). If interpretation proceeds in such sub-sentential phases which are sent to pragmatic interpretation, there is no reason why we should incorporate into the semantics a generalized version of *mod*, i.e., a function which modulates both the inputs and outputs of the compositional operations. Such output modulations would be redundant. Furthermore, we will see that in actual case studies most of the modulation operations required to cover the data operate on lexical items.

The second reason to favor C^{mod*} is especially pertinent to those concerned with the compositionality of prototypes. There is plenty of evidence for—and some initial models to explain—the sort of context-sensitivity allowed by *mod*∗ [34, 115, 172]. Consider a series of studies presented by Roth and Shoben [172] that test the ways in which the prototype of words like *animal* shifts across contexts. In context such as "Stacey volunteered to milk the animal whenever she visited the farm", subjects showed a prototype effect (e.g., ease of categorization reflected by reading times) for *cow* and *goat* relative to *horse* and *mule*. However, in contexts such as "Fran asked her father to let her ride the animal", subjects showed the opposite prototype effect, favoring *horse* and *mule* [172]. A second experiment shows that the prototype shifts can be quite subtle. For example, when processed in contexts such as "During the midmorning break the two secretaries gossiped as they drank the beverage", subjets report a goodness of exemplar distribution that ranks coffee as a more typical beverage than tea, and tea as more typical than milk. When processed in contexts such as "Before starting his day, the truck driver had the beverage and a donut" coffee remains the most typical beverage, but now milk ranks as more typical than tea. These experiments suggest that the prototypes associated with lexical terms for natural and artifact kinds are modulated—sometimes substantially, other times subtly—as a function of the broader context.

Barsalou [5] presents a detailed account of the context-sensitivity of lexical prototypes. Barsalou agrees that lexical prototypes are relatively unstable: they change across contexts as a function of, among other things, the preceding discourse and relevant goals. According to Barsalou, this 'instability' of prototypes is partly explained by distinguishing between the knowledge associated with words in long term memory, and the token prototypes extracted to working memory to perform particular processing tasks. The prototypes extracted to working memory are informationally pruned, relative to their corresponding long-term memory databases. This processes of selective activation is influenced by various contextual factors, such as preceding discourse, salient cues, goals, and relevant task constrains. What is crucial to note here is that this type of lexical-level context-sensitivity is basically a byproduct of general properties of the processes that underlie the extraction of information from long-term to working-memory. In this sense, we get this type of meaning modulation for free.

However, this same process cannot be used to explain modulation once the prototypes or lexical items are in working memory (as e.g. when a prototype for a non-terminal node is combined with another prototype). This means that if we also want to account for modulation at non-terminal nodes as required by the general implementation of *mod* favored by Recanati (and although they are not explicit about this, presumably also by those who favor *Modal C*), we would have to include additional context-sensitive and top-down meaning modulation operations on working memory units for which we have no independent evidence. Furthermore, the general *mod* operations envisaged by Recanati and other contextualist are just the sort of operations that are extremely taxing to working memory. For on this view, the operations of default language processing in everyday tasks involve holding single lexical items and phrases in mind, comparing them with each other and the context, and performing modulation operations on them, operations which are responsive to other single items and phrases, also held in mind and operated upon. We can say then, that on the general implementation of *mod*, 'calibrations' performed between items in working memory are general feature of ordinary language processing. One problem with this view is that it is well-known that even very simple calibration operations on two or more items in working memory can be extremely taxing. In particular, calibrations are taxing in a way that does not seem to correspond to the usual easy flow of everyday language processing.[12]

More importantly, the general implementation of *mod* seems to over-generate, i.e., to entail that there should be certain meaning shifts that are not available (Asher [1]). Consider the following examples:

(5.1) a. Mary stopped the apple.

b. Mary stopped eating the apple.

[12] Although we will not rely heavily on this particular criticism of *mod*, it is worth pointing out some evidence which, albeit only suggestive, is quite interesting. Kahneman has familiarized many readers with the famous Add-n tasks used to study cognitive effort. In the basic paradigm, subjects are presented strings of three or four digits, e.g., 4176. They can in addition be given some simple overlapping task, such as following a rhythm. After the string has disappeared, they have to perform some operation on each of the digits, such as Add-1, and then report the results, in this case, 5287. Using a pupil dilatation paradigm, Kahneman was able to measure the degree of cognitive effort required of each type of task. The example we mentioned would have been somewhat effortful, but if it involved an Add-3 rule, it would have reached the limit for most of us. In his recent account, Kahneman adds an anecdote that is very suggestive. In the breaks, some of the participants engaged in casual chats with the experimenters. Sometimes the participants did not move from their place, so Kahneman and colleagues could see the pupil dilatation that corresponded to the causal conversations. What surprised them was that it was extremely small, specifically, much less than any of the tasks that involved operations on multiple items held in working memory. The level of effort measured during casual chats was roughly the same as that required to hold two items in mind, without performing any operations on them.

(5.2) a. John's dog ate Hamlet.

b. John's dog ate the copy of Hamlet.

In most contexts, the meaning expressed by (5.1b) would be the most relevant or salient interpretation of (5.1a). Likewise, the meaning expressed by (5.2b) would be the most relevant or salient interpretation of (5.2a). However, those meanings are unavailable. If a relevance seeking *mod* function could apply to certain key items at any level of interpretation, these readings would be easily obtained. So not only do the 'calibration' operations of the general *mod* function impose rather severe demands on working memory, but, in addition, such a function would generate meanings that are not available. Hence general *mod* faces both explanatory and descriptive problems. If we can model prototype compositionality using only *mod**—as I will argue in the next sections—there is no reason to postulate that ordinary language processing generally involves something like the unconstrained *mod*.

Having replaced C with C^{mod*}, we should spell out the corresponding revisions to our account of prototype combinatorics. Fortunately, these are rather straightforward:

(Semantic Locality) If α is a complex expression constituted by $\{\gamma, \beta\}$, only features derived from the modulated* prototype of γ or the modulated* prototype of β can go into the prototype of α.

(Uniform Modification) If α is a complex expression constituted by $\{\gamma, \beta\}$, and of the form $[_Z X\ Y]$, only operations defined for all expressions of type Z can determine the way in which (i) features of γ and β are inherited by α and (ii) the weights on the dimensions of the prototype of α are re-adjusted.

According to Semantic Locality, the features of the prototype of a particular complex expression e in context c are derived from the modulated* prototypes of the lexical constituents of e in c. According to Uniform Modification, the combinatorial operations themselves are not context-sensitive.

5.6 Lexical prototypes revisited

The next step in our defense of prototype compositionality is to revisit the notion of 'prototypes'. As we said above, many discussions, critical or sympathetic, of compositionality and prototype theory basically 'define' prototypes as structured representations of features that are typical, diagnostic, or some function of both, relative to their associated categories. Feature x is typical, relative to category A, if the probability is high that an entity has x if it belongs to A, and x is diagnostic if the probability is high that an entity belongs to A if it has x. For our purposes, the key feature of this account is that prototypes are individuated via some statistical central tendency relation to their associated categories. The most prominent central tendency is typicality, but in some cases (e.g., perceptual or magnitude concepts) it can be an average value. Let us call this view 'externalist', since it emphasizes the way in which prototypes encode statistical central tendencies of the external categories they represent.[13]

However, to assume that, in general, prototypes either do or aim to encode the statistical central tendencies of their associated categories is clearly mis-

[13] The tendency to individuate prototypes via the statistical route might come from two sources. First, early prototype theorists tried to defend a pure form of the theory, namely, that concepts are just prototypes. From this perspective, it might seem that prototypes can only perform the required referential function if they somehow encode the actual statistical central tendencies of the category they represent. However, most current prototype theorists do not hold that prototypes are the sole components of concepts, nor do they think that, in most cases, they have an extension determining role [65, 104, 142, 144, 171]. Secondly, some the earliest and most influential studies in support of prototype theory came from perceptual categorization and learning [170]. When presented with categories of dot patterns made by distortions of a central tendency, or of stick figures in which an average figure is central, subjects learned categories faster if the better examples, particularly the most central members, are presented first, and were faster at categorizing items which are most similar to the most central members. (for an overview, see Murphy [118]).

taken. As emphasized by Rosch [171], it is more accurate to think of prototypes as a broad class of types of graded conceptual structures. Prototypes with features that encode central tendencies form an important subclass, but other types of prototypes use features that are e.g. diagnostic or salient, but not typical, of their associated categories. These features are often the extreme values of dimensions. Most people think of the best examples of cities as the largest and most cosmopolitan, rather than average cities. Similar points apply to concepts such as SAINT and GENIUS: intuitively, our prototypes for these concepts use the most extreme features associated with sainthood and brilliance. Consider also certain concepts for social roles such as PRESIDENT or MOTHER: the prototypes associated with these concepts often encode expectations that can be rather uncommon for actual members of the class. Indeed, even perceptual prototypes such as those associated with color and geometrical shape terms might not capture any statistical central tendencies of their corresponding real world categories (e.g., in what sense is fire-engine red either typical or an 'average' of instances of red?).

One factor that determines whether we acquire prototypes with non-typical features is the way in which we acquired it. For example, we often abstract prototypes using salient features of the members of the category we first encountered, and subsequent experiences with other category members often show that the features initially extracted can be rather unusual. Similarly, repetitive and/or emotionally charged media bias is partly responsible for certain social stereotypes which do not encode central tendencies. Unfortunately, these misleading prototypes can be hard to modify. These points are nicely illustrated by a famous experiment conducted by Hamill and Nisbett [62]. Subjects were presented with a video interview of a supposed prison guard. Half the subjects viewed a humane guard, and half a brutal one. Subjects were then given statistical information regarding the representativeness of the prison guard in the video relative to the class of prison guards. It was clearly emphasized by the experimenters that this information was reliable. Still, subjects who saw the video of the humane guard used, in subsequent tasks, a prototype of prison guards in general as humane, even when they were explicitly told that the humane guard in the video was highly atypical. Similarly, subjects who saw the brutish guard then used a prototype of prison guards in general as brutish, even when they were told the brutish guard in the video was highly atypical. In short, subjects extracted prototypes for the category of prison guards based on the initial, emotionally salient encounters, and were highly insensitive to subsequent, reliable information regarding the degree of typicality encoded in their original prototypes.

For our purposes, the crucial point is the following: once we acknowledge that the way in which we acquire a prototype can determine whether or not it encodes central tendencies of the associated category, we are forced to accept that the linguistic-compositional way of obtaining prototypes might often *not* result in structures that encode central tendencies. If someone made, for the first time ever, neon-green carrots and they happen to be quite healthy, then my linguistically determined prototype for *neon-green carrots* lacks an inductively useful feature of the corresponding category. For all I know, if we count, among all the pet fishes, those that currently dwell in Asia, then maybe their typical feature is that they live in artificial garden ponds rather than small in-house tanks. Of course, once the prototypes for (esp. novel) complex NPs are produced by FL, we often immediately begin to modify them—using general reasoning and whatever tools we have at our disposal—so that they do encode such central tendencies, since this (sometimes) increases their usefulness for categorization and induction.[14]

[14]Interestingly, the reason why concepts for social roles sometimes do not encode statistical

Let us call this pluralistic view of prototypes, ‘internalist’. The distinction between externalist and internalist views of prototypes is crucial in discussions of compositionality. To know whether in general we can compose the prototype of a complex phrase from the prototypes of its constituents, we have to individuate the input and output prototypes. In many cases, especially those involving the prototypes of complex expressions, the internalist and externalist views suggest different ways of individuating prototypes, which often result in importantly different assumptions about the features of particular prototypes which FL has to account for. Indeed, even researchers who in theory accept something like the internalist view of prototypes—e.g., Hampton and Jonsson [68]—in practice (e.g., when interpreting experimental data, such as when they decide what subjects reported ‘intuitions’ are about) individuate the prototypes of complex concepts in a way that exaggerates the extent to which they encode central tendencies. As we will see next, this common slippage makes the compositionally challenge seem insurmountable.

5.7 *Pet fish* and *beach bicycle* reconsidered

We have now presented the two main components of our revised framework for dealing with prototype combinatorics. The first component is C^{mod*}. Since we are dealing with intersective NPs, this amounts to a context-sensitive version of DS. The second component is an internalist and pluralist account of prototypes which emphasizes that the features associated with simple and complex expressions often do not encode statistical central tendencies. Call this framework ‘CCIP’ (for Context-sensitive, Compositional, and Internalist Prototypes). The

central tendencies might be because they often have a social-normative use (Leslie [102]). For example, one crucial use of identifying someone as a FATHER is to impose norms of how he ought to behave. Of course, such classification can also often be made to predict how he will in fact behave. But in these cases, the role of reliable predictions might have a lower priority relative to that of norm imposition.

task now is to use CCIP to address the emergent features objection to the compositionality of prototypes. We will do this in two steps. First, we reconsider the issue of emergent features. We will illustrate the steps that, according to CCIP, determine the prototypes of complex expressions and show that CCIP can be easily reconciled with the usual examples of emergent features. Second (in the next section), we consider some experimental evidence that supports key aspects of CCIP.

Consider again the infamous *pet fish*. Strictly, the prototype for *fish* is probably quite complex, but let us assume for simplicity that a typical trout is indeed a really good exemplar of the prototype associated with the standing meaning of *fish*. The prototype associated with the standing meaning of *pet* likely includes some typical pet functions, even if in some occasions we also use more static perceptual features associated with typical pets. Given those choices, what is a reasonable candidate for a compositionally determined prototype for *pet fish*? If we assume a context-*insensitive* version of DS, then this prototype would seem to include perceptual information about some rather ordinary fish, and functional information about the typical functions associated with pets. A good exemplar of this would be a trout that keeps one company or that lives in one's house. As critics point out, this result would seem problematic.

However, the constrained context-sensitive version of DS adopted by CCIP entails that when processing *pet fish* the prototypes of both *pet* and *fish* are modulated* by the context. Recall Roth and Shoben [172]'s results: in contexts that invoke *milk*, the prototype for *animal* is more like an encoding of a cow and a goat than of a horse and a mule, while in contexts that invoke *riding*, the opposite holds. There is no reason to think that terms such as *pet* and *fish* are less context-sensitive. It follows that in the context of talking about fish, the prototypical functions associated with *pet* involve more aesthetic or ornamental uses than companionship or cuddling uses. Similarly, in the context of talking about pets, the perceptual features associated with *fish* will be closer to those typical of smaller ornamental fish than to those typical of wild fish.

Still, even after modulation* of the prototypes associated with the constituents of *pet fish*, the linguistically determined prototype, assuming a process of combination that respects DS, is unlikely to encode at least some prominent central tendencies of the category of pet fish. So the features of the linguistically determined *pet fish* prototype—namely, that they are something like smallish ornamental fish—are unlikely to fully correspond to the features we associate, especially after some reflection, with *pet fish*. From the perspective of CCIP, there is nothing problematic about this. Our intuitions when asked and given time to think about the 'typical' features of *pet fish* are about the representation associated with *pet fish* after the prototype assigned by FL has been modified based on pragmatic considerations and using general beliefs. Since we have extensional knowledge about pet fish—e.g., their typical small size, bright colors, living conditions, and decorative uses—we can use that information to modify the perceptual and functional features of the linguistically determined prototype. This subsequently modified prototype associated with *pet fish* is closer to encoding the central tendencies of the category of pet fish.

Similar points can be made to deal with *arctic bicycle*, *Harvard-educated carpenter*, and other examples of emergent features that have actually been obtained experimentally (Johnson and Keil [80], Kunda et al. [96]). For example, in the context of talking about arctic or snow vehicles, the prototype of *bicycle* could be modified* to fit certain general facts about snow vehicles, such as that they need special tires. Still, it is likely that the prototype assigned by FL to *arctic bicycle* is something like an ordinary bicycle with a special function, say that of being usable in snow and ice. This is just an informal way of saying that the combinatorial processes follow DS. This prototype might well fail to encode some important central tendencies of the category of arctic bicycles. For example, maybe most arctic bicycles do not look like ordinary bicycles: have special pedaling systems and two slides with a manual propeller instead of wheels. To generate a richer and more inductively useful concept, the linguistically determined prototype for *arctic bicycle* has to be pragmatically modified. In this case, most of us do not have any relevant extensional beliefs, since acquaintance with exemplars of arctic bicycles is rare. However, we have plenty of relevant general beliefs, e.g., that vehicles designed for snow-use need specially powerful traction systems and tend have special types of tires. We also know something about why this is the case. We can use these general beliefs to modify the prototype associated with *arctic bicycle* so that it includes emergent features such as SPIKED TIRES. This process is comparatively slow and deliberate. This is why subjects usually take a couple of minutes to generate emergent features for novel NPs.

In short, if we adopt CCIP the cases of emergent features found in the literature do not threaten prototype compositionality. According to CCIP, there are two sources of emergent features. First, the prototypes associated with lexical terms are often modulated* to select the subset of the lexically available information that is relevant in the linguistic and discourse context. These shifts—which can be quite subtle—may result in the addition or elimination of features relative to the default or standing prototypes of lexical terms. Secondly, the compositionally determined prototypes associated with complex expressions are often taken as input by general pragmatic processes which modify them so that they better encode features relevant for the current task (which often requires aiming for statistical central tendencies). These two sources of emergent features are perfectly compatible with the compositionality of prototypes.

5.8 Experimental findings reconsidered

We have seen that there are good theoretical and empirical reasons to accept CCIP. In particular, there are good reasons to adopt the CCIP account of emergent features over accounts based on non-compositional constraints such as *Modal C*. The question I now want to ask is whether there is any direct empirical evidence for CCIP. I will discuss two important experiments which support CCIP over accounts which assume something like *Modal C*.

5.8.1 Time-sensitive compositionality experiment

Taken as evidence agains the compositionality of prototypes, the main shortcoming of the classic emergent features experiments is that they do not control for the time-course of the (combinatorial) operations which determine the emergent features of phrases. Insofar as subjects could reason, without significant time constraints, about the sorts of features which different classes could have, some of these features could be added after the combinatorial operations of FL have already been carried out. According to CCIP, this is precisely what is happening. Studying the time-course of the combinatorial operations of language is extremely challenging. However, McElree et al [110] carried out a very clever study which is directly relevant to this discussion.

To study the time-course of composition, McElree et al [110] used a speed—accuracy trade-off paradigm. Participants were presented with sentences consisting of a simple or complex NP as subject and a predicated property. Their task was to verify whether the predicated property was true (or false) of the subject. The participants were required to respond within six predefined time windows, and the dependent measure was response accuracy within each time window. Consider the following representative examples:

(5.3) a. Boiled celery is green. [TN]

b. Boiled celery is soft. [TP]

(5.4) a. Water pistols have triggers. [TN]

b. Water pistols are harmless. [TP]

In the (a) cases, the predicated property is true of both the head noun, and the noun phrase ('TN' condition). In the (b) examples, the predicated property is true only of the noun phrase ('TP' condition). In the experiment, the noun-phrase appeared on a screen for 600 sec. Then the predicate appeared. Then either 300, 500, 700, 900, 1,500 or 3,000 msec after the presentation of the predicate a tone was heard that signaled to participants that they had to verify ('yes' or 'no') whether the predicate applied to the noun-phrase. Participants were trained to respond within 100 and 300 msec after hearing the tone.

The results were as follows. By 2 sec after the presentation *of the predicate* subjects were extremely accurate at verifying the predicates; in particular, by that time they were as accurate in the TP as in the TN condition. However, at all the earlier time windows, they were significantly more accurate for the TN condition than for the TP condition (i.e., they were significantly *less* accurate for the emergent features). The 600 msec given to view the subject NPs is a safe assumption for the time required to combine the constituents of the NPs (note that this 600 msec is additional to the time given to process the predicate, which were varied). So this experiment suggests that the features which are true only of the noun phrase—i.e., the emergent features—were computed (a couple of seconds) *after* the output of the initial combinatorial operations.

Importantly, the emergent features used in these experiment were all of the extension feedback variety. To illustrate with (5.3b) and (5.4b), note that most of us believe that boiled celery is soft and water pistols are harmless because we have seen some relevant exemplars (cf., water guns could of course be dangerous, and some foods get stiff/crispy when boiled). We can easily extract the emergent feature by using those exemplars. So it is arguably that this extension feedback based emergent features are the best candidates for the view that emergent features are computed by the non-compositional combinatorial processes of FL. In contrast, computing emergent features for novel NPs via reasoning or general beliefs seems to involve more complicated computations. It is safe to assume that their time-course would show an a significantly greater delay than even the extensional-feedback based emergent features.

Ultimately, that conjecture requires empirical support. However, to see why it seems like the best bet, recall some examples obtained by Johnson and Keil [80]: *hospital bycicles* had the emergent features METERS/DIALS and STATIONARY, and *arctic bicycles* had the emergent feature SPIKED TIRES. In a subsequent study, Johnson and Keil tried to tap into the reasoning processes behind the generation of emergent features by asking subjects to fill in reasoning schemes such as the following:

(R) Since the ARCTIC ________
and since BICYCLES ________
then arctic bicycles have spiked tires.

Note that the reasoning scheme ends with the emergent feature obtained in the previous study. In this case, most participants reasoned as follows:

(R) Since the ARCTIC *is covered with ice and snow*
and since BICYCLES *need traction*
then arctic bicycles have spiked tires.

In the reasoning scheme for *hospital bicycles*, most participants said that since hospitals *need to monitor patient's health* and since bicycles *require physical exertion*, hospital bicycles have meters and dials. As these examples illustrate, the reasoning behind the generation of emergent features for novel NPs can be quite subtle. Of course, in some cases we might use heuristics that bypass these results. Still, it is worth noting that subjects usually take several minutes to generate these sorts of emergent features. Hence, it is safe to assume, with CCIP, that the computation of these emergent features comes after the initial combinatorial operations of FL.

5.8.2 The modifier effect

We next consider an experiment, due to Connolly et al. [30], which was originally presented as decisive evidence *against* the compositionality of prototypes. According to [30], violation of DS is even more widespread than the classic examples of emergent features suggest. Even in cases where there are no obvious extensional (cf. pet fish living in tanks) or general beliefs (cf. Harvard-educated carpenters being idealistic), they argue that subjects do not seem to respect DS when constructing prototypes for [A N] expressions. To support that conjecture, they devised the following experiment. Subjects were presented with sets of four sentences, and asked to score each sentence on a scale of 1-10 for its likely truth, with 1 = 'highly unlikely' and 10 = 'highly likely'. Each set included sentences of four types (A-D):

(A) A sentence S with an unmodified familiar N as subject and a predicate that is typical of instances of the subject: *Squirrels eat nuts.*

(B) Modifies S by adding a modifier to the subject N that is a typical feature of the N: *Tree-dwelling squirrels eat nuts.*

(C) Modifies S by adding a modifier to the subject N that is a non-typical feature (but neither bizarre nor contradictory) of the N: *Nicaraguan squirrels eat nuts.*

(D) Modifies S by adding an additional modifier to the one in condition C:

Black Nicaraguan squirrels eat nuts.

According to Connolly et al. [30], models which accept DS predict that the likelihood of truth scores in each of these conditions should be roughly equivalent. For if you assume that features of the head N which are not directly implicated in the modification should be inherited, then *Nicaraguan squirrels* should be judged to be just as likely to eat nuts as *black Nicaraguan squirrels*, since in neither case are dietary habits directly implicated in the modification.[15]

That prediction is incorrect. For each set of sentences, there was a systematic deviation away from the likelihood of truth scores obtained in the baseline condition A. Sentences of Type B had reliably lower scores than those of Type A, although the modifiers in B stood for typical features of the head Ns. Furthermore, the introduction of one (in C) and two (in D) non-typical modifiers resulted in progressively lower scores. Here is another set of examples, including the (representative) average likelihood of truth score assigned to each type of sentence:

(1-A) Ducks have webbed feet (8.3)

(1-B) Quacking ducks have webbed feet (7.7)

(1-C) Baby ducks have webbed feet (7)

(1-D) Baby Peruvian ducks have webbed feet (6.5)

Connolly et al. [30] argue that these results suggest that subjects do not follow DS when determining the prototypes of complex NPs: features related to the type of feet of *ducks* do not seem to be directly inherited into the modified NPs,

[15] Connolly et al. [30] choose predicates that are not about the dimension of the subject head N that is being modified. For example, *black* modifies a dimension of *squirrel* that has nothing to do with eating habits, which is the reason why, they argue, it should not affect the confidence in *black squirrels eat nuts* relative to *squirrels eat nuts*.

even though the modifiers do not directly affect the dimension related to feet.[16] Assuming compositional models of prototype combinatorics for intersective NPs have to follow DS, these results seem to undermine such models.

For us, the crucial question is this: given CCIP, what results should we expect in these experiments? Arguably, we should expect something very close to the actual results. Consider the representative stimuli (1-A) to (1-D) above. The basic pattern is simple: a progressive decrease in confidence from sentences of Type A to those of Type D. In addition, note that even the lowest scores are still well within the 'more likely than not' range.

Let us begin by considering the aspects of this pattern which are easily explained. Firstly, as Connolly et al. [30] acknowledge, there is a Gricean explanation for the decrease in confidence when comparing sentence of Type A to sentences of all the other Types. When presented with sentences where the subject head Ns are modified, most readers likely assume that the modifications are significant. That is, the very act of modification implicates an interpretation of the head N that deviates somewhat from the unmodified head N. This is a reasonable explanation for some decrease in confidence in sentences of Type B-D compared to the unmodified sentences of Type A, but it does not explain the pattern of decreasing confidence from sentences of Type B to C (and perhaps not from B/C to D, but this is more debatable).

In addition, models of prototype combinatorics that follow DS use combinatorial mechanisms that reinforce the Gricean explanation (Hampton and Jonsson [68], Prinz [144]). Specifically, Hampton [64]'s feature pooling model

[16]The results could signal a violation of DS for two reasons. First, they could signal a violation of Semantic Locality. For example, if when computing the prototype of *baby ducks* or *baby Peruvian ducks* some emergent feature is added regarding their type of feet, this would explain why subjects are less confident that *baby ducks* and *baby Peruvian ducks* have webbed feet compared to *ducks*. Second, the results could also signal a violation of Uniform Modification. For example, if the way in which weights are readjusted in the case of *baby ducks* is different to the case of *quacking ducks*, this would explain why subjects' drop in confidence with respect to webbed feel is different in each case. In either case, DS would be violated.

and Smith et al. [180]'s SMM both assume that weights for dimensions and features are readjusted systematically when prototypes are combined. According to both models, in [A N] combinations the weight of the dimension of the N which takes the modifier is increased, which means that there is a decrease (the extent depends on the model) in the relative weight assigned to other dimensions and features. For example, both models predict that the dimension which specifies type of feet in the prototype of *quacking ducks* is less important than in the prototype of *ducks*.

Now, to satisfy Uniform Modification these weight readjustments have to be carried out in a uniform way. This entails that weight readjustment (just as the Gricean points) explains the relative pattern of confidence decrease from Type A to B/C to D cases, but not the decrease reliably shown between B and C cases (e.g., between *quacking ducks* vs.*baby ducks have webbed feet*). Since the NPs in both Type B and C cases have the same syntactic structure and number of modifiers, the decrease in confidence when comparing one typical (Type B) and one non-typical but non-conflicting modifier (Type C) has to be explained in some other way. We now turn to that.

According to CCIP, pragmatic modulation can operate on sub-sentential phrases. So CCIP entails that responses to whether the predicative VP *have webbed feet* is likely true of the subject NP, are sensitive to the prototype of the subject NP after it has already been modified by general beliefs to ensure that, among other things, it represents a coherent concept and, in some cases, one that encodes statistical central tendencies.[17] What general beliefs could account for the difference between Type B (NP with a typical modifier) and

[17]This point can be illustrated with a simple example. Relative to CCIP, what task is a subject performing if he is asked whether it is likely that *stone lions can be found in museums*? Most subjects immediately answer that this sentence is very likely true. This suggests that what they are doing is deciding whether the VP predicate applies to sculptures of stone lions and not to real lions made of stone. If they were considering the literal interpretation of *stone lions*, they would quickly respond that the sentence is very likely false.

Type C (NP with a non-typical but non-conflicting modifier)? Consider again the representative stimuli, now focusing only on B and C. The brackets indicate 'chunks' or phases that are sent for pragmatic interpretation:

(1-B) [Quacking ducks] [have webbed feet] (7.7)

(1-C) [Baby ducks] [have webbed feet] (7)

The key difference between Type B and C cases is that, since the latter involve non-typical modifiers, the subject NPs seem to stand for 'special' or 'atypical' subclasses of the head N. Being atypical, they are less likely than normal members (subjects of Type B) to have the typical properties (note: in all cases the predicates stood for typical predicates). Consider another representative example. Subjects were slightly more confident that *Expensive pearls are white* (Type B) than that *South Sea pearls are white* (Type C). Pearls are typically expensive and white. However, *South Sea pearls* seems to stand for a special subclass, so we cannot be so sure that they are white.[18]

5.9 Objections and clarifications

CCIP issues in an optimistic view of the future success of compositional models of prototype combinatorics. This is not to suggest, of course, that the task is easy. The construction of detailed models of prototype combinatorics is an active area of research; and a short journey through the literature shows that coming up with all the relevant details is extremely difficult. The task here is only to defend a general framework that, if adopted by particular models, insulates them against certain very general objections to the entire enterprise.

[18]Consider other atypical modifiers used by Connolly et al [30] to generate Type C cases. We find *Peruvian ducks*, *Norwegian lambs*, *Brazilian doves*, and *European ostriches*. We also find *white coyotes*, *albino crocodiles* and *yellow roosters*. The most obvious explanation here is that subjects interpret these atypical locations and physical attributes as signaling significant or even atypical subgroups. Hence the (slight) decrease in confidence observed in Type C, when compared to Type B cases, is entirely expected.

Before closing this discussion, there are two important objections to CCIP that we should directly address.

Recall that CCIP entails a particular distinction between the prototypes for complex expressions determined by FL and the subsequent pragmatically modified prototypes. Unlike more traditional views of the semantics-pragmatics divide, this distinction cuts across what speakers intuitively think is the literal meaning of expressions. We argued that the subjects in 'slow' experimental paradigms—such as those that ask them (under no time constraints) to list 'typical' features of complex NPs—often base their intuitions on pragmatically modified prototypes. Hence from the CCIP perspective, many influential studies advertised as being about language processing should be re-classified as studies about general reasoning.[19] But here's the rub. One might grant the conceptual coherence of this distinction and still object to its usefulness: for if CCIP is correct, doesn't it follow that we have no way of distinguish between the prototypes determined by FL and those subsequently affected by general pragmatic processing?

Fortunately, there are experimental paradigms which can be used to determine which features of the prototypes informally associated with complex expressions are determined by FL. For example, we can use cross-modal lexical priming studies[20] to see if fast processing of terms such as *pet fish* or *beach bi-*

[19]Take, for example, Johnson and Keil [80]'s famous studies in support of emergent features in conceptual combination. In one experiment, the typical features for complex expressions are determined as a function of how often they appeared in the feature-lists produced by subjects. The result is that complex expressions tend to have typical features which were also emergent features, in the sense that they do not appear in the lists of typical features of their constituents. As in most other such studies, subjects were given no time constraint to come-up with the feature-lists for simple and complex expressions. This same point applies to the influential studies presented in by Kunda et al [96] and Murphy [116]. From the perspective of CCIP, whatever this experimental paradigms are tracing, they are *not* cleanly picking out the features of the prototypes for complex expressions as determined by FL.

[20]See, for example, the cross-modal priming paradigm used by Swinney et al. [186] to argue against the view, defended by Springer and Murphy [184], that in some conceptual combinations such as *boiled celery* emergent features were activated prior to features of the constituents.

cycle primes some of their allegedly emergent features such as LIVES IN TANKS and WIDE TIRES, before processing the full sentences. If they do not, this supports the view that at least some of what are often taken as typical emergent features are computed post-linguistically. We can also use imaging techniques with high temporal resolution, such as EEG, to test whether emergent features are computed during language processing. For this aim, a useful event-related potential is the N400, which is known to inversely correlate with semantic expectations. Famously, sentences which predicate typical features of the subject such as *Dutch trains are yellow* generate less N400 activity at the predicate than sentences such as *Dutch trains are white/sour*, which do not predicate typical features. So we can use sentences that predicate emergent features of subject NPs (e.g., *Beach bicycles have wide tires*) and compare these to sentences that predicate features typical of *both* the head Ns and the complex NPs (e.g., *Beach bicycles have round tires*). If we get a significantly larger N400 when processing the predicates in the former cases this would suggest that emergent features are *not* usually computed during online processing. Finally, if projects along the lines of Pylkkanen [153] are eventually successful in isolating brain areas involved in purely semantic compositional processes, we might be able to use more location-sensitive imaging techniques to see if particular instances of phrasal processing are purely semantic or also involve pragmatic processes.

The second objection I want to address aims at the entire enterprise of accounting for prototype combinatorics. According to this objection, prototype theorists can avoid this whole conundrum by rejecting the 'uniformity assumption', namely, the idea that what is constitutive of lexical concepts—say, prototypes, conceptual roles, or asymmetric dependencies—should also be constitutive of complex concepts. This move was originally proposed by conceptual role theorists such as Horwich [76, 77] and Recanati [162] in response to Fodor's argument that conceptual roles are not compositional. Fodor's argument against the compositionality of conceptual roles is analogous to his argument against prototypes: the inference from *brown cows* to *dangerous* might be part of the conceptual role of *brown cows*, even if the inference to *dangerous* is not part of the conceptual role of either *brown* or *cows.* According to Horwich and Recanati, Fodor unjustifiably assumes that what is constitutive of lexical concepts should also be constitutive of complex concepts. In particular, Recanati argues that once conceptual roles determine the reference for lexical terms, all we need to say about complex terms is that their reference is determined compositionally: to know the meaning of *brown cows* is just to know that it is that concept whose extension is the intersection of the brown things and the cows. This response might seem especially attractive to prototype theorists who are also dual-content theorists. For if we hold that prototypes are one component of meaning, the other being reference determining cores, we can also hold that although simple concepts have prototypes, complex concepts do not: prototypes do not compose, only reference-determining conceptual components. Indeed, some theorists have proposed just this kind of response to save dual-content prototype theories from the emergent features challenge (see e.g., Margolis and Laurence [106]).

However, if we reject the uniformity assumption as a solution to the compositionality challenge, not only does prototype theory loose much of its explanatory power, but the claim that prototypes are conceptual components begins to seem like an unwarranted stipulation. In particular, one of the reasons why prototype theory is beginning to seem attractive to linguists and philosophers of language is because there are modifiers that seem to operate on the prototypes of nouns and verbs (Taylor [197], Sasson [176], Leslie [102], and chapter 4). These accounts are only available if we assume that prototypes are compositional components of meaning.

Although we cannot go into details here, a few examples will give the gist of these accounts. Take modifiers such as *typical* and *perfect*. Although a gun might not have been made to shoot, and an umbrella might not been made to cover one from the rain, intuitively a typical gun ought to have been made to shoot, and a typical umbrella to work as a cover for the rain. Similarly, although there are guns that might be lousy shooting instruments, and umbrellas that work badly as rain covers, perfect guns shoot well and perfect umbrellas are effective rain covers. In these cases, modifiers such as *typical* and *perfect* seem to operate on subcomponents of the prototypes of the head Ns. Other modifiers that seem to access the prototypes of Ns are privative Adjs such as *fake* and *counterfeit* (see chapter 4 above). In the same spirit, Leslie [102] recently argued that certain modifications of social-role concepts seem to access the associated prototypes. Take the case of *real*. Although friends can betray and husbands can be terrible providers, at least in some common idiolects of English, real friends don't betray and real husbands are the main providers for their families. Leslie's idea is that in complex expressions such as *real friends* and *real husbands*, the prototypes associated with the head Ns can be accessed by the modifiers. Importantly, these prototype modifiers can operate on complex NPs: consider e.g. *fake plastic gun*, *real tough husband*, *counterfeit gold Rolex* and *typical faithful priest*. It follows that these accounts presuppose that phrases such as *plastic gun*, *tough husband*, etc., also have prototypes.

The point then is this. These attractive accounts treat the modifiers as operators on prototypes. Since these operators also apply to complex NPs, the latter must also have associated prototypes. It follows that these accounts are only available if we hold that, in general, complex expressions also have prototypes. So prototype theory would loose much of its explanatory power if, as a 'solution' to the emergent features challenge, we drop the uniformity assumption. This is indeed one of the main reasons for trying to meet the challenge directly, even if we accept a dual-content theory, and to insist that the complex expressions have, at least in general, compositionally determined prototypes. CCIP provides linguists and philosophers of language with the framework to seriously explore accounts of modifiers such as *typical*, *real* and *fake* that treat them as operators on prototypes.

5.10 Conclusion

Recent discussions of prototype combinatorics agree on one thing: that prototypes are not strictly compositional. Critics insist that this is a decisive objection against the role of prototypes as conceptual components (Fodor [44, 47], Connolly et al. [30, 55]). Sympathizers respond that the objection's force is undercut by the result that non-compositional combinatorial procedures can be productive and systematic (Hampton and Jonsson [68], Prinz [144]). I argued that this response is based on a simplistic account of the reasons for accepting compositionality, and that its proponents adopt a non-compositional MDC that is too unconstrained. In addition, I argued that parties to the debate tend to fall on an problematic externalist way of individuating prototypes. As a result, the prototypes they assign to complex expressions seem to be compositionally underivable from those of the constituents.

The CCIP framework addressed these two shortcomings. There is something like an emerging consensus that we cannot defend a processing constraint that requires that the meaning of token complex expressions be computed from the standing, unmodified meaning of the parts (Pagin [122], Szabo [192], Recanati [164]). So most linguists and philosophers of language now agree that if FL is compositional, it is so in way that allows some degree of context-sensitive modulation. Furthermore, it is increasingly appreciated that prototypes are informationally rich and relatively flexible structures which often do *not* encode features that are 'typical' of their corresponding category (Barsalou [5], Taylor [197], Rosch [171]). Once these revised notions of compositionality and prototype individuation are properly implemented, the 'emergent features' argument against the compositionality of prototypes looses its force. More importantly, the main experimental evidence presented against the theory turns out to directly support it.

The idea that, at least for artifact and natural kind terms, linguistic meaning might include bits of general knowledge was originally defended by Putnam alongside his more influential views on the directly referential components. It is only a slight exaggeration to say that the compositionality objection basically eliminated any interest among semanticists and philosophers of language in developing the 'prototype/stereotype' component of Putnam's original theory of meaning. The recent effort by prototype theorists to address the challenge of emergent features by dropping compositionality has also proven to be an ineffective recruiting strategy. In contrast, CCIP allows semanticists and philosophers to keep a principle that, for various good reasons, most are unwilling to drop—i.e., some form of strict compositionality. At the same time, it opens space for a renewed exploration of richer theories of linguistic meaning and concepts, one more intimately informed by recent advances in cognitive psychology. That this integrated approaches can be extremely fruitful is evident from work such as Sasson [176] and Leslie [102].

Bibliography

[1] Nicolas Asher. *Lexical Meaning in Context: A Web of Words.* Cambridge University Press, Cambridge, UK, 2011.

[2] Giosue Baggio, Michel Van Limbalgen, and Peter Hagoort. The processing consequences of compositionality. In Markus Werning Wolfram Hinzen and Edouard Machery, editors, *The Oxford Handbook of Compositionality*, chapter 32, pages 656–672. Oxford University Press, New York, 2012.

[3] Mark C. Baker. *The Atoms of Language.* Basic Books, New York, 2001.

[4] J. R. Barclay, J. D. Bransford, J. J. Franks, N. S. McCarrell, and K. Nitsch. Comprehension and semantic flexibility. *Journal of Verbal Learning and Verbal Behavior*, 13(4):471–481, 1974.

[5] L. W. Barsalou. The instability of graded structure: Implications for the nature of concepts. In Neisser, editor, *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*, pages 101–140. Cambridge University Press, Cambridge, 1987.

[6] Laurence Barsalou. Ad hoc categories. *Memory & Cognition*, 11:211–27, 1983.

[7] Lawrence W. Barsalou. The instability of graded structure: Implications for the nature of concepts. *Concepts and conceptual development: Ecological and intellectual factors in categorization*, pages 101–140, 1987.

[8] Ned Block. Advertisement for a semantics for psychology. *Midwest Studies in Philosophy*, 10(1):615–78, 1986.

[9] Ned Block. Holism, hyper-analyticity and hyper-compositionality. *Philosophical Issues*, 3:37–72, 1993.

[10] Paul Bloom. Intention, history, and artifact concepts. *Cognition*, 60(1):1–29, 1996.

[11] Paul Bloom. Theories of artifact categorization. *Cognition*, 66(1):87–93, 1998.

[12] Paul Bloom. *How Children Learn the Meaning of Words*. The MIT Press, Cambridge, MA, 2002.

[13] Paul Bloom. Water as an artifact kind. In Eric Margolis and Stephen Laurence, editors, *Creations of the Mind: Theories of Artifacts and their Representation*, chapter 9, pages 150–156. Oxford University Press, Oxford, 2007.

[14] Emma Borg. *Minimal Semantics*. Oxford University Press, Oxford, 2004.

[15] Emma Borg. *Minimal Semantics*. Oxford University Press, Oxford, 2004.

[16] Emma Borg. Minimalism versus contextualism. In G. Preyer and G. Peter, editors, *Context-Sensitivity and Semantic Minimalism: New Essays on Semantics and Pragmatics*. Oxford University Press, Oxford, 2007.

[17] Tyler Burge. *Foundations of Mind*, volume 2. Oxford University Press, Oxford, 2007.

[18] Federica Busa and Pierrette Bouillon, editors. *The language of word meaning.* Cambridge University Press, Cambridge, 2001.

[19] Elizabeth Camp. On the generality constraint. *Philosophical Quaterly*, 54:209–231, Aug 2004.

[20] H. Cappelen and E. Lepore. Precis of insensitive semantics. *Philosophy and Phenomenological Research*, 73(2):425–434, 2006.

[21] Robyn Carston. Enrichment and loosening: complementary processes in deriving the proposition expressed. *Linguistische Berichte*, 8:103–127, 1997.

[22] Sergio E. Chaigneau. *Studies in the Conceptual Structure of Object Function.* PhD thesis, Emory University, 2002.

[23] Noam Chomsky. *Knowledge of Language: Its Nature, Origin, and Use.* Praeger Publishers, Connecticut, 1986.

[24] Noam Chomsky. Language and nature. *Mind*, 104, 1995.

[25] Noam Chomsky. *New Horizons in the Study of Mind Language.* Oxford University Press, Oxford, 2000.

[26] Noam Chomsky. Beyond explanatory adequacy. Technical report, MIT, Cambridge, MA, 2001.

[27] Noam Chomsky. Derivation by phase. In M. Kenstowitz, editor, *Ken Hale: A Life in Language*, pages 1–52. MIT Press, Cambridge, MA, 2001.

[28] Noam Chomsky. Reply to gopnik. In Louise M. Antony and Norbert Hornstein, editors, *Chomsky and his Critics*, chapter 11, pages 316–325. Blackwell Publishing Ltd, 2003.

[29] John Collins. Horwich's schemata meet syntactic structures. *Mind*, 112(447):399–431, July 2003.

[30] A. C. Connolly, J. Fodor, L. Gleitman, and H. Gleitman. Why stereotypes don't even make good defaults. *Cognition*, 103(1):1–22, 2007.

[31] V. J. Cook and Mark Newson. *Chomsky's Universal Grammar: An Introduction*. Blackwell Publishing Ltd, Oxford, 2007.

[32] Seana Coulson and Gilles Fauconnier. Fake guns and stone lions: Conceptual blending and privative adjectives. In B. Fox, D. Jurafsky, and L. Michaels, editors, *Cognition and Function in Language*. CSLI, Palo Alto, CA, 1999.

[33] Seana Coulson and Gilles Fauconnier. Fake guns and stone lions: Conceptual blending and privative adjectives. In B. Fox, D. Jurafsky, and L. Michaels, editors, *Cognition and Function in Language*. CSLI, Palo Alto, CA, 1999.

[34] D. A. Cruse. *Lexical Semantics*. Cambridge University Press, Cambridge, UK, 1986.

[35] Martin Davies. Tacit knowledge and semantic theory: can a five per cent difference matter? *Mind*, 96:441–462, 1987.

[36] Michael Dawson. *Understanding Cognitive Science*. Blackwell Publishers Ltd, Oxford, 2001.

[37] Josh Dever. Compositionality. In Ernest Lepore and Barry Smith, editors, *Handbook of Philosophy of Language*. Oxford University Press, Oxford, 2006.

[38] Josh Dever. Compositionality. In Ernest Lepore and Barry Smith, editors, *The Oxford Handbook of Compositionality.* Oxford University Press, Oxford, 2006.

[39] Michael Dummett. *The Seas of Language.* Oxford University Press, Oxford, 1993.

[40] J. Elman, E. Bates, M. Johnson, and A. Karmiloff-Smith. *Rethinking Innatness: A Connectionist Perspective on Development.* MIT Press, Cambridge, MA, 1996.

[41] Gareth Evans. *Collected Papers.* Oxford University Press, Oxford, 1985.

[42] Jerry Fodor. *Concepts: Where Cognitive Science Went Wrong.* Oxford University Press, Oxford, 1998.

[43] Jerry Fodor. *Concepts: Where Cognitive Science Went Wrong.* Oxford University Press, Oxford, 1998.

[44] Jerry Fodor. *Concepts: where cognitive science went wrong.* Oxford University Press, New York, 1999.

[45] Jerry Fodor. Language, thought and compositionality. *Mind & Language*, 16:1–15, February 2001.

[46] Jerry Fodor. *LOT 2: The Language of Thought Revisited.* Oxford University Press, Oxford, 2008.

[47] Jerry Fodor and Ernest Lepore. The pet fish and the red herring: why concepts still can't be prototypes. *Cognition*, 58(2):329–43, 1996.

[48] Jerry Fodor and Ernest Lepore. *The Compositionality Papers.* Oxford University Press, Oxford, 2002.

[49] Jerry Fodor and Ernest Lepore. *The Compositionality Papers.* Oxford University Press, Oxford, 2002.

[50] Jerry Fodor and Ernest Lepore. *The Compositionality Papers.* Oxford University Press, Oxford, 2002.

[51] Jerry Fodor and Zenon Pylyshyn. Connectionism and cognitive architecture: A critical analysis. Technical report, Rutgers Center for Cognitive Science, 1988.

[52] Jerry Fodor and Zenon Pylyshyn. Connectionism and cognitive architecture: A critical analysis. Technical report, Rutgers Center for Cognitive Science, 1988.

[53] Bradley Franks. Sense generation a quasi-classical approach to concepts and concept combination. *Cognitive Science*, 19:441–505, Jan 1995.

[54] Bradley Franks. Sense generation: A quasi-classical approach to concepts and concept combination. *Cognitive Science*, 19(4):441–505, 1995.

[55] Lila R. Gleitman, Andrew C. Connolly, and Sharon Lee Armstrong. Can prototype representations support composition and decompositions? In Markus Werning, Wolfram Hinzen, and Edouard Machery, editors, *The Oxford Handbook of Compositionality*, chapter 20. Oxford University Press, Oxford, 2012.

[56] Sam Glucksberg. *Understanding Figurative Language: From Metaphors to Idioms.* Oxford University Press, New York, 2001.

[57] Sam Glucksberg. The psycholinguistics of metaphor. *Trends in Cognitive Sciences*, 7(2):92–96, 2003.

[58] Adele E. Goldberg. *Constructions at Work: The Nature of Generalizations in Language.* Oxford University Press, Oxford, 2006.

[59] Adele E. Goldberg. Constructionist approaches. In Thomas Hoffman and Graeme Trousdale, editors, *The Oxford Handbook of Construction Grammar*, chapter 2. Oxford University Press, Oxford, 2013.

[60] Adele E. Goldberg. Explanation and constructions. *Mind & Language*, forthcoming.

[61] Alison Gopnik. The theory theory as an alternative to the innatness hypothesis. In Louise M. Antony and Norbert Hornstein, editors, *Chomsky and his critics*, chapter 10, pages 238–255. Blackwell Publishing Ltd, Oxford, 2003.

[62] Ruth Hamill, Richard E. Nisbett, and Timothy DeCamp Wilson. Insensitivity to sample bias: generalizing from atypical cases. *Journal of Personality and Social Psychology*, 39(4):578–589, 1980.

[63] JA Hampton. Emergent attributes in combined concepts. *Creative thought: An investigation of conceptual structures and processes*, pages 83–110, 1997.

[64] James Hampton. Conceptual combination. In K. Lamberts and D. Shanks, editors, *Knowledge, Concepts, and Categories*, pages 133–60. The MIT Press, Cambridge, MA, 1997.

[65] James Hampton. Concepts as prototypes. In B. H. Ross, editor, *The psychology of learning and motivation: advances in research and theory*, volume 46, pages 79–113. Elsevier, 2006.

[66] James A. Hampton. Conceptual combination and fuzzy logic. In Radim Belohlavek and George J. Klir, editors, *Concepts and Fuzzy Logic*, chapter 8. The MIT Press, Cambridge, MA, 2011.

[67] James A. Hampton and Martin L. Jonsson. Typicality and compositionality: The logic of combining vague concepts. In Marcus Werning, Wolfram Hinzen, and Edouard Machery, editors, *The Oxford Handbook of Compositionality*, chapter 18, pages 385–402. Oxford University Press, New York, 2012.

[68] James A. Hampton and Martin L. Jonsson. Typicality and compositionality: the logic of combining vague concepts. In Markus Werning, Wolfram Hinzen, and Edouard Machery, editors, *The Oxford Handbook of Compositionality*, chapter 18. Oxford University Press, Oxford, 2012.

[69] Heidi Harley. Lexical decomposition in modern syntactic theory. In Markus Werning, Wolfram Hinzen, and Edouard Machery, editors, *Oxford Handbook of Compositionality*, chapter 15. Oxford University Press, Oxford, 2012.

[70] Irene Heim and Angelina Kratzer. *Semantics in Generative Grammar*. Blackwell Publishers Ltd, Oxford, 1998.

[71] Irene Heim and Angelina Kratzer. *Semantics in Generative Grammar*. Blackwell Publishers Ltd, Oxford, 1998.

[72] James Higginbotham. Languages and idiolects: their language and ours. In Ernest Lepore and Barry Smith, editors, *The Oxford Handbook of the Philosophy of Language*. Oxford University Press, Oxford.

[73] Wolfram Hinzen. Syntax in the atom. In Markus Werning, Wolfram Hinzen, and Edouard Machery, editors, *Oxford Handbook of Compositionality*, chapter 16. Oxford University Press, Oxford, 2012.

[74] Thomas Hoffman and Graeme Trousdale, editors. *The Oxford Handbook of Construction Grammar*. Oxford University Press, Oxford, 2013.

[75] Lotte Hogeweg. Rich lexical representations and conflicting features. *International Review of Pragmatics*, 4:209–231, 2012.

[76] Paul Horwich. Concept constitution. *Philosophical Issues*, 9:15–19, 1998.

[77] Paul Horwich. *Meaning*. Oxford University Press, Oxford, 2004.

[78] Michael Israel. Literally speaking. *Journal of Pragmatics*, 34:423–432, 2002.

[79] Pauline Jacobson. Direct compositionality. In Wolfram Hinzen Marcus Werning and Edouard Machery, editors, *The Oxford Handbook of Compositionality*, chapter 5, pages 109–28. Oxford University Press, Oxford, 2012.

[80] Christine Johnson and Frank Keil. Explanatory understanding and conceptual combination. In F.C. Keil and R.A. Wilson, editors, *Explanation and Cognition*, pages 327–359. The MIT Press, Cambridge, 2000.

[81] Kent Johnson. On the systematicity of language and thought. *Journal of Philosophy*, 101(3):111–139, 2004.

[82] Marrtin L. Jonsson. *On Compositionality: Doubs on the Structural Path to Meaning*. PhD thesis, Lund University, 2008.

[83] Martin L. Jonsson. *On Compositionality: doubts on the structural path to meaning*. PhD thesis, Lund University, 2008.

[84] Martin L. Jonsson and James A. Hampton. On prototypes as defaults (comments on connolly, fodor, gleitman and gleitman 2007). *Cognition*, 106:913–23, 2008.

[85] Frank C. Keil, Marissa L. Greif, and Rebekkah S. Kerner. A world apart: how concepts of the constructed world are different in representation and

in development. In Eric Margolis and Stephen Laurence, editors, *Creations of the mind: theories of artifacts and their representation*, pages 231–248. Oxford University Press, Oxford, 2007.

[86] Deborah Keleman and Susan Carey. The essence of artifacts: developing the design stance. In Eric Margolis and Stephen Laurence, editors, *Creations of the mind: theories of artifacts and their representation*, chapter 12, pages 212–230. Oxford University Press, Oxford, 2007.

[87] Perfors A. Kemp, C. and J. B. Tenenbaum. Learning overhypotheses. In *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society.*, 2006.

[88] C. Kennedy and B. Levin. Measure of change: The adjectival core of degree achievements. In Louis McNally and Christopher Kennedy, editors, *Adjectives and Adverbs: Syntax, Semantics, and Discourse*. Oxford University Press, 2008.

[89] Chistopher Kennedy and Louise McNally. Color, context, and compositionality. *Synthese*, 174:79–98, 2010.

[90] Joshua Knobe, Sandeed Prasada, and George E. Newman. Dual character concepts and the normative dimension of conceptual representation. *Cognition*, 127:242–257, 2013.

[91] Bernard W. Kobes. Semantics and psychological prototypes. *Pacific Philosophical Quarterly*, 70(1):1–18, 1989.

[92] Hilary Kornblith. How to refer to artifacts. In Eric Margolis and Stephen Laurence, editors, *Creations of the Mind: Theories of Artifacts and their Representation*, chapter 8, pages 138–149. Oxford University Press, Oxford, 2007.

[93] Angelika Kratzer. *Modals and Conditionals*. Oxford University Press, Oxford, 2012.

[94] Manfred Krifka, Francis Jeffry Pelletier, Gregory N. Carlson, Alice ter Muelen, Gennaro Chierchia, and Godehard Link. Genericity: An introduction. In Gregory N. Carlson and Francis J. Pelletier, editors, *The Generic Book*, chapter 1, pages 1–124. The University of Chicago Press, Chicago, 1995.

[95] Saul Kripke. *Naming and Necessity*. Harvard University Press, Cambridge, MA, 1980.

[96] Ziva Kunda, Dale T. Miller, and Theresa Claire. Combining social concepts: The role of causal reasoning. *Cognitive Science*, 14:551–577, 1990.

[97] Ran Lahav. Against compositionality: the case of adjectives. *Philosophical Studies*, 57:261–279, 1989.

[98] George Lakoff and Mark Johnson. *The metaphors we live by*. The University of Chicago Press, Chicago, 1980.

[99] George Lakoff and Mark Johnson. *Metaphors we live by*. The University of Chicago Press, Chicago, IL, 1980.

[100] Richard Larson and Gabriel Segal. *Knowledge of Meaning*. The MIT Press, Cambridge, 1995.

[101] Richard K. Larson and Gabriel M. A. Segal. *Knowledge of Meaning*. The MIT Press, Cambridge, Massachusetts, 1995.

[102] Sarah Jane Leslie. "Hilllary Clinton is the only man in the Obama administration": Dual character concepts, generics and gender. *Analytic Philosophy*, page forthcoming, 2014.

[103] Peter Ludlow. *The Philosophy of Generative Grammar*. Oxford University Press, Oxford, 2011.

[104] Edouard Machery. Concepts: A tutorial. In Radim Belohlavek and George J. Klir, editors, *Concepts and Fuzzy Logic*, chapter 2. The MIT Press, Cambridge, Massachusetts, 2011.

[105] Barbara C. Malt and Steven A. Sloman. Artifact categorization: The good, the bad, and the ugly. In Eric Margolis and Stephen Laurence, editors, *Creations of the Mind: Theories of Artifacts and their Representation*, chapter 6, pages 85–123. Oxford University Press, Oxford, 2007.

[106] E. Margolis and S. Laurence. Concepts and cognitive science. In E. Margolis and S. Laurence, editors, *Concepts: core readings*, chapter 1, pages 3–82. The MIT Press, Cambridge, Massachusetts, 1999.

[107] Eric Margolis. How to acquire a concept. *Mind & Language*, 13(3):347–369, 1998.

[108] Eric Margolis and Stephen Laurence. *Creations of the Mind: Theories of Artifacts and their Representation*. Oxford University Press, Oxford, 2007.

[109] Adee Matan and Susan Carey. Developmental changes within the core of artifact concepts. *Cognition*, 78(1):1–26, 2001.

[110] Brian McElree, Gregory L. Murphy, and Tamara Ochoa. Time course of retrieving conceptual information: a speed-accuracy trade-off study. *Psychonomic Bulletin & Review*, 13(5):848–853, 2006.

[111] M.S. McGlone, S. Glucksberg, and C. Cacciari. Semantic productivity and idiom comprehension. *Discourse Processes*, 17:167–190, 1994.

[112] Louise McNally. Lexical representation and modification within the noun phrase. *Recherches Linguistiques de Vincennes*, 34:191–206, 2006.

[113] Julius Moravcsik. *Meaning, Creativity, and the Partial Inscrutibility of the Human Mind*. CSLI Publications, California, 1998.

[114] Marcin Morzycki. *Modification*. Cambridge University Press, Cambridge, UK, forthcoming.

[115] G. L. Murphy and J. M. Andrew. The conceptual basis of antonymy and synonymy in adjectives. *Journal of Memory and Language*, 32:301–319, 1993.

[116] Gregory L. Murphy. Noun phrase interpretation and conceptual combination. *Journal of Memory and Language*, 29:259–288, 1990.

[117] Gregory L. Murphy. *The Big Book of Concepts*. MIT Press, Cambridge, MA, 2002.

[118] Gregory L. Murphy. *The big book of concepts*. The MIT Press, Cambridge, Massachusetts, 2002.

[119] Geoffrey Nunberg, Ivan A. Sag, and Thomas Wasow. Idioms. *Language*, 70(3):491–538, 1994.

[120] Peter Pagin. Communication and the complexity of semantics. In Markus Werning, Wolfram Hinzen, and Edouard Machery, editors, *The Oxford Handbook of Compositionality*, chapter 25. Oxford University Press, Oxford, 2012.

[121] Peter Pagin and Jeffrey Pelletier. Content, context, and composition. *Context-Sensitivity and Semantic Minimalism: New Essays on Semantics and Pragmatics*, page 25, 2007.

[122] Peter Pagin and Jeffrey Pelletier. Content, context, and composition. *Context-Sensitivity and Semantic Minimalism: New Essays on Semantics and Pragmatics*, page 25, 2007.

[123] Peter Pagin and Dag Westerstahl. Compositionality. October 2008.

[124] Barbara Partee. Lexical semantics and compositionality. In L. Gleitman and M. Liberman, editors, *An Invitation to Cognitive Science: Volume 1*, volume 1. The MIT Press, 1998.

[125] Barbara Partee. *Compositionality in Formal Semantics*. Blackwell Publishers Ltd, Oxford, 2004.

[126] Barbara Partee. Compositionality and coercion in semantics: The dynamics of adjective meaning. In Gerlouf Bouma, Irene Kraemer, and Joost Zwrts, editors, *Proceedings of the Colloquium "Cognitive Foundations of Interpretation"*, Amsterdam, 2007. Royal Netherlands Academy of Arts and Sciences.

[127] Barbara Partee. Compositionality and coercion in semantics: The dynamics of adjective meaning. In Gerlouf Bouma, Irene Kraemer, and Joost Zwrts, editors, *Proceedings of the Colloquium "Cognitive Foundations of Interpretation"*, Amsterdam, 2007. Royal Netherlands Academy of Arts and Sciences.

[128] Barbara Partee. Privative: Subsectives plus coercion? In R. Baule, U. Reyle, and T.E. Zimmerman, editors, *Presupposition and Discourse: Essays Offered to Hans Kamp*. Elsevier, Amsterdam, 2010.

[129] Barbara Partee. Privative: Subsectives plus coercion? In R. Baule, U. Reyle, and T.E. Zimmerman, editors, *Presupposition and Discourse: Essays Offered to Hans Kamp*. Elsevier, Amsterdam, 2010.

[130] Barbara Partee and Hans Kamp. Prototype theory and compositionality. *Cognition*, 57:129–191, 1995.

[131] Barbara Partee and Hans Kamp. Prototype theory and compositionality. *Cognition*, 57(2):129–191, 1995.

[132] Amy Perfors, Joshua B. Tenenbaum, and Terry Regier. Poverty of stimulus? a rational approach. *Cognition*, 118:306–338, 2011.

[133] S. T. Piantadosi, N. D. Goodman, B. A. Ellis, and J. B. Tenenbaum. A bayesian model of the acquisition of compositional semantics. In *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society*, 2008.

[134] Paul Pietroski. The character of natural language semantics. In A. Barber, editor, *Epistemology of Language*. Oxford University Press, 2003.

[135] Paul Pietroski. Minimalist meaning, internalist interpretation. *BIOLINGUISTICS*, 2(4), 2008.

[136] Paul Pietroski. Semantic monadicity and conceptual polyadicity. In Markus Werning, Wolfram Hinzen, and Edouard Machery, editors, *The Oxford Handbook of Compositionality*, chapter 6. Oxford University Press, Oxford, 2012.

[137] Paul Pietroski. Semantic monadicity with conceptual polyadicity. In Marcus Werning, Wolfram Hinzen, and Edouard Machery, editors, *The Oxford Handbook of Compositionality*, chapter 6, pages 129–148. Oxford University Press, Oxford, 2012.

[138] Guillermo Del Pinal. *The Architecture of the Faculty of Language: Compositional Operations and Complex Lexical Representations.* PhD thesis,

Columbia University in the City of New York, New York, forthcoming 2013.

[139] Steven Pinker. Language acquisition. In *An Inivitation to Cognitive Science*, volume 1, chapter 6. The MIT Press, 1995.

[140] Steven Pinker. *The Stuff of Thought: Language as a Window into Human Nature*. the Penguin Group, New York, 2007.

[141] Steven Pinker. *The Stuff of Thought: Language as a Window into Human Nature*. the Penguin Group, New York, 2007.

[142] J. J. Prinz. *Furnishing the mind: concepts and their perceptual basis*. MIT Press, Cambridge, Massachusetts, 2002.

[143] Jesse Prinz. *Furnishing the Mind*. MIT Press, Cambridge, MA, 2002.

[144] Jesse Prinz. Regaining composure: a defense of prototype compositionality. In Markus Werning, Wolfram Hinzen, and Edouard Machery, editors, *The Oxford Handbook of Compositionality*, chapter 21. Oxford University Press, Oxford, 2012.

[145] Jesse Prinz. Regaining compusure: A defense of prototype compositionality. In Markus Werning; Wolfram Hinzen; Edouard Machery, editor, *Oxford Handbook of Compositionality*, chapter 21, pages 437–453. Oxford University Press, Oxford, 2012.

[146] Geoffrey K. Pullum and Barbara C. Scholz. Empirical assessment of stimulus poverty arguments. *The Linguistic Review*, 19(9-50), 2002.

[147] Geoffrey K. Pullum and Barbara C. Scholz. Recursion and the infinitude claim. In Harry van der Hulst, editor, *Recursion and Human Language*, chapter 5, pages 113–139. De Gruyter Mouton, Germany, 2010.

[148] James Pustejovsky. *The Generative Lexicon.* The MIT Press, Cambridge, 1995.

[149] James Pustejovsky. *The Generative Lexicon.* The MIT Press, Cambridge, MA, 1995.

[150] James Pustejovsky. Co-compositionality in grammar. In Markus Werning, Wolfram Hinzen, and Edouard Machery, editors, *Oxford Handbook of Compositionality*, chapter 17. Oxford University Press, Oxford, 2012.

[151] James Pustejovsky. Co-compositionality in grammar. In Markus Werning, Wolfram Hinzen, and Edouard Machery, editors, *Oxford Handbook of Compositionality*, chapter 17, pages 371–384. Oxford University Press, Oxford, 2012.

[152] Hillary Putnam. Is semantics posible? *Metaphilosophy*, 1(3):187–201, 1970.

[153] Liina Pylkkanen, Jonathan Brennan, and Douglas K. Bemis. Grounding the cognitive neuroscience of semantics in linguistic theory. *Language and Cognitive Processes*, 26(9):1317–1337, 2011.

[154] Liina Pylkkanen, Jonathan Brennan, and Douglas K. Bemis. Grounding the cognitive neuroscience of semantics in linguistic theory. *Language and Cognitive Processes*, 26(9):1317–1337, 2013.

[155] Liina Pylkkanen and Brian McElree. The syntax-semantic interface: Online composition of sentence meaning. In Mathew J. Traxler and Morton Ann Bersbacher, editors, *Handbook of Psycholinguistics*, pages 537–577. Elsevier, New York, NY, 2006.

[156] Zenon Pylyshyn. *Computation and Cognition: Towards a Foundation for Cognitive Science.* MIT Press, Cambridge, 1986.

[157] Zenon Pylyshyn. The role of cognitive architecture in theories of cognition. *Architectures for intelligence: the twenty-second Carnegie Mellon Symposium on Cognition*, 1991.

[158] B. Yankama N. Chomsky R. Berwick, P. Pietroski. Poverty of the stimulus revisited. *Cognitive Science*, 35:1207–1242, 2011.

[159] Andrew Radford. *Minimalist Syntax: Exploring the Structure of English*. Cambridge University Press, Cambridge, 2004.

[160] François Recanati. *Literal Meaning*. Cambridge University Press, Cambridge, 2004.

[161] Francois Recanati. *Literal Meaning*. Cambridge University Press, Cambridge, 2004.

[162] Francois Recanati. Indexical concepts. In Manuel Garcia-Carpintero and Josep Macia, editors, *Two-Dimensional Semantics*. Claredon Press, Oxford, 2006.

[163] Francois Recanati. *Truth-Conditional Pragmatics*. Oxford University Press, Oxford, 2010.

[164] Francois Recanati. *Truth-conditional pragmatics*. Oxford University Press, Oxford, 2010.

[165] Georges Rey. Concepts and stereotypes. *Cognition*, 15(1):237–262, 1983.

[166] Lance J. Rips. Similarity, typicality, and categorization. In S. Vosniadou and A. Ortony, editors, *Similarity and Analogical Reasoning*, pages 21–59. Cambridge University Press, Cambridge, 1989.

[167] P. Robbins. How to blunt the sword of compositionality. *Nous*, 36:313–334, 2002.

[168] Phillip Robbins. Minimal semantics and modularity. In G. Preyer and G. Peter, editors, *Context Sensitivity and Semantic Minimalism: New Essays on Semantics and Pragmatics*. Oxford University Press, 2007.

[169] Tom Roeper. The acquisition of recursion:how formalism articulates the child's path. *Biolinguistics*, 5(57-82), 2011.

[170] E. Rosch. Natural categories. *Cognitive Psychology*, 4:532–547, 1973.

[171] Eleanor H. Rosch. Slow lettuce: Categories, concepts, fuzzy sets, and logical deduction. In Radim Belohlavek and George J. Klir, editors, *Concepts and Fuzzy Logic*, chapter 4. The MIT Press, Cambridge, Massachusetts, 2011.

[172] E. M. Roth and E. J. Shoben. The effect of context on the structure of categories. *Cognitive Psychology*, 15:346–378, 1983.

[173] Daniel Rothschild and Gabriel Segal. Indexical predicates. *Mind and Language*, 24:467–493, 2009.

[174] Daniel Rothschild and Gabriel Segal. Indexical predicates. *Mind & Language*, 24(4):467–493, 2009.

[175] Galit W. Sassoon. Adjectival vs. nominal categorization processes. *Belgian Journal of Linguistics*, 25:104–147, 2011.

[176] Galit W. Sassoon. Adjectival vs. nominal categorization processes: The rule vs. similarity hypothesis. *Belgium Journal of Linguistics*, 24:104–147, 2011.

[177] Barbara C. Scholz and Geoffrey K. Pullum. Systematicity and natural language syntax. *Croation Journal of Philosophy*, 21:375–402, 2007.

[178] Gerhard Schurz. Prototypes and their composition from an evolutionary point of view. In Marcus Werning, Wolfram Hinzen, and Edouard Machery, editors, *The Oxford Handbook of Compositionality*, chapter 26. Oxford University Press, Oxford, 2012.

[179] Stephen C. Sloman and Barbara C. Malt. Artifacts are not ascribed essences, nor are they treated as belonging to kinds. *Language and Cognitive Processes*, 18(5-6):1–20, 2003.

[180] E. E. Smith, D. L. Medin, L. J. Rips, and M. Keane. Combining prototypes: a selective modification model. *Cognitive Science*, 12:485–527, 1988.

[181] Kenny Smith and Simon Kirby. Compositionality and linguistic evolution. In Wolfram Hinzen Markus Werning and Edouard Machery, editors, *The Oxford Handbook of Compositionality*, chapter 25. Oxford University Press, Oxford, 2012.

[182] Neil Smith. Introduction: Syntax symposium. *Mind & Language*, 28:377–391, September 2013.

[183] Dan Sperber and Ira A. Noveck, editors. *Experimental Pragmatics*. Palgrave Macmillan, New York, NY, 2004.

[184] K Springer and G. L. Murphy. Feature availability in conceptual combination. *Psychological Science*, 3:111–117, 1992.

[185] Jason Stanley. Context and logical form. *Linguist and Philosophy*, 23(4):391–434, 2000.

[186] David Swinney, Tracey Love, Mathew Walenski, and Edward E. Smith. Conceptual combination during sentence processing. *Psychological Science*, 18(5):397–400, 2007.

[187] Zoltan Szabó. *Problems of Compositionality*. Garland Publisher Inc, New York and London, 2000.

[188] Zoltan Szabó. Adjectives in context. In R. Harnish and I. Kenesei, editors, *Perspectives on Semantics, Pragmatics, and Discourse*, pages 119–46. John Benjamins, Amsterdam, 2001.

[189] Zoltan Szabó. Review of the compositionality papers. *Mind*, 113:340–344, 2004.

[190] Zoltan Szabó. Compositionality. *Stanford Encyclopedia of Philosophy*, 2007.

[191] Zoltan Szabó. Structure and conventions. *Philosophical Studies*, 137(3), 2008.

[192] Zoltan Szabó. The determination of content. *Philosophical Studies*, 148:253 – 272, 2010.

[193] Zoltan Szabó. The case for compositionality. In Markus Werning, Wolfram Hinzen, and Edouard Machery, editors, *The Oxford Handbook of Compositionality*, chapter 3. Oxford University Press, Oxford, 2012.

[194] Zoltan Szabo. The case for compositionality. In Markus Werning, Wolfram Hinzen, and Edouard Machery, editors, *The Oxford Handbook of Compositionality*, chapter 3. Oxford University Press, Oxford, 2012.

[195] Zoltan Szabó and Jason Stanley. On quantifier domain restriction. *Mind & Language*, 15(2-3):219–61, 2000.

[196] P. Tabossi and F. Zaron. The activation of idiomatic meaning. In Schenk Everaert, Van den Linden and Schreuder., editors, *Idioms: Processing, Structure, and Interpretation*, pages 145–163. Hillsdale, New Jersey, 1995.

[197] John R. Taylor. *Linguistic Categorization.* Oxford University Press, Oxford, 2009.

[198] Amie L. Thomasson. Realism and human kinds. *Philosophy and Phenomenological Research*, 67(3):580–609, 2003.

[199] Amie L. Thomasson. Artifacts and human concepts. In Eric Margolis and Stephen Laurence, editors, *Creations of the Mind: Theories of Artifacts and their Representation*, chapter 4, pages 52–73. Oxford University Press, Oxford, 2007.

[200] Michael Tomasello. *Constructing a Language: a Usage-Based Theory of Language Acquisition.* Harvard University Press, Cambridge, MA, 2003.

[201] Charles Travis. On what is strictly speaking true. *Canadian Journal of Philosophy*, 15(2):187–229, 1985.

[202] Charles Travis. On constraints on generality. *Proceedings of the Aristotelian Society*, 44:165–188, 1994.

[203] Charles Travis. On constraints on generality. *Proceedings of the Aristotelian Society*, 94:165–168, 1994.

[204] Charles Travis. Insensitive semantics. *Mind & Language*, 21(1):39–49, 2006.

[205] Bob van Tiel. Embedded scalars and typicality. *Journal of semantics*, 31:147–177, 2014.

[206] Carl Vikner and Per Anker Jensen. A semantic analysis of the english genitive. interacion of lexical and formal semantics. *Studia Linguistica*, 56(2):191–226, 2002.

[207] Carl Vikner and Per Anker Jensen. A semantic analysis of the English genitive. Interacion of lexical and formal semantics. *Studia Linguistica*, 56(2):191–226, 2002.

[208] Daniel Weiskopf. Compound nominals, context, and compositionality. *Synthese*, 156(1):161–204, 2007.

[209] Daniel Weiskopf. Atomism, pluralism, and conceptual content. *Philosophy and Phenomenological Research*, LXXIX(1):131–163, 2009.

[210] Deidere Wilson and Robyn Carston. A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In Noel Burton-Roberts, editor, *Pragmatics*, pages 230–259. Palgrave, London, 2007.

[211] Dieter Wunderlich. Lexical decomposition in grammar. In Marcus Werning, Wolfram Hinzen, and Edouard Machery, editors, *Oxford Handbook of Compositionality*, chapter 14. Oxford University Press, Oxford, 2012.

[212] Fei Xu. Rational statistical inference and cognitive development. In Stephen Laurence Peter Carruthers and Stephen Stich, editors, *The innate mind: Foundations and the future*, volume 3, chapter 10, pages 199–215. Oxford University Press, New York, 2007.