# LUND UNIVERSITY

**On Compositionality**

Jönsson, Martin

2008

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

# On Compositionality

Doubts about the Structural Path to Meaning

Martin L. Jönsson

Compositionality: Doubts about the Structural Path to Meaning
Martin L. Jönsson

*For Pernilla*

# Preface

What follows is my Ph.D. thesis in Philosophy. I wrote it mostly while at the philosophy department at Lund University, and, discounting time spent teaching and on leave abroad, I have worked on it for just over four years. The finished thesis is, of course, only the end result of my work. It therefore contains only hints of the process leading to its completion, and of the intellectual debts I have incurred while working on it. To rectify this to some small degree, I want to take this opportunity to briefly describe some of preceding events that were of particular importance to me and to acknowledge some of the places where my debts lie. But before I do this I want to make a quick but important remark on the methodology and aim of the thesis.

After having initially reviewed some of the literature on compositionality I was struck by the fact that it was fairly compartmentalized; areas that seemed clearly related weren't discussed together. For instance, the discussion in formal semantics of the triviality of compositionality seems to be connected with the discussion in philosophy of what compositionality could do, but the two debates were seldom covered by the same work; and investigations into what compositionality could do were rarely coupled with in-depth discussion of the reason for holding compositionality true and for rejecting alternatives. In this thesis I have tried to overcome this compartmentalization and have attempted to integrate several strands of research in order to answer the main research question as satisfactorily as possible. I'm hoping that the juxtaposition and integration that feature in the thesis resulted in the intended cross-fertilization rather than – as is always the risk when one attempts to combine different forms of research – forcing one or several discussions into molds that they do not quite

fit. However, my integrative approach might invite reading the thesis in a way that was not intended, and I would like to explicitly discourage this reading. Since I describe and discuss theories in several other research traditions than the philosophical one, a reader might have the impression that the following work is an attempt to work in these traditions. It is not. The methodology and the results presented here are not those of (say) formal semantics or cognitive psychology, and this thesis is not a work in either tradition. I'm hoping that the results will be of *methodological* importance to these traditions, but I have not attempted to directly contribute to the *subject matter* of these disciplines. The work, both in its conclusions and its methodology, is philosophical through and through. The theories from other disciplines and traditions that I consider are often versions that I have simplified in order to make the philosophical points I want to make as clear as possible. The project might be classified as philosophy of language, but it might also be classified as philosophy of linguistic science (as a subspecies of the philosophy of science). In any case, it would be erroneous to classify it as a work *in* one of the linguistic sciences.

My first attempt at independent work in philosophy also dealt with composition. My bachelor's thesis in theoretical philosophy, written in 2000, dealt with the regress that Francis Bradley had suggested threatened any metaphysical account countenancing complex entities. This convergence on the topic of composition seems accidental in retrospect. I have at no point been aware of being especially interested in matters pertaining to composition that could have acted as a common cause in my choice of topics, and it seems to me that I pursued the two projects for quite different reasons. The link between the two projects that does exist derives instead from the encouragement I received, after finishing my work on the Bradley regress, in particular from Johannes Persson, Nils-Eric Sahlin, Marcus Ivarsson and Anna-Sofia Maurin, which acted as a part of the motivation for pursuing additional projects in philosophy, and for which – even though it was given a very long time ago – I'm still very grateful.

I was first notified that I had been accepted as a Ph.D. student in the spring of 2002, and it was mutually agreed that I would start working at the start of September that year. I spent much of the summer practicing capoeira in Brazil on a very memorable, but rather

unphilosophical trip, and I returned with only the vaguest ideas about what I wanted to pursue as a research topic. Since I wrote my bachelor's thesis in philosophy, I had developed an interest in questions concerning meaning and concepts, and I had worked on the semantics of thought and psychological theories of categorization as part of additional undergraduate theses in philosophy and cognitive science. As I read up on the literature on meaning and concepts I became aware that compositionality was often assumed in order to make headway in inquiries. However, in the literature I came across, it seemed to me somewhat unclear not only what compositionality amounted to, but also what, exactly, licensed its use it as a constraint on semantic theorizing. Shortly after I had begun looking into it more carefully, I had the opportunity to discuss some of the issues with Jerry Fodor – a leading researcher in the field and someone whose writings had spurred my interest in the topic. We talked at a cognitive science symposium in Stockholm in the summer of 2003, and he encouraged me to pursue compositionality as the topic in my thesis.

A few months later, in October, I was fortunate enough to attend Dag Westerståhl's talk 'What we now know about compositionality' at a conference in honor of Peter Pagin on his fiftieth birthday. Professor Westerståhl's talk, and the tradition he drew attention to, provided a firm foundation on which matters pertaining to compositionality could be discussed, and this influenced my thinking from then on. Professor Pagin's own very interesting work on compositionality, in particular his 'Communication and Strong Compositionality', also shaped my thinking substantially. I'm grateful to both men for writing such wonderful things and for being so helpful in clarifying them in conversation.

Professor Pagin's and Professor Westerståhl's work led me to the work of Wilfrid Hodges, who had developed a very influential framework they both made use of. My contact with Professor Hodges has also been very beneficial for me and I'm grateful for his patience and helpfulness in explaining some of the intricacies of the systems he has developed and for taking the time to read and comment on the initial chapters of the thesis.

My investigation of supposed problems with non-compositional accounts of meaning lead me to contact James Hampton and propose that we together investigate some empirically driven compositionality

arguments against prototype theory. I spent a substantial part of the spring of 2005 at City University in London working with Professor Hampton. It was an amazingly productive stay. I benefited enormously from the collaboration and from all the opportunities I got to tap into Professor Hampton's genius.

# Contents

*Words -- so innocent and powerless as they are, as standing in a dictionary, how potent for good and evil they become in the hands of one who knows how to combine them.*
– Nathaniel Hawthorne

# 1

# Introduction

## 1.1 The Feats of Speakers of Natural Languages

The ease with which we speak our native language does not betray the enormous communicative power we wield by speaking it. Everyday speech is often without effort, delay, or conscious deliberation. Nonetheless, the underlying competence is an extraordinary resource. We can speak *of* any word that we can speak *with* – not only can I say that Sweden has nine million citizens, but also that 'Sweden' has six letters. We can speak with the *same* words in different contexts with *different* effects – in one context I can speak of John by uttering 'He is really nice', while in another I can use the same expression to speak of James. We can speak of *things we have never encountered,* with *phrases we have never before heard* – not only can I speak, using familiar expressions, of pink things, of flying things, and of horses, but I can also speak, using the novel phrase 'pink flying horses', of pink flying horses.

Each of these capacities bestows a certain economy on the vocabulary of natural languages. The first capacity, pertaining to the *universalizability* of language, makes it unnecessary to introduce into the vocabulary a distinct word for every word. The second, pertaining to the *context sensitivity* of language most clearly illustrated by

personal pronouns such as 'I', 'you', and 'he', and by demonstrative pronouns such as 'this' and 'that', makes it unnecessary to burden the vocabulary with different words for different things that in different contexts bear similar relations to the speaker. Finally, language being such that combinations of familiar words can stand for something distinct from the things that each component word stands for, means that the vocabulary need not incorporate a word for every thing or type of thing. This *lexical* economy is mirrored in an obvious way in the phonetics and orthography of natural languages, inasmuch as the voiced and written expressions of natural languages often decompose into a small set of primitives (phonemes and letters, respectively).

   The last of the three capacities is particularly interesting. Not only does it enhance lexical economy, but, as has already been mentioned, it also makes speech of unfamiliar things possible. This, in turn, makes communication all the more efficient. Thus you can tell me about, warn me of, or encourage me to seek out, things I have never encountered. This ability to produce and understand novel expressions is certainly an intriguing cornerstone of our linguistic competence.

   In the philosophy of language there is near unanimity that this ability is best explained in terms of the *Principle of Compositionality*:

  (C)   The meaning of a complex expression is determined by
        the meanings of its parts and its mode of composition.

The following essay is concerned entirely with this principle.

### *1.2 Genesis of the Principle of Compositionality*

Compositionality was introduced under that name in 1963 by Jerrold Katz and Jerry Fodor in their 'The Structure of a Semantic Theory'. The principle as such is much older, however. It has often been attributed to Frege. Some passages in Frege's work and personal correspondence suggest a sympathetic attitude to something like *C*.[1]

---

[1] Passages in Frege suggestive of *C* are often couched in terms of thoughts – the entities that for Frege corresponded to the senses of sentences – being built up in the same way as the sentences expressing them. E.g. Frege writes: '…thoughts have parts out of which they are built up. And these parts, these building blocks, correspond to groups of sounds, out of which the sentence expressing the thought is built up, so

What is certain is that Frege believed in a principle closely related to *C*, namely the *Substitutivity of Co-Referentials*:[2]

(S)     The reference of a sentence remains unchanged when a word in it is substituted by another word with the same reference.

This principle can be traced back almost to the dawn of philosophy. A version of it operates as a tacit premise in the *Electra Paradox* – a paradox introduced 400 BC by Eubulides of the Megerian School that was widely discussed among the ancient stoics around 300 BC.[3]

---

that the construction of the sentence out of parts of a sentence corresponds to the construction of thought out of the parts of a thought. And as we take a thought to be the sense of a sentence, so we may call a part of a thought the sense of that part of the sentence which corresponds to it' (Frege [1914] 1979: 225); and 'We can regard a sentence as a mapping of a thought: corresponding to the whole-part relation of a thought and its parts, we have, by and large, the same relation for the sentence and its parts' (Frege [1919] 1979: 255). The passage most often cited to bring out Frege's belief in compositionality is the following: 'It is astonishing what language can do. With a few syllables it can express an incalculable number of thoughts, so that even a thought grasped by a terrestrial being for the very first time can be put into a form of words which will be understood by somebody to whom the thought is entirely new. This would be impossible, were we not able to distinguish parts in the thought corresponding to the parts of a sentence, so that the structure of the sentence serves as an image of the structure of the thought' (Frege [1923] 1977).

   Some observers, including Janssen (2001), deny that Frege believed in compositionality. Janssen's view seems to be based mainly on the belief that Frege never accepted the notion of 'the meaning in isolation of a sub-sentential expression'.

[2] I assume here that 'bedeutung' can be translated with 'reference'. Consider: 'Nehmen wir einmal an, der Satz habe eine Bedeutung! Ersetzen wir nun in ihm ein Wort durch ein anderes von derselben Bedeutung, aber anderem Sinne, so kann dies auf die Bedeutung des Satzes keinen Einfluß haben.' (Frege 1892: 32).

[3] Diogenes Laertius (2006: 98) attributes the Electra paradox to Eubuilides. The statement of the paradox found in Laertius (2006) is of the following form:

i)      Electra sees Orestes.
ii)     Electra knows that Orestes is her brother.
iii)    Electra does not know that the man she sees is Orestes.
C)      Electra does know and does not know her brother at the same time.

The gist of the paradox seems to be captured in the following, more stringent, version of the argument.

i)      The man Electra sees is Orestes.
ii)     Electra knows that Orestes is her brother.
iii)    Electra does not know that the man she sees is her brother.

The Principle of Compositionality significantly predates Frege as
well. There is, for instance, evidence from linguistic theorizing of the
thirteenth century which suggests a compositional picture of
meaning; and it has been suggested that that the Principle of
Compositionality has been an important element of semantic theory
ever since the work of Priscianus Caesariensis (fl. AD 500). The idea,
closely connected to the Principle of Compositionality, that general
rules are needed to arrive at the meaning of complex expressions, can
be found as far back as 200–100 BC in passages in Sanskrit.[4] So the
Principle – or at least ideas closely connected to it – has been around
for a very long time.

During the twentieth century Alfred Tarski's work on a recursive
specification of the semantics of first-order logic paved the way for
compositional approaches to the semantics of natural language.[5]

---

C)        Electra both knows and does not know that Orestes is her brother.

If this is the argument, the second conjunct of the conclusion only follows on the
assumption that co-referential expressions (in this case 'Orestes' and 'the man she
sees') can be substituted without changing the truth values of the sentences they
occur in. Hence, a version of *S* (in which 'the reference of a sentence' is replaced with
'the truth value of sentence') is a tacit premise in the argument. See Wedberg (1958)
for discussion of the connections between the paradox and the thinking of ancient
stoics.
[4] This paragraph draws upon (a) an extended version of Dever (2006) available at
https://webspace.utexas.edu/deverj/personal/papers/compositionalitylong.pdf,
accessed 2008-01-14 (the claim about linguistic theorizing in the thirteenth century);
(b) Szabó (2000: vii) (the claim about Priscianus Caesariensis); and (c) the following
translation of a passage from Patañjali's 'Paspaśā' (which seems to express the idea
that general rules are needed for understanding certain expressions):

> Bṛhaspati addressed Indra during a thousand divine years going over
> the grammatical expressions speaking each particular word, and still
> he did not attain the end. With Bṛhaspati as the instructor, Indra as
> the student, and a thousand divine years as the period of the study,
> the end could not be attained, so what of the present day when he
> who lives a life in full lives at most a hundred years?...Therefore the
> recitation of each particular word is not a means for the
> understanding of grammatical expressions. – But then how are
> grammatical expressions understood? Some work containing general
> and particular rules has to be composed[…]
>
>                                                    (Staal 1969: 501-502)

I became aware of Staal's discussion through Pagin (2007). See Hodges
(forthcomming) for a recent, very good discussion of the history of compositionality.
[5] See Tarski ([1933] 1956).

Through the influential work of Donald Davidson and Richard Montague, the Principle of Compositionality was placed at the center of semantic theorizing, and through their work also, together with the work of their followers, a large class of natural language constructions came to be given compositional treatment.[6] Today, much of the work being done in formal semantics deploys a compositional framework.[7] The situation is similar in contemporary philosophy of language, where the principle is widely assumed to be true.

## *1.3 The Aim of the Inquiry*

The Principle of Compositionality has been investigated in a variety of ways. Many, of course, have tried to determine whether or not it is true,[8] but it has also been approached *historically* (in inquiries into its original conception and whether this or that prominent scholar believed in it)[9], *formally* (in attempts to give it a precise statement in a formal language and relate it via proofs and definitions to other notions defined in that language)[10], and, as it were, *constructively* (in accounts, couched in psychological or linguistic theories, of the way in which the meanings of particular linguistic constructions are composed).[11] It has even been approached by way of *conceptual analysis*, the aim being to clarify the principle by exploiting 'not entirely pre-theoretical, but still, reasonably innocent[…]intuitions about what it is for a language to be compositional' (Szábò 2000:6).

   The approach that will be adopted in this essay, although drawing on previous work, will have a different focus. It will concern itself with the commonly supposed *instrumental* value of compositionality. Many among its proponents have entertained the idea that the

---

[6] See, e.g., Davidson ([1965] 2001; [1967] 2001; [1970] 2001) and Montague ([1970] 1974a; [1970] 1974b; [1973] 1974).
[7] See, e.g., Heim and Kratzer (1998), Chierchia and McConnell-Ginet (1990), and Larson and Segal (1995) for representative textbooks on formal semantics.
[8] See, e.g., Hintikka (1980), Pelletier (1994; 2000), and Lahav (1989).
[9] This approach is applied to Frege by Janssen (2001) and Pelletier (2001), and to Tarski by Hintikka and Sandu (1999).
[10] See, e.g., Hodges (2001), Pagin (2003), Fernando (2001) and Hendricks (2001) for this approach.
[11] See Larson and Segal (1995) and Heim and Kratzer (1998) for constructive approaches in formal semantics. See Murphy (2002) for an overview of work that has been done on the topic in psychology.

Principle of Compositionality can be used as a source of information about *meaning*. Here is a strong statement of this position.

> Over the last few years, we have just about convinced ourselves that compositionality is the sovereign test for theories of lexical meaning. So hard is this test to pass, we think, that it filters out practically all of the theories of lexical meaning that are current in either philosophy or cognitive science. Among the casualties are, for example, the theory that lexical meanings are statistical structures (like stereotypes); the theory that the meaning of a word is its use; the theory that knowing the meaning of (at least some) words requires having a recognitional capacity for (at least some) of the things that it applies to; and the theory that knowing the meaning of a word requires knowing criteria for applying it. Indeed, we think that only two theories of the lexicon survive the compositionality constraint: viz., the theory that all lexical meanings are primitive and the theory that some lexical meanings are primitive and the rest are definitions.
>
> (Fodor and Lepore [2001] 2002: 43)

The underlying idea seems to be the following. If the Principle of Compositionality is true, then, whatever the meanings of the parts of a complex expression *e* are, it must be the case that they, together with *e's* mode of composition, determine *e's* meaning. Hence, if the principle is true, a theory that concerns itself with the assignment of meanings to expressions can be evaluated in terms of whether or not it accords with the principle. If the theory turns out not to accord with compositionality, so the argument goes, we learn something about meanings – i.e. that they are not correctly described by that theory, or even perhaps that they are not the kind of thing invoked by that theory.

The present thesis aims to dissect this line of reasoning, in an attempt to illuminate whether there really is such a path from compositionality to meaning. More precisely, the goal of inquiry in this thesis is *to ascertain to what extent the Principle of Compositionality can be justifiably imposed as a constraint on semantic theories and thereby provide information about what meanings are.*

The inquiry is of considerable importance, since the previous reasoning, if correct, has far-reaching consequences. Assigning meanings to expressions is something that theories in many scientific

disciplines are concerned with. Hence, if we learn that the principle is true, this will have consequences not only for the philosophy of language, but also for various neighboring sciences, such as psychology and linguistics. We will then be in a position to justifiably abandon theories in these disciplines that conflict with compositionality, and to justifiably terminate research programs that depend on such theories. So, much hangs in the balance.

Each of the first four subsequent chapters presents a challenge to the idea that compositionality actually can be justifiably imposed as a constraint on semantic theories (and thereby inform us about meaning). Because each contains a discussion of whether the challenge in question can be overcome, each chapter brings us closer to the goal of inquiry.

Chapter Two is an attempt to determine whether the Principle of Compositionality is *trivial*. For if it is, it cannot be used as a constraint on semantic theories, since – the apparent contradiction notwithstanding – a trivial constraint is not a constraint at all.

Chapter Three and Chapter Four ask whether there are reasons to think that the Principle of Compositionality is *true*. If there are not, we cannot justifiably impose the principle as a constraint on semantic theories (and cease to consider theories that do not accord with the principle), since a non-compositional theory could still be *correct*.

Chapter Five is concerned with the converse issue: it asks whether there are any reasons to think that the Principle of Compositionality is *false*. If there are, and if these override any reasons that might exist for thinking that the principle is true, then there is no reason to think that a semantic theory that does *not* accord with the principle is incorrect – we will instead have a reason to think just the opposite.

Following the discussion of these challenges, Chapter Six, which importantly builds on the results of the previous chapters, explores the ways in which information about what meanings are can be derived by imposing compositionality constraints on semantic theories (given that these impositions are just).

The main conclusion of the thesis will be that the view that a semantic theory must accord with compositionality is *unjustified*, and that the information semantic theories in violation of compositionality provides us with therefore is limited. The explanatory reasons normally given for thinking that the Principle of Compositionality is true provide us with little reason to think that *C*

rather than some other competing determination principle – which is like *C* except that it feature additional determinants over and above the meanings of the parts of an expression and its mode of composition – is true.

   As a corollary to the main result of the inquiry, I will, after having explicated the Principle of Compositionality in a somewhat novel way, defend the position that this construal is to be preferred over others. Making the principle more precise is necessary since the statement of the principle given above is in many ways ambiguous: almost all the words in it permit multiple reasonable interpretations. Even though much philosophical work has already been put into teasing apart different interpretations, existing explications seem to be either too weak to be explanatorily satisfactory or too strong to be compatible with prominent characteristics of natural languages.

## *1.4 Meaning*

Since linguistic meaning is central to the aim of inquiry, some remarks should be made about it at the outset. The meanings we shall primarily be concerned with are those invoked when one holds, for instance, that 'doe' means the same as 'female deer', that 'Lisa caressed John' means the same as 'John was caressed by Lisa', that 'hug' and 'mug' mean different things, that 'Martin likes chips' and 'I like chips' mean different things, and that 'Jane had a book stolen' and 'John likes banks' can mean several things. The relevant meanings are meanings of *expressions*, i.e. meanings of *utterance-types*. As the term suggests, an utterance-type is something that can be uttered on many occasions. *Sentences*, *words* and *phrases* are particular kinds of expression. An *utterance* is something that cannot be uttered on more than one occasion; utterances are particular, non-repeatable events existing at particular points in space-time. What, exactly, the existence of expressions amounts to depends on the metaphysical theory of types one adopts. This will be left open in what follows. It should be noted though that the meaning of an expression is somehow connected with the pattern across the meanings of actual and potential utterances of that expression. Given the meanings of all the utterances of 'Lisa caressed John', the meaning of 'Lisa caressed John' is thereby constrained, if not determined. For simplicity, I will

assume that the meaning of an expression determines, given contexts of utterances, what utterances of that expression mean and vice versa.

It is important to realize that statements of the form '*e* means *m*' are elliptical. Expressions mean things only relative to a language, relative to a group of speakers or, in the limiting case, relative to a single speaker. For instance, 'Banks bar smoking' does not mean anything *simpliciter*.[12] Relative to English, or to a group of monolingual English speakers, it means that banks prohibit smoking. Relative to Swedish, or to a group of monolingual Swedish speakers, it means that someone named 'Banks' (e.g. Gordon Banks) was wearing a tuxedo.[13] Relative to a bilingual English-Swedish speaker it means both. So the statements that we shall be concerned to elucidate are of the form '*e* means *m* in language *l*' or '*e* means *m* to *S*' or '*e* and *e*\* mean the same thing to *S*'. Although the relativization is very important it will on occasion be inexplicit, as this helps to facilitate the exposition.

Meaning claims like those mentioned above pass without notice in everyday conversation, because our understanding of them for everyday purposes is usually unproblematic. However, when they are attended to more carefully, it becomes difficult to see what exactly they amount to – and, in particular, what the meanings that they invoke really are. Among the numerous candidates that have been suggested in the philosophical literature or invoked in semantic theorizing, are: things in the world, sets of things in the world, truth values, functions involving things in the world and truth values, fuzzy sets of things in the world, truth-conditions, abstractly conceived properties, possible worlds, sets of possible worlds, functions from possible worlds to truth values, functions from possible worlds to things or sets of things, inferential roles, verification conditions, mental representations, concepts, prototypes, schemata, thoughts, characters, and contents. In the view of many compositionalists, it is by delimiting this plurality of options that compositionality can inform us about what meanings are.

---

[12] I owe the example to Pernilla Asp.
[13] Another example, from Cappelen and Lepore (2005), is 'MIA LOVES FRED'. Relative to English, or to a group of monolingual English speakers it means that Mia loves Fred. Relative to Norwegian, or to a group of monolingual Norwegian speakers it means that Mia is being promised peace. The problem with this example is, of course, that it only has these two meanings when written in capital letters.

It is worth emphasizing that not all semantic theorizing is concerned with the metaphysical question of what meanings are. Hence, the invocation of a certain kind of semantic value, as part of a semantic theory, is not necessarily a metaphysical proposal. For instance, most semantic research done in linguistics and cognitive psychology is concerned with other questions. Whereas semantic research in linguistics might instead be concerned with the exploration of the relations between the semantic values of structurally related expressions, semantically related research in cognitive psychology might seek to predict how subjects perform in certain tasks that involve complex expressions on the basis of their performance of tasks involving the parts of these expressions. Theories in both these traditions might be seen, by the relevant theoreticians, as instruments of prediction, and possibly explanation, but the theoreticians in question might have no wish to commit themselves to the existence of the entities invoked by their theories.

Having said this, I should emphasize that as we progress with the question of what compositionality tells us about meaning, information might emerge about whether the entities invoked by the aforementioned kinds of semantic theory *can* be meanings. It might turn out that some of these entities cannot be meanings, or that others fit the role perfectly. Naturally, it is not very interesting to learn that things no one has ever supposed to be meanings cannot be meanings; but it is interesting to know how the space of candidate meanings is constrained by considerations of compositionality, and we might have to endure being told things that we already know if we are to digest all the consequences of this restriction.

In order not to unduly bias the inquiry, I will proceed by making as few assumptions as possible about what meanings can be. This means that in the various explications that will be given of *C* 'meaning' will be left unexplicated, and the properties of *C* will be examined without framing it in terms of some particular semantic theory. It does seem reasonable to leave the principle open in this way to avoid bias, although, as we shall see in Chapter Three, this approach complicates the process of assigning explanatory properties to the principle.

Although it is obviously pointless to try to list all the assumptions that will *not* be made about meanings in this inquiry, one deserves special mention. Historically speaking, it has been common in the philosophy of language to assume that *sentences* express complete

propositions, i.e. that sentences can be true or false. It seems that Frege, in 'Über Sinn and Bedeutung', makes this assumption, for instance.[14] And it is a frequent assumption in Paul Grice's attempts to explain the meaning of a sentence in terms of what users of that sentence mean (or should mean) by uttering that sentence on particular occasions.[15] The assumption has become increasingly antiquated, however, as philosophers of language have noted a number of ways in which the sentences of a natural language depend for their truth-evaluability on the context of utterance. For instance, the sentence 'I'm ready' is not truth-evaluable, since not all utterances of it have the same truth-value. The truth-value of an *utterance* of this sentence will depend on particular characteristics of the context: on whether *the speaker* is indeed ready at *the time of utterance*, and perhaps on whether he or she is ready with regard to some contextually salient event. So it seems that the sentence itself has no truth-value, i.e. fails to express a complete proposition. In view of this I will not assume that sentences express complete propositions. Indeed, the weaker assumption – suggessted by David Kaplan's work – that the meanings of sentences determine functions from contexts of utterance to propositions might also face difficulties, so this assumption will not be made either.[16] The reason for mentioning that the first of these assumptions will not be made is that is has been used as a simplifying assumption in much literature on compositionality.[17] Because the thesis does not make this assumption it is somewhat distanced from the results of earlier work. This should be kept in mind in what follows, since it means that departures from the conclusions of other writers need not mean that those writers are mistaken, but only that different results emerge when theorists simplify in different ways.

---

[14] Frege ([1892] 1997: 157ff.) holds explicitly that many sentences have truth-values (and that these constitute the Bedeutung of the relevant sentences).

[15] See Grice ([1957] 1989; [1968] 1989; [1982] 1989) for three contrasting proposals with the common assumption that the meanings of sentential utterance-types are propositions.

[16] See Kaplan ([1977] 1989). The disclaimer is not meant to deny that meanings determine functions from contexts to propositions. I am merely cautiously trying to constrain the inquiry as little as possible.

[17] See, e.g., Schiffer (1987) and Fodor and Pylyshyn (1988). The assumption continues to be made in quite recent work on compositionality such as Fodor (2001) and Pagin (2003).

## *1.5 Languages and Semantic Theories*

Both *languages* and *semantic theories* have been held to be compositional. This prompts, by way of pre-inquiry, discussion of: 1) what these entities amount to, and 2) whether, on explication, compositionality applies to them.

David Lewis has proposed that *a language* can be understood as a function from sentences to meanings. On his intended set-theoretic construal this means that a language is a set of ordered pairs, where the first member of each pair is a sentence, and the second is a meaning.[18] Given this minimal characterization of a language, *C* does not apply to languages. In order for *C* to apply to some *x*, it is at least required that the basic notions invoked by *C* – 'expression', 'meaning of an expression', 'complex expression', 'part of an expression', and 'mode of composition of an expression' – are interpretable in terms of *x*. But since Lewis's construal of a language does not contain any information about the structural relations between expressions, or for that matter information about what sub-sentential expressions there are, it deprives talk of, say, 'a complex expression' or 'a part of an expression' of sense.[19]

We can retain Lewis's basic idea that a language is a set-theoretic object and at the same time get something that *C* can apply to by changing the kind of object a language is supposed to be. This is what I will do. In what follows, I will assume that a language is *an ordered pair of a partial syntactical algebra and a function (an 'interpretation function') from expressions to meanings*.[20] On this construal, languages contain sufficient information for *C* to apply to them; the structural relations can be read off from the syntactical algebra and the algebra can encompass sub-sentential expressions.

On the set-theoretical interpretation of a language that unites Lewis's conception and my own, a language is an abstract object

---

[18] See Lewis ([1975] 2001). Davidson ([1992] 2001) uses 'essentially the same' concept as Lewis. Lewis assumes that sentences are sequences of *types of marks or sounds*, but since this is too narrow if, say, sign-languages are to be recognized as proper languages, I will not follow Lewis in making this assumption.

[19] Lewis was well aware of this. See Lewis ([1975] 2001: 572ff.).

[20] This account will be complicated slightly in the next chapter, where reasons will be given for thinking that the second member of the ordered pair is really a function from *grammatical terms* to *semantic ranges*: see Section 2.1. See the appendix for a clarification of the technical terms (such as 'partial algebra').

which does not exist anywhere in space-time. This means, for instance, that there are no direct causal connections between languages and the people that speak them. Instead, speakers are related to the languages they speak in some other way – plausibly, via *utterances* of expressions in the relevant language, i.e. via expression-tokens that do exist in space-time. A language, on the present conception, might exist whether or not anyone speaks it, has spoken it, or will ever speak it. I will leave this question open, but at the very least, on the current conception, there is no essential difference between, on the one hand, languages, such as Swedish or English, that are spoken by entire nations and, on the other hand, completely unique idiolects. I will also leave open the question whether a language contains a finite or an infinite number of expressions.

Since the inquiry concerns *C*'s deployment as an adequacy constraint on semantic theories, it is semantic theories rather than languages that will be in focus in what follows. It is therefore important that these are characterized in such a way that *C* applies to them as well. I will assume that a *semantic theory* is an *n*-tuple consisting of *at least*: 1) a partial syntactic algebra, 2) an interpretation function, and 3) a set of rules.[21] The syntactic algebra determines what the theory will treat as an expression and the way in which such expressions are structurally related. The interpretation function determines what meanings are assigned to which expressions, and the set of rules contains the descriptions in which the theory is framed. A semantic theory of this kind provides clear interpretations of the basic notions invoked by *C* – 'expression', 'meaning of an expression', 'complex expression', 'part of an expression', and 'mode of composition of an expression' – and it is thus clear that *C* can be applied to semantic theories. Importantly, by also containing a set of rules, a semantic theory of this kind permits formulations of the compositionality principle going beyond the basic ones.[22] My reason

---

[21] The interpretation function of a semantic theory is the same kind of thing as the interpretation function of a language. In the discussion of Section 2.1 I give reasons for thinking that the former should be construed as a function from *grammatical terms* to *semantic ranges*; but until we arrive at that discussion it can be thought of as a function from expressions to meanings.

[22] This conception of a semantic theory draws on ideas implemented in a framework developed by Wilfrid Hodges (2001) for the express purpose of discussing compositionality. Exact deviations from Hodges will be spelled out in the next chapter.

for refraining from simply identifying semantic theories with the first three members of the *n*-tuple is that I want to leave it open whether the meaning assignment effected by the theory draws on resources other than those corresponding to its first three members.

On my construal, semantic theories, like languages, are abstract set-theoretical objects, but semantic theories and languages differ in important respects: the partial syntactic algebra and the interpretation function which together constitute a complete language form only part of a semantic theory. This conception of a semantic theory can be seen as a syntactic-semantic construal of semantic theories, in the sense that it contains both parts that correspond to the things the theory is about – meanings and expressions – and a part that is used to describe those things – the rules. So a rough way to think of theories like this is in terms of *interpreted descriptions*. Since semantic theories contain descriptions as parts, they are fairly fine-grained. This conception of semantic theory might be novel, but I think that it is permissive enough to encompass a wide range of the things referred to, in a non-technical sense, by 'semantic theory'. Standard truth-theories of natural languages, for instance, and psychological models of prototype combination, fit this mold.

The core motivation for construing a semantic theory as the *n*-tuple described above is that there is no built-in requirement that a semantic theory must be compositional. Such a restriction would be inappropriate given the present goal of inquiry. If a linguistic theory is analytically, necessarily compositional, then compositionality cannot be used as an adequacy constraint on semantic theories, since it will then be trivially satisfied. Although it seems appropriate in the present context, my decision not to define semantic theory in terms of compositionality departs from common practice in the philosophy of language.[23] An additional motivation for adopting this kind of

---

[23] Consider, e.g., how Robert Stalnaker, characterizes a 'descriptive-semantic theory'.

> A descriptive-semantic theory is a theory that says what the semantics for the language is, without saying what it is about the practice of using that language that explains why that semantics is the right one. A descriptive semantic theory assigns semantic values to the expressions of the language, and *explains how the semantic values of complex expressions are a function of the semantic values of their parts*
>                                        (Stalnaker 1997: 535, emphasis added)

conception of a semantic theory is that it is compatible with drawing the semantics-pragmatics distinction in several different ways.

To illustrate various points connected with semantic theories as clearly as possible I will, in the remainder of the essay, repeatedly provide descriptions of simple semantic theories along the following lines.

---

This amounts to a compositional characterization of a descriptive semantics, if one assumes – surely reasonably – that Stalnaker admits the possibility that the meanings of complex expressions with different modes of combination can be explained in different ways, and that 'is a function of' is understood along the lines of 'is determined by'.

Another telling example is a recent article by Jeffrey King and Jason Stanley, where the authors describe three ways of drawing the pragmatics-semantics distinction, and where all the distinctions, though importantly different, share the assumption that semantics is compositional:

> According to the first strategy for making a distinction between semantics and pragmatics, the semantic interpretation of a complex expression *e* is the result of composing the standing meanings of the lexical items in *e* in accordance with the *semantic composition rules* corresponding to the syntactic structure of *e*.
>
> (King and Stanley 2005: 115, emphasis added)

> According to a more contemporary conception of the semantics-pragmatics distinction, there are two levels of semantic value. The first is the non-relativized notion of the standing meaning of an expression. The second is the relativized notion of the referential content of an expression in a context[…]a complex expression relative a context *c* has a referential content that is the result of combining the referential contents of its constituent terms relative to the context *c* in accord with the *semantic composition rules* corresponding to the syntactic structure of that expression. The result of this latter process is a genuine level of semantic value.
>
> (Ibid: 116, emphasis added)

> Given Perry's distinction between automatic and intentional context-sensitive expressions, one might adopt the following distinction between semantics and pragmatics. The semantic content of a sentence in a context is a function of (and only of) the referential contents of the automatic indexicals in the sentence relative to that context, together with the standing meanings of all other lexical items in the sentence, together with *context-independent composition rules*.
>
> (Ibid: 120, emphasis added)

$t_1$    *Syntax:*                    *Semantics:*
        N   → 'dog',      $\mu$('dog')     = {x : x is a dog}
              'cat'       $\mu$('cat')      = {x : x is a cat}
        A   → 'brown',    $\mu$('brown')   = {x : x is brown}
              'black'     $\mu$('black')    = {x : x is black}
        NP  → A·N         $\mu(_{NP}$ A N) = $\mu$(A) $\cap$ $\mu$(N)

These descriptions should be understood in the following way. The syntax is described in terms of syntactic rules which are either 1) *explicit syntactic rules* (e.g. 'N → dog, cat') specifying that certain expressions (in this example, 'dog', and 'cat') are *simple expressions* belonging to a certain *syntactic category* (here, the N-category), or 2) *implicit syntactic rules* (e.g. 'NP → A·N') specifying that members of certain syntactic categories (here, members of the A-category and the N-category) can be concatenated to form *complex expressions* belonging to a certain syntactic category (here, the NP-category).[24] Syntactic categories of the latter kind correspond to *modes of composition*. The name of a syntactic category can also be treated as a variable over the expressions in that category.

   The semantics of these theories will be described in terms of semantic rules that are either 1) *explicit semantic rules* (e.g. "$\mu$('dog') = {x : x is a dog}") specifying the meaning (here, the set of all dogs) of a particular expression (here, 'dog') or an *implicit semantic rule* (e.g. '$\mu(_{NP}$ A N) = $\mu$(A) $\cap$ $\mu$(N)') specifying the meaning of expressions *qua* some general description of expressions (here, 'A N', which is read: 'an expression that results from concatenating an expression of the A-category with an expression of the N-category').

   The descriptions of the simple theories, together with the preferred interpretation of these descriptions (as set out in the previous two paragraphs), determine the partial syntactic algebra, the interpretation function and the set of rules that jointly constitute a semantic theory.

---

[24] Many think that this kind of syntactic theory is too simple to model the syntax of natural languages and it should be noted that by exemplifying semantic theories in terms of this simple form of syntax I do not mean to imply that a semantic theory must be couched in terms of this kind of syntax; my point is merely that additional complications often fail to facilitate the relevant compositionality aspect that I want to illustrate with the theory, and that therefore a simpler presentation is more suitable.

[25] On occasions where some details of the semantic theory are left out, and the descriptions given only determine a set of theories, I will permit myself to use the phrase 'the $t_n$- theory', although in fact only a set of theories is determined by the descriptions given, not a unique theory. The claims I make about that theory should then be interpreted as being about *all* the relevant theories.

It is an important benefit of construing a semantic theory as the aforementioned *n*-tuple that compositionality can be discussed independently of the goal one has in mind when doing semantics. Given this conception, a semantic theory can be interpreted *as being about* many things. [26] Two alternatives will be particularly important, and the second will be the main focus of what follows. A semantic theory can be interpreted either so that it is about *a language* or so that it is about *a speaker's semantic competence*, i.e. roughly, *the speaker's knowledge of meanings* – the latter being a certain mental state. Depending on our response to this issue of interpretation,

---

[25] The definition of 'semantic theory' given in the appendix spells out exactly this interpretation of the rules in terms of constraints linking the rules of the theory with the rest of the theory.

[26] Richard Larson and Gabriel Segal, who construe the subject matter of the semantic theory they develop (i.e. what their semantic theory is about) as speakers' linguistic knowledge, list a number of other alternatives.

> To view the subject matter of semantic as linguistic knowledge is to locate the place of semantic theory within the general enterprise initiated by Noam Chomsky (1965, 1975, 1986a), for whom linguistic theory is a theory of the real knowledge of speakers. This project contrasts with a variety of other commonly held views of the subject matter. For example, some have taken semantics to be a theory of the semantic relations holding between expressions (including inferential and thematic relations). Many others have construed semantics as a theory of the relations holding between language and the world. Still others have insisted that since languages are abstract objects, linguistics (including linguistic semantics) should be pursued as a branch of mathematics. Our conception differs from all of these. On our view, semantics is part of a theory of speakers' knowledge. Facts about language-to-language and language-to-world relations may furnish important clues about the content of this knowledge–they may furnish data–but they are not the object of inquiry itself. The object of inquiry is knowledge of language.
>
> (Larson and Segal 1995: 10)

different conditions will presumably need to be met if the theory is to be *correct*. I will now explore this presumption.

Assume to begin with that a semantic theory *t* is interpreted so that it is about a language *l*. When is *t* a correct theory of *l*? Consider first the simple case in which a theory is understood in terms of the same kind of set-theoretical object as the language was identified with, i.e. an ordered pair of syntactic algebra and function from expressions to meanings. One then cannot hold (what would otherwise have been natural) that *t* is correct if it correctly describes *l*, since theories – on this conception – are not the kind of thing that describes anything. But an even simpler formulation is instead available, namely that *t* is correct if *t* = *l*. In the present context we cannot settle for this conception of a semantic theory, since it does not permit clear formulation of the Principle of Compositionality. However, the correctness conditions for theories of the favored kind can be framed in a corresponding way. A semantic theory *t*, if interpreted as being about a language *l*, is correct if the syntactic algebra of *t* and the syntactic algebra of *l* are identical, and if the interpretation function of *t* and the interpretation function of *l* are also identical. When a semantic theory is interpreted as being about a language, it seems reasonable to hold that the semantic (or the syntactic) rules of the theory (e.g. '$\mu(A\cdot N) = \mu(A) \cap \mu(N)$') are irrelevant as far as correctness goes. There is no fact of the matter in this case (at least not on the present conception of language) that could arbiter between theories containing different sets of rules but also contains the same interpretation function (and the same syntactic algebra).

Assume now that a semantic theory *t* is instead interpreted as being about the semantic competence of a speaker *S*. This interpretation places semantic theorizing in a tradition incorporating the work of, notably, Noam Chomsky – a tradition in which it is assumed that linguistic theorizing is really part of psychological theorizing about individual speakers.[27] It is under this interpretation that a semantic theory most clearly lends itself to the explanation of the phenomena compositionality is usually held to explain (e.g. speakers' understanding of novel expressions). On this interpretation of semantic theory, the syntactic algebra and the interpretation function of the theory are still relevant to the correctness of the theory, and it

---

[27] See, e.g., Chomsky (1957; 1965; 1986).

seems very reasonable to assume that there is a fact of the matter about the ways in which expressions are structured and meanings are assigned to expressions. But in this case this might only be a *necessary* condition of correctness. For on the present interpretation of semantic theory, it seems reasonable also to require that *the correctness of a theory depends on whether the semantic rules of the theory* (e.g. '$\mu(A \cdot N) = \mu(A) \cap \mu(N)$') *correctly correspond, in some way, to the mechanism in S which is responsible for S conforming to the interpretation function.* This is vague, and it is controversial whether it can be spelled out more exactly.[28] I will make an attempt to make sense of it when the issue is discussed in more detail in Chapter Three.

A final note on terminology: I shall assume that the notion that *C is true* (relative a certain speaker's linguistic competence, or to a certain language) is equivalent to the notion that any correct theory (there may be several), about that speaker's competence or about that language, accords with *C*.

---

[28] Quine (1972) and Wright (1986), e.g., have expressed concern about this.

*I trust that none will stretch the seams in putting on the coat, for it may do good service to him whom it fits.*
– Henry Thoreau

# 2

# The Non-Triviality of Compositionality

The goal of inquiry in this essay is to ascertain to what extent the principle of compositionality (*C*) can be justifiably imposed as a constraint on semantic theories, and thereby provide information about what meanings are.

(C)    The meaning of a complex expression is determined by the meaning of its parts and its mode of composition.

If *C* turns out to be a *trivial* constraint on semantic theories, it *cannot* be used in this way since it is by being incompatible with semantic theories that *C* can provide semantic information. For if *C* is incompatible with a certain semantic theory, and *C* is likely to be true, it can tell us – at the very least – that meanings are probably not correctly represented by that theory. However, if *C* turns out to be trivial, i.e. if *C* turns out to be in accordance with *all* semantic theories, there would be no incompatibilities between *C* and semantic

theories from which we can learn anything. *C* would then be like, for instance, a requirement on axiomatic biological theories that they must have among their theorems the tautologies that can be stated in the biological vocabulary (e.g. 'Cells emerge only from preexisting cells or it is not the case that cells emerge only from preexisting cells', 'Animals and fungi share the same basic machinery that copies and transcribes DNA into proteins or it is not the case that animals and fungi share the same basic machinery that copies and transcribes DNA into proteins'); all relevant theories conform to the requirement, so we can learn nothing about the relevant domain from imposing it.

That *C* is not trivial *simpliciter*, i.e. that *C* is not in accordance with all semantic theories, will be immediate once *C* has been stated precisely, and the aforementioned threat to its informative potential can thereby be swiftly removed. But *C* might also be trivial in other ways, and its informative potential could thus be threatened from other directions. As a matter of fact, when one turns to the literature on compositionality, it turns out that the view that compositionality (on some construal) is trivial (in some sense) has several proponents.[1]

In order to address the issue of whether compositionality is in some sense trivial, this chapter will begin by introducing a first explication of *C* – what will be referred to as '$C^{FE}$' – and discuss whether *C* (on this explication) is a trivial constraint in either of three different senses. After giving the first explication of compositionality (Section 2.1), and showing that it is not trivial *simpliciter* (Section 2.2), I will introduce two other senses in which compositionality might be held to be trivial (also in Section 2.2). The next two sections (sections 2.3 and 2.4) will each discuss one of these senses, possible reasons found in the literature for thinking that *C* is trivial in that sense and the implications for the informative potential of *C* if *C* turns out to be trivial in that sense. It will be concluded (Section 2.5) that *C* (as explicated in this chapter) is not trivial on any of the three senses of 'trivial' considered in the preceding sections.[2]

---

[1] See, for instance, van Benthem (1984), Hendriks (1993), Horwich (1997; 2001; 2005), Zadrozny (1998), and Lappin and Zadrozny (2000).

[2] My conclusion places me among researchers such as Westerståhl (1998), Kazmi and Pelletier (1998), Dever (1999), Pelletier (2000), Fodor and Lepore ([2001] 2002) and Cohnitz (2005).

### *2.1 A Weak Compositionality Explication*

Although the gist of compositionality can be read off *C*, it is too vague to determine whether a given semantic theory is compositional or not, and thus whether there are semantic theories that are non-compositional, i.e. whether compositionality is trivial. In order to remedy this, a precise version of *C* will be offered in this section. *C* will be explicated in terms of a very weak principle since if it can be shown that compositionality is non-trivial on this explication, it follows that all explications of *C* that we will consider in the rest of the essay are also non-trivial, since they are all logically stronger this principle.

 'is determined by' is the source of a substantial portion of *C*'s vagueness, so my attempt to make *C* more precise will begin by considering this part. The common understanding of compositionality (which will be adhered to in the essay) is in terms of the claim that both of the cited *determinants,* i.e. both the meanings of the parts of an expression, and its mode of composition, are *necessary* to determine the cited *determinable*, i.e. the meaning of that expression, and that they are *jointly sufficient* to determine the cited determinable. Consider (1), another determination claim, to see the force of this.

(1)     The amount of money deposited in and withdrawn
        from one's bank account *determines* one's balance.

It seems to me that a normal utterance of a determination claim like (1) is, strictly speaking, true or false given that the cited determinants (in this case: the amount of money deposited, and the amount of money withdrawn) are *necessary* and *jointly sufficient* for determining the cited determinable (in this case: the balance). So an utterance of (1) could turn out false due to the existence of some other determinant  (for instance: the fact that the bank account bears interest) or due to the fact that the cited determinable does not depend on one or more of the cited determinants (for instance: if it is the case that no matter how much money one withdraws from the

account, the balance is unaffected). Compositionality is normally interpreted along similar lines and I will follow suit in this essay.[3]

---

[3] That compositionality is generally held to entail the claim that *only* the cited determinants enter into the determination of the meaning of complex expressions is clear from the following quotations. 'Very roughly, a system of representations is compositional iff: (a) it contains both syntactically primitive and syntactically complex symbols; and (b) the syntax and content of the complex symbols is *exhaustively* determined by the syntax and content of their primitive constituents.' (Connolly, Fodor et al. 2007: 5, emphasis added), 'Nobody denies that the meaning of a complex expression depends on the meanings of its constituents and on its structure; the bite of compositionality is that *it depends on nothing else*.' (Szabó 2001: 121, emphasis added), 'The Principle of Semantic Compositionality is the principle that the meaning of an expression is a function of, *and only of*, the meanings of its parts together with the method by which those parts are combined. (Pelletier 2004: 133), and 'Compositionality is a tool for limiting what can be relevant to determining the meaning of a complex expression[...]Only *semantic* information can go into the determination of the semantic value of a complex expression' (Dever 2006: 647).

Not being clear on the difference between 'is determined by' and 'depends on' (i.e. the reading you get if you do not think compositionality entails that only the cited determinants enter into the determination of the cited determinable) might lead to undue conviction in the truth of the principle of compositionality. It should thus be noted that it is often the case that when a connection between the meanings of the parts of an expression and the meaning of that expression is claimed to be obvious or self evident the connection actually asserted is not one of *determination* but only one of *dependence*. The following quotations illustrate this.

> Take again 'Socrates loves Plato'. This is a complex symbol, composed of three symbols, namely 'Socrates', 'loves' and 'Plato'. Whatever may be the meaning of the compound symbol, it is clear that it *depends* upon the meaning of the separate words
> (Russell [1919] 1997: 290, emphasis added)

> It is conceded by most philosophers of language, and recently by some linguists, that a satisfactory theory of meaning must give an account of how the meanings of sentences *depend* upon the meanings of words[...]I do not dispute these vague claims, in which I sense more than a kernel of truth.
> (Davidson [1967] 2001: 17, emphasis added)

> It is uncontroversial that, apart from idioms, the meaning of any complex expression-type (such as a sentence) *depends* on the meanings of its component words and on how those words have been combined with one another.
> (Horwich 2005: 9-10, emphasis added)

With this in place, we can now turn to an explication of 'is determined by'. Famous remarks made by David Lewis provide inspiration.[4]

> What sorts of things determine how something depends on something else? *Functions*, of course; functions in the most general set-theoretic sense, in which the domain of arguments and the range of values may consists of entities of any sort whatever, and in which it is not required that the function be specifiable by any simple rule.
>
> (Lewis [1970] 2004: 194, emphasis in original)

A function 'in the most general set theoretic sense', is a *function-in-extension*; a single-valued set of ordered pairs, i.e. a set of ordered pairs $f = \{<x_1,y_1>,\ldots,<x_n,y_n>\}$ such that for each member $x$ of the domain of $f$ (i.e. the set $\{x : <x, y> \in f\}$) it is the case that there exists at most one $y$ such that the ordered pair $<x, y>$ is a member of $f$.[5] There is no restriction on what the members of these pairs can be, they can, for instance, themselves be sets, or functions, or ordered $n$-tuples.

It is important not to confuse a function-in-extension with a *function-in-intension*, i.e. a kind of *rule*.[6] So '3x+4', for instance, is a function-in-intension, and it, and many other functions-in-intension (such as 'x+x+x+4'), determine a certain function-in-extension, namely the set $\{<1, 7>, <2, 10>, <3, 13>\ldots\}$. Although this very weak notion of a function according to which a function is not required to be computable or even describable, might strike one as unsuitable for present purposes, it at least captures the idea common to all determination claims: if the determinants are the same, then the determinable must be the same as well.[7]

In order to arrive at a suitable explication of $C$ in terms of a function-in-extension, we can go via some suggestions made by Wilfrid Hodges, who has developed an influential framework for

---

[4] Note that Lewis made these comments in a different context. He was concerned with characterizing intensions in terms of functions-in-extensions.

[5] See the appendix for more information about the notation.

[6] The terms 'function-in-extension' and 'function-in-intension' are taken from Church (1951).

[7] It is by being very weak that this explication of 'is determined by' is suitable for an explication of $C$ in the present chapter since this allows the chapter's conclusion concerning it to be generalized to other explications of $C$ that we will later be concerned with.

discussing compositionality and related notions.[8] It will be instructive to spell out his formulation of compositionality first and then use it as a contrast for the formulation (which is very close to his) that I want to use in the remainder of this chapter.

In Hodges' framework, a *grammar*, which intuitively represents the 'surface structure' of the expressions of a language, is defined as a triple $<E, A, \Sigma>$ where $E$ is a set of *expressions*, $A$ is a set of *atomic expressions* which is a subset of $E$, and $\Sigma$ is a set of *syntactic rules* where each syntactic rule $\sigma$ is a partial function $\sigma : E^n \rightarrow E$ for some non-negative integer $n$ called the *arity* of $\sigma$. Since Hodges wants to allow for structural ambiguity he also makes use of a *grammatical term algebra*, which intuitively represents the 'deep structure' of the expressions of a language. These terms are then mapped via a function *val* onto the expressions of the grammar (*the values* of these terms), representing the various structural analyses of the expression they are mapped onto. The grammatical term algebra of $<E, A, \Sigma>$ is the set $GT(<E, A, \Sigma>)$ (*GT* for short) of *grammatical terms*, defined inductively by:

i)     Every atomic expression is in $GT$; its *value* is itself.

ii)    If $\sigma$ is a syntactic rule of arity $n$, the terms $t_0, \ldots, t_{n-1}$ are all in $GT$ and have values $e_0, \ldots, e_{n-1}$ respectively, and the expression $\sigma(e_0, \ldots, e_{n-1})$ is defined, then the term '$\sigma(t_0, \ldots, t_{n-1})$' is in $GT$ and its *value* is the expression $\sigma(e_0, \ldots, e_{n-1})$.[9]

The expressions that *val* maps onto themselves are said to have *trivial analyses*. Hodges then defines the interpretation function $\mu$ as a partial function on a subset of $GT$ (rather than on a subset of $E$). This means that if a complex expression has several structural analyses it can have several meanings (given that an expression is assumed to have the meanings of its structural analyses). Since $\mu$ is not required to be total the existence of meaningless expressions is allowed for. $\mu$ maps

---

[8] See Hodges (2001). See, for instance, Pagin (2003; 2004) and Westerståhl (2002; 2004) for work inspired by Hodges.

[9] Hodges makes use of a term-algebra as well, but since this is primarily needed to spell out substitution versions of compositionality (which I will not be concerned with) I have omitted this part in my exposition. I will instead use 'term' as an abbreviated form of 'grammatical term'.

grammatical terms into a set of *meanings*. Nothing is assumed about the elements of this set. If $\mu$ is defined for a certain grammatical term, that term is *$\mu$-meaningful*. Derivatively, we can say that an *expression* which is the value of a $\mu$-meaningful grammatical term is *$\mu$-meaningful*.

Since we will frequently represent semantic theories in a certain way in this essay, it will be helpful to illustrate Hodges' notation by showing how these representations can be translated into it. For that purpose consider the following theory.

| $t_1$ | Syntax: | | Semantics: | |
|---|---|---|---|---|
| | S | $\rightarrow$ 'Bob runs', | $\mu$('Bob runs') | $= true$ |
| | | 'Sue swims' | $\mu$('Sue swims') | $= true$ |
| | C | $\rightarrow$ 'and', | $\mu$('and') | $= \wedge$ |
| | | 'or' | $\mu$('or') | $= \vee$ |
| | S | $\rightarrow$ S·C·S | $\mu(_s\, S_1\, C\, S_2)$ | $= \mu(C)(\mu(S_1), \mu(S_2))$ |

A description like this translates into a description in terms of expressions and grammatical terms in the following way. All strings that can be obtained by the syntactic rules are expressions (i.e. 'Bob runs', 'Sue swims', 'and', 'or', 'Bob runs and Sue Swims', 'Bob runs or Bob runs' etc.). The expressions given by the explicit syntactic rules (i.e. those that feature expressions and not only variables) are the atomic expressions (i.e. 'Bob runs', 'Sue swims', 'and', and 'or'). All the atomic expressions are also grammatical terms. In addition, corresponding to each way in which an expression can be derived by the syntactical rules there is a grammatical term. For instance, 'Bob runs and Sue swims or Bob runs' can be derived in two different ways, and corresponding to these there are the following two grammatical terms.

(2)    '$\sigma(\sigma($'Bob runs', 'and', 'Sue swims'), 'or', 'Bob runs')'
(3)    '$\sigma($'Bob runs', 'and', $\sigma($'Sue swims', 'or', 'Bob runs'))'

It will sometimes be convenient to refer to grammatical terms by using bracketed representations like the following.

(2')    $[_s\, [_s\, [_s\, \text{Bob runs}]\, [_c\, \text{and}]\, [_s\, \text{Sue swims}]]\, [_c\, \text{or}]\, [_s\, \text{Bob runs}]]$

(3')    [$_S$ [$_S$ Bob runs] [$_C$ and] [$_S$ [$_S$ Sue swims] [$_C$ or] [$_S$ Bob runs]]]

The two grammatical terms have the same value, namely 'Bob runs and Sue swims or Bob runs'. All grammatical terms, and (derivatively) all expressions covered by $t_1$ are $\mu$-meaningful (having truth-values as meanings).

   The framework that Hodges describes also gives a certain precise meaning to 'mode of composition'. Each of the syntactic operations in the grammar corresponds to a way in which expressions can be combined. Since expression might be possible to derive with several different operations it is really the grammatical terms that have modes of composition. Derivatively, one can hold that an expression has the modes of composition of the grammatical terms of which it is a value. In theories represented in the manner of $t_1$ there corresponds a mode of composition to each implicit syntactical rule. Hence, the complex terms and expressions covered by $t_1$ only have one mode of composition ('$_S$ S$_1$ C S$_2$').

   After having introduced his framework Hodges gives several formulations of compositionality. The simplest one is along the following lines.

   $\mu$ is *compositional* iff

   i)    Each subterm of a $\mu$-meaningful expression is
         $\mu$-meaningful, and

   ii)   There is a function $r$ such that for every
         complex $\mu$-meaningful term $s = \sigma(e_0,\ldots, e_{n-1})$,

         $$\mu(s) = r(\sigma, \mu(e_0),\ldots, \mu(e_{n-1})).^{10}$$

---

[10] It is assumed here that $r$ does not have to be of a particular arity. There are ways to get around this oddity but since they complicate the formulation I've settled for this version.

   I'm deviating slightly from Hodges' formulation. He does not include i) as a requirement which is part of the formulation of compositionality, but as an assumption which is made in order for ii) to apply. It will still be case that my and Hodges' formulations are such that exactly the same meaning functions are compositional. But on my conception, one can end up with the result that a certain meaning function is not compositional when one ends up with a presupposition failure on Hodges' framework. My point with deviating is just to make salient a point that I will later criticize: that i) needs to be true in order for $\mu$ to be compositional.

Intuitively, in order for there not to exist a function *r* of this kind, there has to exist two complex grammatical terms which have different meanings, but with the same mode of composition and parts with exactly the same meanings. Note that since the existence of *r* is a matter of the existence of a certain function-in-extension this formulation of compositionality is very weak. A theory like $t_1$, for instance, is in accordance with it.[11]

| $t_2$ | *Syntax:* | | *Semantics:* | |
|---|---|---|---|---|
| | A | $\rightarrow$ 'black' | $\mu$('black') | $= m_1$ |
| | N | $\rightarrow$ 'dog' | $\mu$('dog') | $= m_2$ |
| | NP | $\rightarrow$ A·N | $\mu$('black dog')$= m_3$ | |

According to this theory there is only one meaningful complex grammatical term (that underlying 'black dog') and since both its sub-terms ('black' and 'dog') are meaningful i) is satisfied. And since any function that contains the pair <<NP, $m_1$, $m_2$>, $m_3$> satisfy the requirements for *r*, ii) is satisfied as well. So $t_1$ is compositional. The important feature to notice here is that Hodges' compositionality formulation only constrains how meanings *distribute* over expressions, and not how they are 'connected'. So, for instance, theories that do not encompass multiple grammatical terms that share modes of composition will be compositional. Hodges' formulation is very similar to formulations in terms of the existence of a homomorphism from syntax to semantics, which are also very weak.[12]

---

This follows from Hodges' formulation, but, it seemed to be clearer on this formulation than Hodges' original one.

[11] The previous theory was also compatible with this version of compositionality.

[12] The homomorphism formulation is common in the tradition in formal semantics initiated by Richard Montague. See, for instance, Montague ([1970] 1974: 227), Janssen (1986) and Partee (1997). Since Hodges' formulation does not accord with the ordinary definition of a homomorphism it might be worthwhile to spell out the connection between his formulation and formulations in terms of homomorphisms.

Informally, given an arbitrary set *A* of elements, a set *F* of *n*-adic properties of the members of *A,* an arbitrary set *B* of elements, a set *G* of *n*-adic properties of the members of *B,* and a correspondence between the properties in *F* with the properties in *G* of the same acidity, a mapping from *A* to *B* is a homomorphism iff each member *a* of *A* has a property in *F* iff the corresponding member of *B* has the corresponding property in *G*. A homomorphism is thus a property preserving function.

Hodges' formulation is quite similar to the formulation that I will adopt in this chapter, but it deviates in one important respect. Hodges' formulation, and most other formulations of the principle of compositionality, are not pure determination claims. Most formulations of compositionality really involve two different, separable aspects. The first aspect is *the determination claim*; that complex expressions being in a certain way with respect to modes of composition and the meanings of their parts *determine* how they are with respect to their meanings. The second is *a quantified claim*; that expressions have a certain amount of meanings and modes of combination. Adding the following emphasis to *C* highlights this aspect.

---

Homomorphism claims are often made in the context of an algebraic framework. Here, the members of the sets *F* and *G* are restricted to operations. In this framework the homomorphism requirement can be stated more formally. An algebra <*A, F*> is an ordered pair consisting of a *carrier A* which is a set of elements and a set *F* of operations on that set. Normally it is assumed that each operation is total, i.e. that each operation is such that if it is of arity *n* it is defined for all *n*-tuples of elements of the relevant carrier. If this assumption is made, then given two algebras <*A, F*> and <*B, G*>, and a function *C* which maps functions of *F* to functions of *G* with the same arity, a function $\mu$ from *A* to *B* is a *homomorphism* iff for each *n*-ary operation $f \in F$, ordered *n*-tuple $a_1,\ldots,a_n$ and element $a_{n+1}$ such that each $a_i \in A$,

(i)     $f(a_1,\ldots,a_n) = a_{n+1}$ iff $C(f)(\mu(a_1),\ldots, \mu(a_n)) = \mu(a_{n+1})$.

If $\mu$ is one-to-one and onto *B* then $\mu$ is also an *isomorphism*.

Strictly speaking, in order to ask meaningfully whether $\mu$ is a homomorphism <*A, F*>, <*B, G*> and *C* must all be given. However, several other non-trivial questions connected to the homomorphism requirement can be meaningfully asked without all of these parameters being set. For instance; given <*A, F*>, <*B, G*> and *C does there exist a function $\mu: A \rightarrow B$ which is a homomorphism*? Given <*A, F*>, <*B, G*> and $\mu$ *does there exist a function C such that it makes $\mu$ a homomorphism*? Given <*A, F*>, *B* and $\mu$ *does there exist a function C and a set of operations G such that $\mu$ is a homomorphism with respect to <A, F>, <B, G> and C*? It is this last question that corresponds to the compositionality formulation given in the main text.

Note that dropping the requirement that the operations of an algebra are total (which is done in Hodges's framework) complicates the definition of a homomorphism. For even though there might exist a function *C* which maps functions of *F* to functions of *G* with the same arity, it might not be the case that a function $f \in F$ is defined for elements $a_1,\ldots,a_n$ iff *C(f)* is defined for elements $\mu(a_1),\ldots,\mu(a_n)$. The simplest way to adapt the definition of a homomorphism to a partial algebra is just by adding this constraint. (See Grätzer (1979) for additional stronger requirements.)

(C)      *The* meaning of a complex expression is determined by
         *the* meaning of its parts and *its* mode of composition.

Given a Russellian analysis of 'the' and the natural understanding of
'its mode of composition', *C* embodies the claim that complex
expressions and parts of complex expressions have *unique* meanings,
and that complex expressions have *unique* modes of composition.[13]
Returning to Hodges' formulation, we can see that it too embodies
similar quantificational restrictions. In particular in order for an
interpretation function to be compositional, i) subterms of
meaningful expressions cannot be meaningless, ii) atomic expressions
cannot have several meanings unless they have non-trivial structural
analyses, and iii) expressions cannot have multiple meanings unless
they have multiple structural analyses.[14]

  The determination claim and the quantified claim are separable
claims, and since we are concerned with devising a very weak
formulation of compositionality, we have a reason to remove
inessential material from it. More importantly, when we will later
build other more explanatory able explications of compositionality on
top of the formulation given in this chapter, it is important that we
do not build into these formulations material which is not needed for
explanatory reasons. If we do so we might end up in the position once
we start using compositionality as a constraint on semantic theories
where we require that a semantic theory must be in a certain way even
though we are not strictly speaking explanatory licensed to do so.
Finally, by building in inessential material into our compositionality
explication we run the risk of ending up with a formulation that is
more susceptible to criticism than it ought to be. We will see in
Chapter Five that this is a very real threat and that formulations that
embody the aforementioned quantified claims run into problems

---

[13] See Russell ([1919] 2001).
[14] Similar assumptions seem to be quite common. Westerståhl, for instance, remarks
that 'Standard formulations of the compositionality principle presuppose single-
valuedness of meaning.' (Westerståhl 2002: 243). Westerståhl (2004) shares Hodges'
quantification restrictions except for the first one. Substitutional versions of
compositionality avoid some of these commitments.

concerning meaninglessness and non-lexical non-structural forms of ambiguity.[15]

Hodges' formulation is already compatible with a complex expression having any number of modes of composition – since modes of compositions really belong to grammatical terms and a single expression can correspond to any number of grammatical terms with different modes of composition – so we only have to make adjustments to remove the quantificational restrictions put on meanings. One compelling way to do this is through employing the notion of a *semantic range*, the possibly empty set of meanings of a grammatical term. On this conception, saying that a grammatical term has a semantic range does not imply that that term has a unique meaning or even that it has a meaning. By employing the notion of a semantic range we can adjust Hodges' compositionality formulation in order to end up with something quantificationally more innocent. We will assume that the interpretation function, which we will call $\mu'$ to mark the change, assigns semantic ranges instead of meanings to grammatical terms, and that the semantic ranges are drawn from the power set of the set of meanings. The semantic range of an expression is assumed to be the union of the semantic ranges of its structural analyses. Since saying that an expression has a certain semantic range does not entail that it is meaningful we can now simplify the framework, without apparent loss, by making the interpretation function total.[16] This means that the new compositionality formulation – which we will dub $C^{FE}$ – can be simplified in the following way.

---

[15] See, for instance, Pelletier (2000) for the claim that compositionality is incompatible with non-lexical non-structural ambiguity.

[16] It is sometimes remarked that one reason for wanting the interpretation function to be partial is so that a distinction between grammaticality and meaningfulness can be retained. (See for instance Pagin and Westerståhl (forthcoming)). This distinction can still be made in the present context (even though the interpretation function is total) since the grammatical meaningless expressions will be those that are assigned empty semantic ranges.

A semantic theory accords with $C^{FE}$ iff

there is a function $r$ such that for every complex term $s=\sigma(e_0,\ldots, e_{n-1})$,

$$\mu'(s) = r(\sigma, \mu'(e_0),\ldots, \mu'(e_{n-1})).^{[17, 18]}$$

If no extra quantitative restrictions are imposed on the members of the semantic ranges, we speak of *thin* $C^{FE}$. Corresponding to the Russellian interpretation of $C$ is *thick* $C^{FE}$, which adds the requirement that all semantic ranges contain exactly one meaning. Corresponding to other quantitative restrictions there exist other principles.[19]

Like Hodges' original formulation, $C^{FE}$ is compatible with $t_2$.

| $t_2$ | Syntax: | | Semantics: | |
|---|---|---|---|---|
| | A | $\rightarrow$ 'black' | $\mu$('black') | $= m_1$ |
| | N | $\rightarrow$ 'dog' | $\mu$('dog') | $= m_2$ |
| | NP | $\rightarrow$ A·N | $\mu$('black dog') | $= m_3$ |

---

[17] 'FE' in '$C^{FE}$' is short for 'Function-in-Extension'.

[18] There is a certain permissiveness in the formulation of $C$ which is – appropriately – not carried over to the formulation of $C^{FE}$. The expressions 'dog party' and 'party dog' have parts with the same meaning. In addition, they are both noun-noun compounds, yet they have different meanings. Still, normal interpretations of compositionality would not treat 'party dog' and 'dog party' as a counter example to compositionality since, the order of the parts having the two meanings is different. This difference is captured by $C^{FE}$ and by Hodges' formulation since the function $r$ qua an ordinary function is permitted to yield different values to different permutations of arguments.

[19] One could, for instance, formulate a restriction corresponding to the quantitative restrictions in Hodges' formulations. Note though that adding this requirement to $C^{FE}$ does not make this compositionality variant equivalent to Hodges' formulation. $C^{FE}$ 'sees' meaningless complex expressions, but Hodges' version does not.

One could also formulate a quantitative restriction corresponding to a *generic* reading of $C$. Consider the sentence 'The leg of an ant is very strong.' where this kind of reading comes very naturally. The salient reading of the sentence is not such that that it entails that ants only have one leg or that all ants have legs. Instead it says something about the legs of the ants that have legs – that these are very strong. If this kind of reading of $C$ is made more explicit we end up with something like 'If something is a meaningful complex expression then all of its parts have unique meanings and its meaning is determined by the meanings of its parts and its mode of combination.'

The theory is framed in terms of meanings and not semantic ranges. But by collecting the meanings of expressions into sets, the theory entails that 'black' has the semantic range $\{m_1\}$, 'dog' the semantic range $\{m_2\}$ etc. The application of $C^{FE}$ works more or less in the same way as the previous compositionality formulation; there is only one complex grammatical term (that corresponding to 'black dog') so any function containing the pair $<<NP, \{m_1\}, \{m_2\}>, \{m_3\}>$ satisfy the requirements for $r$. Hence $t_2$ is in accord with $C^{FE}$. (Note that this is so even though no rule is given that describes how the meaning of 'black dog' can be derived from the meanings of 'black' and 'dog').

It bears emphasis that the introduction of semantic ranges is not meant to suggest that there are less homonyms than we think that there are. By using thin formulations of compositionality I allow for the possibility that grammatical terms have more (or less) than one meaning, but I do not thereby want to suggest that it is *generally* the case that grammatical terms have more or less than one meaning. It says nothing, for instance about, whether, for instance, 'bank' corresponds to a term with multiple meanings, or whether it corresponds to multiple terms all of which have unique meanings. In many cases it might be attractive to assume that the latter option is correct.

Both Hodges' formulation of compositionality and $C^{FE}$ not only makes 'is determined by' more precise, they also make 'parts' more precise. In particular they explicate 'parts' in terms of '*immediate parts*', rather than '*ultimate parts*'.[20] The 'immediate parts' explication is stronger than the corresponding 'ultimate parts' explication, but there is an important reason to adopt it anyway. If we explicate compositionality in terms of the meanings of ultimate parts it is natural to follow suit and define 'mode of composition' as 'total mode of composition', i.e. the exact structure of the term rather than just its topmost connections (e.g. corresponding to the term underlying 'big

---

[20] In fact, as discussed by Pagin and Westerståhl (forthcoming), one can formulate compositionality in terms of parts at different levels of immediacy; immediate parts, second-level parts, third-level parts etc. One problem with formulating compositionality in terms of intermediate levels is that not all complex terms have parts at these levels and such an account, which is already complicated to formulate, would have to be complemented with a treatment of the exceptions.

We can also note that all these readings of 'parts' assume a 'proper part' reading. Compositionality would be trivial on more lenient readings. The meaning of a complex expression obviously determines itself.

black dog' would not be the A·NP-mode of composition but the A·A·N mode of composition). This seems implausible when we turn to explanations of speakers' ability to understand novel expressions, because it makes it likely that when a speaker is confronted with a novel expression, then the relevant mode of composition will often be novel to the speaker as well. This will make the speakers ability to generalize to novel but obviously familiar expressions much more limited than it actually is. For instance, a speaker understanding (4) and (5) will also generally be able to understand (6) and (7) although all of these expressions correspond to different total modes of composition.

(4)    the father of John
(5)    the father of the father of John
(6)    the father of the father of the father of John
(7)    the father of the father of the father of the father of John

So it seems reasonable to explicate compositionality in terms of immediate parts rather than in terms of ultimate parts.

## 2.2 Triviality Tout Court

Is $C^{FE}$ a trivial constraint on a semantic theory? It is not trivial *tout court*: there are theories which are not in accordance with it. $t_3$, for instance, is one such theory.

| $t_3$ | Syntax: | | Semantics: | |
|---|---|---|---|---|
| | V | → 'seeks' | $\mu$('seeks') | $= m_1$ |
| | N | → 'dogs', | $\mu$('dogs') | $= m_2$ |
| | | 'canines' | $\mu$('canines') | $= m_2$ |
| | VP | → A·N | $\mu$('seeks dogs') | $= m_3$ |
| | | | $\mu$('seeks canines') | $= m_4$ |

Even though the semantic ranges of the parts of the terms corresponding to 'seeks dogs' and 'seeks canines' are the same ($<\{m_1\}, \{m_2\}>$), and the two terms have the same mode of combination (they are both VP's), they have different meanings. This means that there is no function $r$ which maps the mode of combination of a complex term and the meanings of the parts of that term to the meaning of

that term. In order to yield the right meaning for 'seeks dogs' $r$ would have to contain the member <<VP, $\{m_1\}$, $\{m_2\}$>, $\{m_3\}$>, and in order to yield the right meaning for 'seeks canines' $r$ would have to contain the member <<VP, $\{m_1\}$, $\{m_2\}$>, $\{m_4\}$>, but this means that $r$ is not a function since it is not single valued.

So $C^{FE}$ is not trivial *simpliciter*. But it might still be the case that $C^{FE}$ is trivial in some other sense. Some of the literature pertaining to this question suggests, for instance, that versions of compositionality similar to $C^{FE}$ might be trivial in the sense that *for each non-compositional theory there exists another compositional theory which is, at least for some intents and purposes, equivalent to the first theory*. Other literature suggest that versions of compositionality similar to $C^{FE}$ might be trivial *due to the existence of reasonable assumptions about semantic theories that entail these versions*. This would make $C^{FE}$ trivial relative to the *reasonable* semantic theories. These two ideas will be explored in the next two sections.

## 2.3 *τ-Triviality*

If it turned out that for each *non-compositional* semantic theory there existed a *compositional* semantic theory which was in some appropriate sense *equivalent* to the first theory, there is a clear sense in which compositionality would be a trivial constraint on semantic theories (even if it would not be trivial *tout court*). Let's refer to a constraint which is trivial in this way as being *τ-trivial*.

τ-triv.    A constraint $q$ on semantic theories is *τ-trivial* iff for each theory not in accordance with $q$ there exist another equivalent theory which is in accordance with $q$.

Depending on what sense of 'equivalent' that is invoked, it might be the case that if compositionality turned out to be *τ*-trivial, it could turn out that the informative potential of compositionality – its ability to inform us about what meanings are – thereby would be compromised. For instance, if compositionality turned out to be *τ*-trivial in the sense that there always exists, for each non-compositional theory, another compositional theory which is

equivalent to the first theory in the sense that *the two theories agree completely with respect to their meaning assignments* then it might not be the case that we can learn anything about what meanings are by requiring that a semantic theory should be compositional: compositionality could never be used to rule out a semantic theory *qua* its meaning assignment.

Standards of equivalence for semantic theories are interest relative; what counts as two theories being equivalent relative to one set of questions and one set of background assumptions, will not be the same relative to other sets of questions and other sets of background assumptions. So whether $C^{FE}$ is $\tau$-trivial in the present context depends on the concerns and assumptions of the present inquiry. Since semantic theorizing, as described in the previous chapter, is concerned with the correctly describing languages and speakers' linguistic competencies, the standard of equivalence should depend on what are the facts of the matter pertaining to languages and linguistic competences. Since a language is here are understood in terms of an ordered pair consisting of a partial syntactic algebra and an interpretation function, it follows that there is a fact of the matter concerning which expressions contain which other expressions as parts, and which meanings that are assigned to which expressions. It was also assumed that that there are corresponding facts of the matter about a speaker's linguistic competence. This means that in order for two semantic theories to be *empirically* equivalent, they should at least be identical in terms of their assignments of syntactic structure and meaning. Or at least, since we do not have any data concerning what the exact meaning of an expression is but only intuitions about synonymy, ambiguity etc., in order for two theories to be empirically equivalent, they should at least share structure assignments, and be isomorphic with respect to their interpretation functions.

However, if we assume this minimal standard of equivalence, it follows that if two semantic theories are equivalent, they are also the same with respect to whether they accord with $C^{FE}$. Since $C^{FE}$ by definition is only concerned with the structural relations between expressions and how meanings distribute over these expressions, any two theories which are the same with respect to structure assignments and are isomorphic with respect to their interpretation functions are either both in accordance with or not in accordance with $C^{FE}$. Hence,

$C^{FE}$ is not $\tau$-trivial in the present context (given the suggested standard of equivalence).

The following three subsections will be concerned with mathematical proofs that have been used in order to argue that compositionality is in some sense an empty or trivial condition.[21] The discussion of each proof will be used to illustrate exactly how the relevant arguments, if construed as arguments to the effect that $C^{FE}$ is $\tau$-trivial, can be resisted along the previous lines. If my arguments are correct these proofs do not give us any reason to think that $C^{FE}$ is $\tau$-trivial *relative to the present inquiry and its standard of equivalence*. This is of course compatible with the view that $C^{FE}$ is $\tau$-trivial relative some other inquiry (and its standard of equivalence).[22] Had, for instance, 'language' been understood in a weaker sense (e.g. something closer to Lewis's original conception) and the standard of equivalence would have been adjusted accordingly, then $C^{FE}$ would indeed have been $\tau$-trivial. The discussions that follow will also contain other (ultimately unsuccessful) attempts to extract from the relevant proofs reasons for why a principle like $C^{FE}$ is trivial.

### 2.3.1 Zadrozny's Theorem

In his 'From Compositional to Systematic Semantics', Wlodek Zadrozny proves the following theorem, which he thinks demonstrates that 'the standard definition of compositionality is formally vacuous' (Zadrozny 1994: 329).

> ZADROZNY'S THEOREM: Let $M$ be an arbitrary set. Let $A$ be an arbitrary alphabet. Let '$\cdot$' be a binary operation, and let $E$ be the set closure of $A$ under '$\cdot$'. Let $\mu$: $E \rightarrow M$ be an arbitrary function. Then there is a set of functions $F$ and a unique map $f$: $E \rightarrow F$ such that for all $s, t \in E$, $f(s.t) = f(s)(f(t))$, and $f(s)(s) = m(s)$.

---

[21] Westerståhl (1998) drew my attention to these proofs.

[22] It follows immediately from a version of compositionality not being trivial *simpliciter*, and not being necessarily false, that if the standard of equivalence is completely deflated (and any two theories are deemed equivalent) that this version of compositionality is $\tau$-trivial. But this follows from there existing both compositional and non-compositional theories and from deflating the standard of equivalence. No mathematical proof is needed.

Assume that $M$ is a set of meanings, that $A$ is the set of atomic expressions, that '·' is the concatenation operation, that $E$ is the set of expressions generated from $A$ by the concatenation operation, and that $\mu$ is an interpretation function. On this interpretation, the theorem can be seen as guaranteeing, for each non-compositional semantic theory, the existence of a certain compositional semantic theory – what will be referred to as a *z-theory*.[23] Corresponding to the non-compositional theory $t_4$, for instance, there exist the compositional *z*-theory $t_5$.

$t_4$  *Syntax:*                      *Semantics:*
      V   → 'seeks'               $\mu$('seeks')          $= m_1$
      N   → 'dogs',               $\mu$('dogs')           $= m_2$
            'canines'             $\mu$('canines')        $= m_2$
      VP  → A·N                   $\mu$('seeks dogs')     $= m_3$
                                  $\mu$('seeks canines')  $= m_4$


$t_5$  *Syntax:*
      V   → 'seeks'
      N   → 'dogs',
            'canines'
      VP  → A·N

      *Semantics:*
      $f$('seeks')        $= \{\ \langle$'seeks', $m_1\rangle\}$
      $f$('dogs')         $= \{\ \langle$'dogs', $m_2\rangle$,
                               $\langle f$('seeks'), $f$('seeks dogs')$\rangle\}$
      $f$('canines')      $= \{\ \langle$'canines', $m_2\rangle$,
                               $\langle f$('seeks'), $f$('seeks canines')$\rangle\}$
      $f$('seeks dogs')   $= \{\ \langle$'seeks dogs', $m_3\rangle\}$
      $f$('seeks canines') $= \{\ \langle$'seeks canines', $m_4\rangle\}$

The *z*-theory is compositional in two senses. First, it is in accordance with $C^{FE}$ (which follows from its meaning assignment being *hyper-distinct*, i.e. such that formally distinct expressions are assigned

---

different meanings). Second, the *f*-value of each complex expression can be derived from the application of the *f*-value of one of its parts to the *f*-value of the other. For instance, *f*('seeks canines') = *f*('canines')(*f*('seeks')).

Does Zadrozny's theorem give us a reason to think that $C^{FE}$ is $\tau$-trivial? It does, if $t_4$ and $t_5$ are equivalent. They are not equivalent in the present context since the two theories embody different meaning assignments: where 'dogs' and 'canines' are assigned the same meaning by $t_4$ they are assigned different meanings by $t_5$. On the conception of 'language' and 'linguist competence' adopted in this essay, i.e. on the present construal of the two objects semantic theories are about, there is a fact of the matter whether two expressions have the same meaning or not. And it is not consequential if there for each non-compositional theory, exists another compositional theory, *if there is no guarantee that the latter theory will be correct if the former is correct*. Since the two theories are not empirically equivalent there is no such guarantee. Note that $t_4$ and $t_5$ do not only assign different meanings to the expressions, the two theories are not even isomorphic with respect to their interpretation functions. So $t_4$ and $t_5$ are not equivalent, and, since this holds in general for a non-compositional theory and its *z*-theory, Zadrosny's proof has not provided us any reason to think that $C^{FE}$ is $\tau$-trivial.[24] The same consequence follows even if we take a less realistic stance – by assuming less about what there is a matter of fact about – and hold instead that there is certain data that we want a semantic theory to account for (such as synonymy judgments) or certain explanations that it should provide. There is no guarantee that semantic theories with different meaning assignments do this equally well. So there is no guarantee that a compositional alternative to a certain non-compositional theory makes use of appropriate meaning assignments.

There are passages in which Zadrozny seems to suggest that the two theories are equivalent.[25] But it seems likely that he would consider,

---

[24] This holds in general since a semantic theory cannot fail to accord with $C^{FE}$ if its interpretation function is hyper-distinct, but all *z*-theories will always have hyper-distinct interpretation functions since the *f*-values they assign are all defined for exactly one expression, the expression of which they are an *f*-value.

[25] Here is one passage where he describes matters in this way.

for instance, requiring that a theory induce certain semantic relations, to amount to imposing an extra condition on the semantics, and thus that he has some other standard of equivalence in mind.[26] Be that as it may, it is still the case that given that we make this requirement, there is no guarantee that there is an equivalent compositional theory for each non-compositional theory. Hence, Zadrozny's proof does not show that $C^{FE}$ is $\tau$-trivial in the present context.[27]

---

> In view of the above theorems, any semantics is equivalent to a compositional semantics, and hence it would be meaningless to keep the definition of compositionality as the existence of a homomorphism from syntax to semantics without imposing some conditions on this homomorphism.
>
> (Ibid: 334)

The other theorems that Zadrozny makes reference to in the quote are refinements of his first theorem. He shows i) that the theorem goes through even if the concatenation operation is not total and ii) that the meanings of the original theory can be recovered in a different way than having to apply the value of *f* for a certain expression to that expression. It is very unclear in what sense the two theories are supposed to be equivalent. Perhaps his other remarks in terms of one theory's *encoding* the other theory captures his intentions better. See below.

[26] This seems to be what Lappin and Zadrozny maintains in response to the criticism – which is similar to mine – offered by Westerståhl (1998) and Kazmi and Pelletier(1998) on Zadrozny's original article.

[27] Trying to escape the trivializing consequences of Zadrozny's result by arguing against hyper-distinctness is the wrong approach to the problem. It seems that Werning (2005) suggests a ban on hyper-distinct meanings in order to avoid having to say that compositionality is trivial since the identity function on expressions can be used as a meaning function (and the identity function on expressions always result in a compositional semantic theory if compositionality is assumed to be something like Hodges' formulation). However, the reason for why the identity function is not a suitable semantics is that it does not respect our pre-theoretic semantic intuitions and not that it is hyper-distinct. A general ban against hyper-distinct meanings will not avoid the trivial constructing of compositional semantics if these intuitions are disregarded. Consider Werning's ban. It is understood along the lines of the following definition of non-hyper-distinctness; given a set of grammatical terms *E*, a meaning function $\mu$ with domain *E* is called *non-hyper-distinct* if there are grammatical terms *s*, $t \in E$ such that $s \neq t$ and $\mu(s) = \mu(t)$. A ban along the lines of this definition does not avoid the trivial generation of compositional meaning functions since there exists the possibility of defining a trivial non-hyper-distinct function. Consider the following function.

$f(x) = m_1$    for all $x \in E$.

The difficulty with approaching the issue in Werning's manner is that it too narrowly construes the problem. Hyper-distinctness is one way to realize compositionality but there are many others. All theories according to which all expressions mean the same

In other places in his essay, Zadrozny states his conclusion differently. There he holds that he has shown that 'any semantics can be *encoded* as a compositional semantics' (Ibid: 329, emphasis added). This seems to pertain to the fact that the meaning assignment of any non-compositional theory is derivable from the meaning assignment of its *z*-theory. Given $t_4$ and $t_5$, for instance, $\mu$('seeks') = $f$('seeks')('seeks'), $\mu$('dogs') = $f$('dogs')('dogs'), etc. So for a *z*-theory each expression *e* really has two semantic values; its *$\mu$-value*, i.e. the value yielded by $f(e)(e)$, and its *f-value*, i.e. the value yielded by $f(e)$. Given that *$\mu$-values* can be recovered from *f-values*, it follows from Zadrozny's theorem that for each non-compositional theory, there is a compositional theory that encodes that theory. But does this entail that $C^{FE}$ is, in some sense, a trivial constraint?

It does not in the present context, as the following line of thought shows. The topic of the present inquiry ultimately concerns the extent to which compositionality can contribute to answering the question 'What are meanings?'. This question is raised in the context of certain semantic data, in particular, distributional information about meanings, information about which expressions have the same meanings, which expressions have different meanings, which expressions that does not have any meanings, etc.[28] Inquiring into whether compositionality can tell us anything about meanings thus really amounts to inquiring into what we can learn by imposing the compositionality constraint *on things already constrained by the semantic data*. But this means that the meanings invoked by the compositionality constraint should be the same as those which are required to conform to the semantic data. If they are not, then compositionality cannot be used to inform us about meanings in the sense of 'whatever it is that is invoked by statements of synonymy, meaninglessness etc.'.

However, the aforementioned fact – that for each non-compositional theory, there exists a compositional theory that encodes that theory – does not mean that there always exists a semantic theory which is compositional and accords with the

thing are also trivially compositional, and such theories are not hyper-distinct. They are hyper *in*distinct.

[28] This is of course not to say that we have *complete* information about which expressions have the same meanings or which expressions have different meanings etc. only that we have *some* information of this kind.

semantic data *with respect to the same meanings.* In the encoding theories there is a clear division of labor between accounting for compositionality and accounting for the semantic data. If we assume, for the sake of the argument, that $t_4$ accords with the semantic data, it is with respect to the $\mu$-values that $t_5$ accords with the semantic data, but it is with respect to its *f*-values that $t_5$ is compositional. Hence, the two constraints that should apply to the same thing, applies to different things.

Here is another way to put the same point. We are concerned with what the constraint that the *meaning* of a complex expression is determined by the *meanings* of its parts and its mode of composition can tell us about whatever is invoked in holding, for instance, that two expressions have the same *meaning*. But the *z*-theories that issues from Zadrozny's theorem are not such that they provide semantic entities that can be the interpretation of all three of these occurrences of 'meaning'. Instead, the *z*-theories embody an equivocation; with respect to being compositional, *f*-values are considered to be meanings, and with respect to being in accordance with the semantic data, the $\mu$–values are considered to be meanings. If we, in order to avoid the equivocation, hold instead that what is relevant for compositionality *and* for accounting for the semantic data are the *f*-values, then *z*-theories will generally be incorrect. If we instead hold that what is relevant for compositionality *and* for accounting for the semantic data are the $\mu$-values then *z*-theories will only be compositional if they encode compositional theories.

So the existence of compositional theories that encode non-compositional theories does not contribute towards compositionality being a trivial constraint. The *z*-theories of non-compositional theories, like the non-compositional theories that they encode, fail to provide entities that both accord with the semantic data and are compositional. So, to reiterate a point made a few paragraphs back, if we assume that a semantic theory must accord with certain semantic data, Zadrozny's theorem does not provide us with a reason to think that compositionality is a trivial constraint.

### 2.3.2 Janssens Theorem

Consider next a theorem proved by Theo Janssen, which has been used by Herman Hendricks for instance, to argue that

compositionality 'is not an empirical principle, but a methodological one' (Hendriks 1993: 137).[29]

> JANSSEN'S THEOREM: Suppose $\mu: E \rightarrow M$, where $E$ is any recursively enumerable set of strings and $M$ is arbitrary. Then there is a partial algebra $\mathbf{A} = \langle A, F \rangle$ with $E \subseteq A$ and $F$ is a set of operations on $A$, a partial algebra $\mathbf{B} = \langle B, G \rangle$ with $M \subseteq B$ and $G$ is a set of operations on $B$, and a function $h$ from $A$ onto $B$ which is compositional, and such that for all $e \in E$, $h(e) = \mu(e)$.[30]

It follows from Janssen's theorem that corresponding to any non-compositional theory like $t_4$ there exist compositional theories which are just like $t_4$ with respect to their interpretation functions. $t_6$ is one such theory.

| $t_4$ | Syntax: | | Semantics: | |
|---|---|---|---|---|
| | V | $\rightarrow$ 'seeks' | $\mu$('seeks') | $= m_1$ |
| | N | $\rightarrow$ 'dogs', | $\mu$('dogs') | $= m_2$ |
| | | 'canines' | $\mu$('canines') | $= m_2$ |
| | VP | $\rightarrow$ A·N | $\mu$('seeks dogs') | $= m_3$ |
| | | | $\mu$('seeks canines') | $= m_4$ |

| $t_6$ | Syntax: | | Semantics: | |
|---|---|---|---|---|
| | V | $\rightarrow$ 'seeks' | $\mu$('seeks') | $= m_1$ |
| | $N_1$ | $\rightarrow$ 'dogs' | $\mu$('dogs') | $= m_2$ |
| | $N_2$ | $\rightarrow$ 'canines' | $\mu$('canines') | $= m_2$ |
| | $VP_1$ | $\rightarrow$ A·$N_1$ | $\mu$('seeks dogs') | $= m_3$ |
| | $VP_2$ | $\rightarrow$ A·$N_2$ | $\mu$('seeks canines') | $= m_4$ |

We remember that the first theory is not in accordance with $C^{FE}$ since there is no function $r$ which maps the mode of combination of a complex term and the meanings of the parts of that term to the meaning of that term. In order to yield the right meaning for 'seeks

---

[29] The original proof is due to Janssen (1986). Janssen himself is careful not to draw any strong conclusions from the proof about compositionality being trivial.

[30] The statement of the theorem reproduced here is a simplified version given by Westerståhl (1998).

dogs' $r$ would have to contain the member <<VP, $\{m_1\}$, $\{m_2\}$>, $\{m_3\}$>, and in order to yield the right meaning for 'seeks canines' $r$ would have to contain the member <<VP, $\{m_1\}$, $\{m_2\}$>, $\{m_4\}$>, but this means that $r$ is not a function since it is not single valued. But $t_5$ is in accordance with $C^{FE}$ since a set $r$ can contain the member <<VP$_1$, $\{m_1\}$, $\{m_2\}$>, $\{m_3\}$>, and the member <<VP$_2$, $\{m_1\}$, $\{m_2\}$>, $\{m_4\}$>, and still be single valued and hence a function.

   Does Janssen's theorem give us any reason to think that $C^{FE}$ is $\tau$-trivial? It does if $t_4$ and the theories like $t_6$ are equivalent. They are not equivalent in the present context, since the two theories embody different structure assignments; where 'seeks dogs' and 'seeks canines' are assigned the same mode of composition by $t_4$, they are assigned different modes of composition by $t_6$. But on the conception of 'language' and 'linguist competence' adopted in this essay, i.e. on the present construal of the two objects semantic theories are about, there is a fact of the matter whether two expressions have the same mode of composition or not. And – to reiterate the point made in the previous sub-section – it is not consequential if there for each non-compositional theory, exists another compositional theory, *if there is no guarantee that the latter theory will be correct if the former is.* Since the two theories are not empirically equivalent there is no such guarantee. Note that the syntactical adjustments needed to transform $t_4$ into $t_6$ are necessary in general in order for one theory to respect $C^{FE}$ and the other to violate it, as the two theories have identical interpretation functions; if two theories are identical with respect to syntactical structures and meaning assignments then they are identical with respect to whether they accord with $C^{FE}$. So $t_4$ and $t_6$ are not equivalent, and Janssen's theorem gives us no reason to think that $C^{FE}$ is $\tau$-trivial. Again, the same consequence follows even if we take a less realistic stance and hold instead that there is certain data that we want a semantic theory to account for or certain explanations that it should provide. For instance, we might take the fact that 'dogs' and 'canines' can be substituted for each other in all contexts while preserving grammaticality to indicate that they belong to the same syntactic category. There is no guarantee that semantic theories with different syntactic components account for the syntactic data equally well. So there is no guarantee that a compositional alternative to a certain non-compositional theory makes use of an appropriate structure.

Had we been working with a different notion of a language something that corresponds more closely to Lewis's original conception of a language (like a function from expressions to meanings), for instance, then Jansson's theorem would have proved that $C^{FE}$ was $\tau$-trivial if we had adjusted our standard of equivalence accordingly. There would have been no fact of the matter of whether a language had one structure rather than another and semantic theories with identical interpretation functions would have been equivalent. Perhaps it is something like this Hendricks had in mind. In these circumstances, two empirically equivalent theories could differ with respect to whether they accord with $C^{FE}$, thereby suggesting that the principle is not empirical.

### 2.3.3 The Recursion Theorem

Consider finally *the recursion theorem* from which Johan van Benthem concludes that 'by itself, compositionality provides no significant constraint on semantic theory' (van Benthem 1984: 57).

> THE RECURSION THEOREM: Assume that a set $E$ is freely generated from a set $A$ by a set of functions $\Sigma$ s.t. for each function $\sigma_i \in \Sigma$, $\sigma_i: E^n \to E$ for some $n$, which is the arity of $\sigma_i$. Further assume that $M$ is a set, that $\mu: A \to M$, and that $\Delta$ is a set of functions s.t. for each function $\delta_i \in \Delta$, $\delta_i: M^n \to M$ for some $n$, which is the arity of $\delta_i$. Finally, assume that $h: \Sigma \to \Delta$, and that $h$ maps functions on functions of the same arity.
>
> Then there is a unique function $f: E \to M$ s.t.
>
> i)    For each $e$ in $A$, $f(e) = \mu(x)$
> ii)   For each function $\sigma_i \in \Sigma$ of arity $n$ and each sequence $e_1, \ldots, e_n$ of members of $E$,
>      $f(\sigma_i(e_1, \ldots, e_n)) = h(\sigma_i)(f(e_1), \ldots, f(e_n))$. [31]

---

[31] The name of the theorem is from Enderton (2001). The statement of the theorem is a generalization of that given by Enderton which corresponds to the informal statement given by van Benthem. Cfr:

The relevant interpretation of the terms are as before; assume that $E$ is the set of expressions, that $A$ is the set of atomic expressions, that each function $\sigma_i \in \Sigma$ corresponds to a syntactical rule, that $M$ is a set of meanings, and that each function $\delta_i \in \Delta$ corresponds to a semantic rule. That the set $E$ must be *freely generated* from the set $A$ with respect to the functions in $\Sigma$ amounts to the requirement that i) $E$ is generated from $A$ by the functions in $\Sigma$, ii) the functions of $\Sigma$ are one-to-one, iii) the ranges of the functions of $\Sigma$ are disjoint from $A$, and iv) the ranges of the functions of $\Sigma$ are disjoint from each other.

It being possible to extend the interpretation function of a semantic theory $t$ to a homomorphism between its syntactic and semantic algebra amounts to it being possible to extend $t$ to a compositional theory.[32] The recursion theorem thus guarantees that certain semantic theories can be extended to compositional theories.[33]

---

> Suppose that some Algebra <$A$, $O$> with syntactic operations has been chosen, representing disambiguated readings of linguistic items, and that, assuming the severest possible semantic constraint, some meaning algebra <$B$, $M$> is prescribed in advance. Our task is then to see if there exists a homomorphism from the former to the latter algebra. (This is the algebraic version of compositionality.) In an extreme case, even the connections between syntactic and semantic operations in $M$ are fixed, and we have at least a respectable problem. But usually, such a connection is not prescribed: indeed any polynomially definable operation on $B$ may be assigned, in principle to any operation in $O$[...]Finally, even this presentation is too restricted, for actually, the syntactic algebra is *free* (it is freely generated by the basic lexical items). What this algebraic assertion amounts to is this. The construction of <$A$, $O$> is such that, given any connection of operations in $O$ with semantic operations of the same number of arguments, an arbitrary map from basic lexical items to suitable semantic entities will be uniquely extendable to a homomorphism as required. Thus we are entitled to conclude that *by itself, compositionality provides no significant constraint upon semantic theory.*
>
>                              (van Benthem 1984: 57, emphasis in original)

[32] The exact connection between the compositionality explications like $C^{FE}$ and those in terms of a homomorphism between syntax and semantics was discussed in footnote 12.

[33] Note that 'semantic theory' is used differently in this passage compared to the rest of the essay. The operations of the algebras are assumed to be total and the theories include semantic algebras.

The theorem does not guarantee that all semantic theories with partial interpretation functions can be extended to compositional theories though; any theory embedding $t_4$ for instance, cannot be extended into a semantic theory. So if we are in a situation where meanings of complex expressions matter, for instance, if we are involved in semantic theorizing as characterized in the previous chapter, then the recursion theorem gives us no reason to think that compositionality is not a substantial constraint. It is hard to see exactly what van Benthem was after when he maintained that 'by itself, compositionality provides no significant constraint on semantic theory' (Ibid.). For if respecting meaning assignments is considered as an additional constraint on a semantic theory (not part of the conditions of equivalence), then no proof is needed in order to show that compositionality is $\tau$-trivial: just rearrange the meanings of a non-compositional theory until it becomes compositional.

The point made against the previous argument can also be made in response to the following argument to the effect that $C^{FE}$ is $\tau$-trivial. If the syntactic algebra of a non-compositional theory is free then there exist other compositional semantic theories which correspond to the original theory by i) having the same syntax, and ii) assigning the same meanings to atomic expressions. Corresponding to the non-compositional theory $t_4$, for instance, there exist several compositional alternatives, one of which is $t_7$.

---

Note also that requiring that $E$ is freely generated from $A$ is crucial for the proof to go through since it rules out counterexamples like the following.

| | | |
|---|---|---|
| $A = \{a, b\}$ | $M = \{1, 2…\}$ | $\mu = \{< a, 1>,$ |
| $E = \{a, b, c…\}$ | | $< b, 2>\}$ |
| $\Sigma = \{\sigma_i \{< a, b>,…\},…\}$ | $\Delta = \{\delta_i \{< 1, 3>,…\},…\}$ | $h = \{<\sigma_i, \delta_i >,…\}$ |

Given this situation, $\mu$ cannot be extended to a homomorphism $f$ which agrees with $\mu$ for all members of $A$, since, $\mu(\sigma_i(a)) = 2$ but $h(\sigma_i)(\mu(a)) = 3$ and hence $f(\sigma_i(a)) \neq h(\sigma_i)(f(a))$. However, if $E$ is freely generated from $A$ (by the functions of $\Sigma$) then the ranges of the functions of $\Sigma$ must be disjoint from $A$ and the counterexample can be avoided (in the example ran$(\sigma_i) \cap A = \{b,…\}$ which violates this condition). The other conditions included in the definition of 'freely generated' avoid other counterexamples.

| $t_4$ | *Syntax:* | | *Semantics:* | |
|---|---|---|---|---|
| | V | $\rightarrow$ 'seeks' | $\mu$('seeks') | $= m_1$ |
| | N | $\rightarrow$ 'dogs', | $\mu$('dogs') | $= m_2$ |
| | | 'canines' | $\mu$('canines') | $= m_2$ |
| | VP | $\rightarrow$ A·N | $\mu$('seeks dogs') | $= m_3$ |
| | | | $\mu$('seeks canines') | $= m_4$ |

| $t_7$ | *Syntax:* | | *Semantics:* | |
|---|---|---|---|---|
| | V | $\rightarrow$ 'seeks' | $\mu$('seeks') | $= m_1$ |
| | N | $\rightarrow$ 'dogs', | $\mu$('dogs') | $= m_2$ |
| | | 'canines' | $\mu$('canines') | $= m_2$ |
| | VP | $\rightarrow$ A·N | $\mu$('seeks dogs') | $= m_3$ |
| | | | $\mu$('seeks canines') | $= m_3$ |

Does the existence of a theory like $t_7$ give us any reason to think that $C^{FE}$ is $\tau$-trivial? It does if $t_4$ and theories like $t_7$ are equivalent. They are not equivalent in the present context; the two theories embody different meaning assignments, where 'seeks dogs' and 'seeks canines' are assigned different meanings by $t_4$ they are assigned the same meanings by $t_7$. But, on the conception of 'language' and 'linguist competence' adopted in this essay, i.e. on the present construal of the two objects semantic theories are about, there is a fact of the matter whether two expressions have the same meaning or not. And – to reiterate the point made in the previous two sub-sections one more time – it is not consequential if there, for each non-compositional theory, exists another compositional theory, *if there is no guarantee that the latter theory will be correct if the former is.* Since the two theories are not empirically equivalent there is no such guarantee. Note that $t_4$ and $t_7$ not only assign different meanings to the expressions, the two theories are not even isomorphic with respect to their interpretation functions. So $t_4$ and $t_7$ are not equivalent, and, more generally, we have been given no reason to think that $C^{FE}$ is $\tau$-trivial. Again, the same consequence follows even if we take a less realistic stance and hold instead that there is certain data that we want a semantic theory to account for or certain explanations that it should provide. There is no guarantee that semantic theories with different meaning assignments do this equally well. So there is no guarantee that a compositional alternative to a certain non compositional theory makes use of appropriate meaning assignments.

Naturally, if the equivalence conditions invoked by '$\tau$-trivial' are changed to i) having the same syntax, and ii) assigning the same meanings to atomic expressions, then we do have a reason to think that $C^{FE}$ is trivial. However, these are not the equivalence conditions stemming from the construal of semantic theorizing given here, and it is unclear under which circumstances such equivalence conditions would be appropriate.

## 2.4 $\varphi$-Triviality

If i) there exists some constraint $\varphi$ on semantic theories such that there exists good reasons to think that only a semantic theory in accordance with $\varphi$ is likely to be correct, and it is the case that ii) a theory being in accord with $\varphi$ entails that it is compositional, then there is a sense in which compositionality is a trivial constraint; compositionality would be trivial relative to all the theories in accord with $\varphi$ – the theories that are likely to be correct. Let's refer to a constraint which is trivial in this way as being $\varphi$-*trivial*.

$\varphi$-triv.     A constraint $q$ on semantic theories is $\varphi$-*trivial* with respect to a constraint $\varphi$ on semantic theories iff $\varphi$ entails $q$, and there are good reasons to think that only semantic theories in accordance with $\varphi$ are likely to be correct.

If compositionality turned out to be $\varphi$-trivial, and it also turned out that $\varphi$ was not entailed by compositionality, then there is a sense in which the informative potential of the principle would be compromised. For since $\varphi$ excludes more theories than compositionality does, everything we can learn about meaning from compositionality can also be learnt from $\varphi$. So by compositionality being $\varphi$-trivial (given that $\varphi$ is not entailed by compositionality) the informative potential of compositionality becomes uninteresting; what is really interesting in this case is the informative potential of $\varphi$.

There are several general constraints that entail $C^{FE}$ and are not entailed by it in turn. However most of these do not seem to be reasonable requirements on semantic theories. For instance, $C^{FE}$ is trivial with respect to the theories where i) *all expressions have different semantic ranges*; ii) *all expressions mean the same thing*; iii) *there are no*

*complex expressions*; or iv) *all complex expressions have a unique mode of composition.* However, that $C^{FE}$ follow from the corresponding general constraints does not make $C^{FE}$ $\varphi$-trivial, since there are no good reasons to require that semantic theories must respect these constraints. That is, we have no reasons to think that only theories in accordance with these constraints are likely to be correct. In fact, imposing some of these constraints seems to violate the intuitions that we are ultimately concerned with accounting for: there are expressions that have the same semantic range ('doe' and 'female deer') and expressions that have different semantic ranges ('doe' and 'die'). In addition, there does not seem to be any good reasons for assuming that all expressions are simple or that all complex expressions have unique modes of composition.[34]

The rest of this section will be concerned with another constraint with respect to which $C^{FE}$ might be $\varphi$-trivial. The constrain follows from a claim made by Paul Horwich. In his paper 'The Composition of Meanings' Horwich develops a deflationary view of compositionality based on what we can refer to as *the constitution thesis*.[35] He succinctly describes his position in the following way.

> [T]he compositionality of meaning imposes no constraint at all on how meaning properties of words are constituted.
> (Horwich 1997: 503)[36]

---

[34] Note that the requirement on a semantic theory corresponding to the property a language must have to be learnable, is not a requirement with the respect to which $C^{FE}$ is $\varphi$-trivial since that requirement does not entail $C^{FE}$. Note also, that the stronger compositionality explications which will be given in later chapters are not requirements with the respect to which $C^{FE}$ is $\varphi$-trivial since there are, as will be argued in Chapter Three and Chapter Four, no good reasons to think that they are true. See Chapter Four for a discussion connected with learnability.

[35] In Horwich's later writings there is also a prominent, separate but related line of argument stemming from the suggestion that we should drop what he calls *the uniformity thesis*. I will concentrate on the constitution thesis here, and postpone discussion of his latter suggestion until Chapter Six where it can be more fruitfully discussed.

[36] A few pages later he puts the point in an even stronger way.

> [C]ompositionality *per se* provides absolutely *no* constraint upon, or insight into, the underlying nature of meaning.
> (Ibid: 507, emphasis in original)

The background concern of this essay – the questions of what meanings are – have been left at a somewhat intuitive level and it has not been spelt out exactly how something must be related to meanings to *be meanings*. What has been tacitly assumed is a certain form of co-extensionality; if some *m* is supposed to be the meaning of an expression *e* then those expressions that mean the same as *e* must also be assigned *m*. But no other restriction has been imposed: anything assigned to an expression is a candidate meaning. In his investigating of what constraint compositionality poses, Horwich frames the situation differently, namely, in terms of a restriction on how *meaning properties* are *constituted*. The connection between *meaning properties* and *meanings* seems to be straightforward; an expression *having a certain meaning property* (e.g. *x* means DOG) translates into it standing in a certain relation to a corresponding meaning (e.g. DOG).[37] But on Horwich's view, a meaning property *S* being *constituted* by another property *U*, is not only a matter of *S* and *U* being co-extensional. The relation obtains when '*S* and *U* are co-extensional, *and when facts about S are explained by this co-extensionality*' (Ibid: 505, fn.3, emphasis added). Horwich suggests that this kind of relation holds, for instance, between water and $H_2O$ since, i) water and $H_2O$ are co-extensional, ii) water being co-extensional with $H_2O$ explains why it boils at 100 degrees centigrade, why it is transparent etc.

With these remarks in mind, it seems that even though Horwich is concerned with a slightly different issue than we are, his claim about the compositionality constraint – that it does not put any constraint on how *the meaning properties of words are constituted* – seems to have implications for what we have been discussing in this chapter. In particular, Horwich's claim seems to entail that compositionality (in his sense) puts no constraint on *which meanings that can be assigned to*

---

The quote I've put in the main text seems to fit better with the rest of his paper so I've taken it as the point of departure in my discussion.

[37] I will follow Horwich in adopting the convention of treating a capitalized English expression as the name of its meaning. For instance, DOGS BARK = the meaning of 'dogs bark'.

Horwich also talks about the constitution of meaning *facts*. 'x means DOGS BARK' express a meaning *property*, and '"dogs bark" means DOGS BARK' express a meaning *fact*.

*words*. Things are complicated by Horwich using 'compositionality' differently than me though and we need to go over his position before being able to spell out the exact implications for $C^{FE}$.

In 'The Composition of Meanings' Horwich is primarily concerned with what assumptions are needed in order to obtain an explanation of the meaning of a sentence in terms of the meaning of its parts and its structure. He seems to suppose that the compositionality constraint just is the requirement that such an explanation be provided. In his attempt to detail the assumptions needed to meet the constraint, he suggests as one of these constraints, *the constitution thesis*, the claim that the meaning property of a complex expression consist in the property of *being the result of combining words with certain meanings into a schema with a certain meaning*. After having put forth this thesis, he concludes that it, together with lexical and structural assumptions, is sufficient to meet the compositionality constraint. The thesis is important to Horwich, partly because he thinks that it can be used to demonstrate that the machinery invoked by Davidsonian truth theories to meet the compositionality requirement has been much more complicated than what is actually needed.[38]

Horwich illustrates the constitution thesis in the following way.

> [T]he fact that 'dogs bark' means DOGS BARK […] is constituted by whatever is the fact regarding its mode of construction and the meanings of its constituents. This turns out to be the fact that the sentence results from putting words meaning what 'dog' and 'bark' mean into a schema meaning what '*n*s v' means […] Thus the meaning property
>
> x means DOGS BARK
>
> *consists in* what I shall call the "construction property":
>
> x results from putting terms whose meanings are DOG and BARK, in that order, into a schema whose meaning is *N*S V.
>
> (Ibid: 504-505, emphasis in original)

---

[38] Horwich (2001; 2005) pursues a related line of argument against a different aspect of the Davidsonian project.

The idea seems to be that while the meaning properties of simple expressions are constituted substantially by use properties, reference properties etc., the meaning properties of complex expressions are constituted trivially by the construction properties of these expressions, i.e. in virtue of being the result of combining various syntactic entities with certain meanings. Horwich generalizes his example in a straightforward manner.

> [Assume that "e" is an arbitrary complex expression and that]
>
> > "e" is the result of applying combinatorial procedure P to the primitives [some of which might be schemata] < "w1",…, "wn">
>
> My proposal is that the meaning property of 'e'– namely, "x means E"– is constituted by the construction property
>
> > x results from applying combinatorial procedure P to the primitives whose meanings are <W1,…,Wn>
>
> > > (Ibid: 507)

According to Horwich, if we assume the constitution thesis, the compositionality requirement can now be met given only lexical and structural assumptions. For given that, for instance, i) 'dog' means DOG, ii) 'bark' means BARK, iii) '*n*s *v*' means *N*S V and that iv) 'dogs bark' results from putting the terms 'dog' and 'bark' in the schema '*n*s *v*' it logically follows that v) 'dogs bark' results from putting terms meaning DOG and BARK into a schema meaning '*N*S V'. And given v) it follows from the aforementioned instance of the constitution thesis that vi) 'dogs bark' means DOGS BARK. Horwich concludes that he has, in virtue of this derivation, given an explanation of the meaning of a sentence in terms of the meaning its parts and its structure.

  Now given that the constitution thesis is adopted, it falls out as a corollary that compositionality (in Horwich's sense) can be satisfied without thereby assuming anything about the constitution of the meanings of primitives. For in order to be a part of the construction property of a complex expression, nothing at all is assumed about the constitution of the meanings of the relevant syntactic entities. When

one holds that the construction property of, say, 'dogs bark' is the property '*x* results from putting terms whose meanings are DOG and BARK, in that order, into a schema whose meaning is *N*S V.', nothing at all has been said about the meanings of 'dog', 'bark' or the schema '*ns v*', only that the syntactic entities have the meanings that they have. So given that we assume the constitution thesis, Horwich's deflationary claim follows; compositionality (in Horwich's sense) imposes no constraint at all on how meaning properties of words are constituted.[39]

Having seen this, we can now return to the question of why adopting the constitution thesis runs the risk of making $C^{FE}$ $\varphi$-trivial. Given that we assume the constitution thesis we know that no two complex expressions that have the same structure and have parts with the same meanings can have different meanings. For if two expressions differ in meaning, then we know that they must have meaning properties which are constituted differently. But then, by the constitution thesis, these expressions must either have different structures or have parts with different meanings, and this contradicts our original assumption that the two expressions were alike in these two respects. But, if such expression pairs are ruled out, then $C^{FE}$ must hold. So it follows that if there are good reasons to think that only semantic theories in accordance with the constitution thesis is likely to be correct, then $C^{FE}$ is $\varphi$-trivial (where $\varphi$ is the constitution thesis): all theories which only invoke meaning properties that are constituted in accordance with the constitution thesis are also in accord with $C^{FE}$.

However, there are good reasons to resist the constitution thesis, so Horwich's deflation of his version of the compositionality principle does not gives us good reasons to think that $C^{FE}$ is $\varphi$-trivial. First and

---

[39] This conclusion does not follow from many other accounts that might also appropriately be called deflationary. Assume that $E$ is a complex expression with two atomic parts $e_1$ and $e_2$. Consider next the account of the meaning of $E$ in terms of being 'the meaning that results from applying the meaning of $e_1$ to the meaning of $e_2$'. On such an account not all meanings of simple expressions can be non-functions. Or consider the account of the meaning of complex expressions in terms of ordered sequences of the meanings of their parts and their syntactic mode of combination. On this account, if we assume standard set theory (ZFC), and $E= e_1 e_2$ and that $\mu(e_1 e_2) = M = <\sigma, \mu(e_1), \mu(e_2)>$ then the meaning of $e_1$ or $e_2$ cannot be $M$ or something which has $M$ as an element. Or consider the account of meaning of complex expressions in terms of mereological sums of the meanings of their parts. Then given that unrestricted composition is not assumed there will be constraints on what the meanings of the expressions that are parts of the same expressions can be.

foremost, there are reasons to suspect that a theory adopting the constitution thesis will fail to accord with some of the semantic data that we have assumed that semantic theorizing is concerned with, in particular synonymy and ambiguity data. Adopting the constitution thesis in the present context would thus be inappropriate. In addition, since the constitution thesis embodies a particular thesis about how the meanings of complex expressions are constituted, it should be adopted only after comparison with alternative theses about meaning constitution, and it seems that the theories adopting the constitution thesis generally have two important shortcomings when compared with theories embodying other more substantive assumptions.

Before I spell out these arguments, I want to emphasize a subtlety in the dialectic. I'm not arguing against Horwich's view that the constitution thesis entail that his compositionality version is in his sense trivial. I'm arguing that it does not follow from the constitution thesis that my compositionality version ($C^{FE}$) is trivial in my sense ($\varphi$-trivial). The first two of the arguments presented below are arguments against both triviality claims but the others might be question begging if construed as arguments against the first claim.

Let's go over the semantic data first. Horwich's proposal violates synonymy intuitions as it stands. It follows from the constitution thesis that expressions that are structurally different, have meaning properties that are constituted differently, and hence have different meanings. But this contradicts the semantic data that that we are concerned with since it entails, for instance, that 'doe' does *not* mean the same as 'female deer', and that 'Lisa caressed John' does *not* mean the same as 'John was caressed by Lisa'. John Collins, who makes a similar point in response to Horwich, lists additional examples of synonymy relations holding across expressions of different structures.[40] Here are a few of these.

| | |
|---|---|
| Dative Movement: | Bob gave flowers to Ann. |
| | Bob gave Ann Flowers. |
| | |
| Nominal Extraposition: | A fly is in my soup. |
| | There is a fly in my soup. |

---

[40] See Collins (2003).

| Clausal Extraposition: | Bob is easy to please. |
|                        | It is easy to please Bob. |

| Topicalization: | I like strawberries. |
|                 | Strawberries, I like. |

| Ellipsis: | Bob went to the bank and |
|           | Mary went to the bank. |
|           | Bob went to the bank and |
|           | Mary did too. |

So by adopting the constitution thesis (as stated above), Horwich goes against many of the synonymy patterns of natural language, and it is thus inappropriate to adopt the thesis in the present context.

Horwich admits that his position, as stated, does not fit well with the kind of meanings which is invoked in judgments like the aforementioned and that his thesis gives rise to fairly fine grained meanings. But he also seems to think that this is not very compromising since

> [W]e are certainly not prevented from recognizing, in addition, a coarser grained kind of meaning, characterized in terms of some similarity relation between fine-grained meanings[...]One way of implementing this idea would be to distinguish between the syntactic and semantic structure of an expression (perhaps identifying the latter with "logical form") and to attribute the same semantic structure to certain expressions whose syntactic structure are different.
>
> <div align="right">(Horwich 1997: 512)</div>

However, since Horwich does not go into any more detail about how the proposal is supposed to be fleshed out, it is unclear exactly what it amounts to and to what extent it is compatible with all the relevant semantic data. What is clear is only that the simple original proposal has to be significantly complicated in order to get the synonymy relations right, and this seems damagingly concessive by itself, since Horwich advances simplicity as one of main argument in favour of adopting the constitution thesis.

> Assuming this constitution thesis, it is clear how, paralleling the reasoning for "dogs bark", we can explain why [a complex expression] means what it does from the fact about what its primitive constituents mean and from the fact about how it is constructed from those primitives. *The great simplicity of such an account is what justifies the constitution thesis (by "inference to the best explanation").*
>
> (Horwich 1997: 507, emphasis added)

In addition to these problems with synonymy, there seems to be even harder problems with ambiguity. Some sentences are ambiguous, even though their different meanings cannot be traced to a difference in the meaning of lexical items or a difference in syntactic structure. But if we adopt the constitution thesis, explaining ambiguities in either of these two ways seem to be the only options available. Sentences (8) and (9) can be used to illustrate the problem.

(8)   John lifted the piano.
(9)   The philosophers lifted the piano.

The problem for the constitution thesis is that there does not seem to be any account, in terms of the two sentences construction properties, for why the first sentence only has one meaning, but the second has two – the second sentence could either mean that the philosophers lifted the piano one at the time or that the philosophers lifted the piano together. The sentences seem to be normal declarative sentences built up by flanking an intransitive verb with two noun phrases and it is thus likely that they should be assigned the same syntactic schema, something like '[NP [VP NP]]'. Thus whatever meaning that is assigned to the schema there will be no difference between the construction properties with respect to the meaning of the schemas. In addition, even though there will be differences between the meanings of the words involved in the sentences, there does not seem to by any reason to think of any particular word in (9) that it is ambiguous and that the difference between (8) and (9) could thereby be explained. (9) is not like (10).

(10)   He went to the bank.

The ambiguity of (10) can be traced to the ambiguity of 'bank'. The ambiguity of (10) could thus be explained by the proponent of the constitution thesis by holding that it results from putting two different meanings (those that correspond to 'bank₁' and 'bank₂') into the meaning of the schema of (10). However, this does not seem plausible with (9): none of the relevant words seems to have a corresponding ambiguity. So on the constitution thesis there does not seem be an explanation of the difference between (8) and (9) and since ambiguity judgments belongs to the semantic data that we're concerned with in the present context, this means that we have a reason not to adopt the constitution thesis.[41]

In addition to the problems the deflationist face concerning synonymy and ambiguity, it seems that there are two other general problems with adopting an account featuring the constitution thesis. The first of these is that one often requires that meaning properties are constituted in a certain way so that certain semantic work can be performed, but if the constitution thesis is adopted not much of the relevant work can be carried out. For instance, on a truth conditional account, the semantic characterizations of sentences manifest, and allow us to explain, the entailment relations that hold between sentences. We know, for instance, that if 'John is a happy clown' is true then 'John is a clown' is true. On a simple truth theoretical account, these sentences might be characterized in the following way.

(11)   'John is a happy clown' is true iff
        John $\in$ {x : x is happy} $\cap$ {x : x is a clown}
(12)   'John is a clown' is true iff John $\in$ {x : x is a clown}

Here the entailment relation between (11) and (12) can be explained in terms of the set inclusion relation between {x : x is a clown} and {x : x is happy} $\cap$ {x : x is a clown}. But on an account featuring the constitution thesis there is no corresponding explanation, since meanings of complex expressions are not constituted in any substantial way.

Another example of semantic work is the prediction of typicality orderings for things falling under general terms. If the meaning

---

property of, say, 'sweet fruit', is constituted by a prototype instead of being constituted trivially by way of the constitution thesis then we have, for instance, not only an explanation of why fruits vary in how typical fruits they are but also an explanation of why sweet fruits vary in how typical sweet fruits they are.[42] But again, on an account featuring the constitution thesis there is no corresponding explanation. Due to the existence of semantic work of this kind, other proposals will often be preferred over adopting the constitution thesis. This seems to be a quite damaging critique since Horwich defines 'constitution' as not mere co-extension but as co-extension that would *provide explanations of facts pertaining to that which is being constituted*. Entailment relations and typicality orderings seems to be facts of exactly this kind.

The second general problem with adopting the constitution thesis rather than other more substantial constitution assumptions, is that one is committed to account for the constitution of the meaning properties of *schemas* and it is very unclear how this can be done.[43] Given the constitution thesis, one trivially discharges the constitution of meaning properties of complex expressions. But one trades this burden for the burden of discharging the constitution of the meaning of schemas. Since, schemas (or modes of composition) are not normally assigned meanings, this seems to be an important idiosyncrasy with Horwich's position. It is doubtful that the trade benefits the deflationist since there is an abundance of possibilities available as constituting the meaning properties of expressions, but none that I know of for constituting the meaning properties of schemas. Given Horwich's position on constitution it is unclear what this amounts to since whatever constitutes schemas must explain facts about these schemas, but what facts are those exactly? So until an account of this is given, Horwich's position is importantly incomplete.

To conclude, there seems to be good reasons for not assuming the constitution thesis, so $C^{FE}$ is not $\varphi$-trivial with respect to it.[44]

---

[42] See Smith and Osherson et al. (1988) for one such account.

[43] Cfr. Collins (2003).

[44] Assuming the constitution thesis is not unlike presenting a compositional theory of meaning, and holding that since this theory exists it is trivial to satisfy compositionality. True, the theory Horwich presents is compatible with the meaning properties of lexical items being constituted by anything. But this follows from even

## 2.5 Concluding Remarks

$C^{FE}$ is a minimal, precise version of compositionality. It is not trivial *tout court* and there does not seem to be any good reasons for thinking that it is $\tau$-trivial or that it is $\varphi$-trivial. Whether this leaves us in a situation where we actually can learn anything about meanings from it, or from stronger explications of compositionality, is not thereby settled, and we will return to this question in Chapter Six. But had things turned out differently, we would have had no reason to return to this question.

In resisting the position that $C^{FE}$ is $\tau$-trivial or $\varphi$-trivial, we made frequent use of the assumption of the present inquiry that it is a matter of fact whether the structure and meaning assignments of a theory are correct. But as was emphasized throughout the previous sections, what is crucial for the arguments is really just the weaker assumption that there is semantical and syntactical data, such as synonymy and ambiguity intuitions, which should be respected by adequate semantic theories.[45] And since I think it is doubtful that one is really doing natural language semantics if one's theory is not evaluated with respect to this kind of data, the conclusion that $C^{FE}$ is not $\tau$-trivial, or $\varphi$-trivial in the context of providing natural language semantics should be fairly untendentious.

We can conclude this chapter by noting that the conclusions drawn here should caution us when we approach solutions to what has been dubbed 'composition extension problems', the formal semantical problems of uncovering the conditions under which a partial semantics can be extended to total compositional semantics. Hodges, for instance, has proved that if a partial semantics is compositional,

---

simpler theories. Consider, for instance, the theory according to which the meaning property of a complex expression is constituted by being that complex expression. This theory is also trivial and is flawed in the same way as Horwich's constitution thesis; it cannot be reconciled with the various semantic data without additional complication.

[45] Daniel Cohnitz puts the matter nicely in the following way

> [T]here is a trivial sense in which every meaning function for a language can be made compositional. This is the sense in which we can hold to every theory we like come what may. However, that is not a special formal feature of compositionality, but of theories in general. We can always choose between giving up a theory, giving up the conflicting data, and even giving up logic.
>
> (Cohnitz 2005: 31)

*co-final*, i.e. such that each meaningless expression is a part of some meaningful expression, and *Husserlian*, i.e. such that the substitution of synonymous expressions within larger expressions preserve meaningfulness, then that semantics can be extended to a total compositional semantics.[46] In a similar vein, Westerståhl has proved that if a partial semantics is compositional and closed under subterms, i.e. such that any part of a meaningful expression is itself meaningful, then that semantics can be extended to a total compositional semantics.[47] What should be noted with these results is that they, even though they are not used in any way to suggest that compositionality can be trivialized, are importantly similar to the recursion theorem The common denominator is the difficulty with using the results in the enterprise of providing semantics for natural languages: what is often of interest is whether there is a compositional extension in conformity with the semantic data. And this remains an open question in spite of these proofs.

---

[46] See Hodges (2001).
[47] See Westerståhl (2004).

# 3

# The Explanatory Utility of Compositionality; Linguistic Creativity

The goal of inquiry in this essay is to ascertain the extent to which the principle of compositionality (*C*) can be justifiably imposed as a constraint on semantic theories and thereby provide information about what meanings are.

(C)    The meaning of a complex expression is determined by the meaning of its parts and its mode of composition.

It cannot be justifiably required that a semantic theory should accord with *C* if it turns out that there are no reasons for thinking that *C* is true. For if we have no reasons to think that *C* is true, we have no reasons to think that a theory not in accordance with it is incorrect. This prompts us to look closer at the reasons to think that *C* is true, and we will do so in this chapter and the next.

Whether *C* is true depends, *qua C* being a determination claim, on whether the cited determinable (that which is being determined) depends *only* on the cited determinants (the things that determine), and on it depending on *both* the cited determinants, i.e. whether, in this case, the meaning of a complex expression depends *only* on the meaning of its parts and its mode of composition, and on *both* the meanings of its parts and its mode of composition. So when we concern ourselves with settling whether there are reasons to believe that *C* is true we should thus *not* be concerned with reasons to think that *something or other* determines the meanings of complex expressions, but the reasons to think that *exactly those factors cited by C*, rather than some other constellation of factors, determine the meaning of complex expressions.

The reasons to think that *C* is true usually cited in the literature involve certain explanations that one supposes that *C*, or rather, a theory in accordance with *C*, can provide. It is often assumed that a theory in accordance with *C* can explain *Linguistic Creativity*, speakers' ability to produce and understand *novel* expressions;[1] *Productivity*, speakers' ability to understand and produce *an infinite number* of expressions;[2, 3] and *Systematicity*, speakers being such that

---

[1] I've adopted 'linguistic creativity' due to the following passage by Chomsky.

> The most striking aspect of linguistic competence is what we may call the "creativity of language", that is the speaker's ability to produce new sentences, sentences that are immediately understood by other speakers although they bear no physical resemblance to sentences which are "familiar".
>
> (Chomsky 1966: 11)

That the sentences that we immediately understand bear no physical resemblance to sentences which are familiar seems to be an exaggeration though and, as will become apparent in the next section, I have left this part out of my own understanding of 'linguistic creativity'. Evans (1981) also uses 'linguistic creativity' when he, like me, wants to distinguish between this phenomenon and (what I refer to as) 'productivity'.
[2] Donald Davidson ([1965] 2001; [1967] 2001) has famously advanced arguments in favour of compositionality in terms of 'learnability'. I will not discuss these separately from productivity since it seems to me that they pertain to the same supposed phenomenon. I hope that Davidson's concerns are addressed even though learnability will not be explicitly discussed.
[3] The ability to produce and understand an infinite number of expressions and the ability to produce and understand *novel* expressions are often conflated.

the set of expressions that they are able to understand and produce *constitute clusters of syntactically and semantically connected expressions*.[4] Since these are by far the most frequently cited reasons for believing in *C*, I will confine myself in this essay to discussing them when discussing possible reasons for thinking that *C* is true.

  If *C* is able to explain why speakers have at least one of these abilities, then given that i) we have good reasons to think that speakers actually have this ability and ii) it turns out that other, competing, determination claims are unable to explain why speakers have this ability, then we have a good reason to think that *C* is true (i.e. that the correct semantic theories accord with *C*), and thus have a good reason to impose on semantic theories the requirement that they should accord with *C*. However, if *C* is able to provide one of the relevant explanations, and there are good reasons to believe in the relevant explanandum, *but it turns out that other equally reasonable principles can also provide the explanation*, there will be no reason to

---

> Defenders of the principle of compositionality generally appeal to our ability to understand a potentially *infinite* array of *new* sentences we have never heard before.
>
> >                              (Szabó 2000: vii, emphasis added).

> One of the most remarkable (and most remarked upon) features of human language is our ability to use and understand expressions *never before uttered*. Out of a finite stock of words we are able to understand any of a potentially *infinite* list of expressions.
> >                    (Barwise and Perry 1983: 31, emphasis added)

It is understandable that the two abilities are often cited together, since if speakers are able to understand an infinite number of expressions and speakers' exposure to expressions at any given time is restricted to a finite number then they must be able to understand an infinite number of *novel* expressions. Hence, they must be able to understand novel expressions. However, it does not follow that if speakers are able to produce and understand novel expressions then they must also be able to produce and understand an infinite number of expressions. So the two abilities are distinct. And the distinction seems to be important since there are more compelling reasons to believe that speakers have the former ability than that they have the latter. Consequently, I will treat the two abilities separately and I will mark the difference terminologically. I will strictly use 'linguistically creative' to describe speakers having the first ability and 'productive' to describe speakers having the second.

[4] The set of sentences 'Mary loves Carl', 'Carl loves Mary', 'Carl loves Carl' and 'Mary loves Mary' might constitute one such cluster. A speaker who is able to produce and understand either of the first two sentences is likely to be able to produce and understand the rest of the sentences in the cluster.  Systematicity has been advanced by, for instance, Fodor (1987) and Fodor and Pylyshyn (1988).

think that *C*, specifically, is true. We will only have reason to think that *one out of a set of principles is true*. And we can then justifiably require at most that a semantic theory accords with one or the other of these principles. In addition, if *C* is able to provide the relevant explanation, and it turns out that other competing determination claims are unable to explain why speakers have this ability but that *there are no good reasons to believe in the relevant explanandum*, the explanation will give us no reason to think that *C* is true and thus *no* justification for requiring semantic theories to accord with *C*.

In light of these considerations, this chapter and the next will initially be spent discussing i) whether a certain phenomenon *is likely to be real* (as opposed to merely apparent); ii) whether semantic theories according with *C can* explain the phenomenon; and iii) whether semantic theories are *required* to accord with *C* in order to do so. Linguistic Creativity will be discussed in this chapter. Productivity and Systematicity will be discussed in the next.

The discussion of linguistic creativity will be divided into two parts. The first part (Section 3.1) will discuss speakers' ability to *understand* novel expressions, and the second part (Section 3.2) will discuss speakers' ability to *produce* novel expressions. The first part will, after an initial discussion of why it is very plausible to think that speakers understand novel expressions, depart from an example due to Richard Larson and Gabriel Segal of a by now quite generally acknowledged kind of explanation of why speakers are able to understand novel expressions (Subsection 3.1.1). The discussion will then proceed by slowly articulating what the explanatory situation is like so that the second and third questions can be answered. This will be done first by introducing a new, explanatory more able explication of compositionality – what will be referred to as '$C^{Fl}$' – (Subsection 3.1.2), proposing a particular explanatory standard (Subsection 3.1.3), clarifying what 'understanding an expression' amounts to (Subsection 3.1.4), and then finally answering the two questions we set out to answer (Subsection 3.1.5). We will then turn to the second half of linguistic creativity, speakers' ability to *produce* novel expressions, and what is required, given the previous explanatory set-up, to explain this (Section 3.2). After the discussion of the previous sections it can be concluded (in Section 3.3) that i) even though *C* (as explicated in this chapter) can be recruited to explain why speakers understand novel expressions, it is not the case that ii) *C* is required in

order to explain why speakers understand novel expressions. Hence, the fact that speakers understand and produce novel expressions, does not give us any reasons to think that *C* specifically is true, and thus no justification for requiring that a semantic theory accord with compositionality. We only obtain a reason to think that some principle out of a set of determination principles is true, and thus justification for requiring that a semantic theory must accord with one of these. It is also concluded that *C* is neither sufficient, nor necessary, in order to explain why speakers are able to produce the expressions that they are able to produce.

## 3.1 Explaining why Speakers Understand Novel Expressions

It is widely believed – and very plausible – that speakers are *linguistically creative*, i.e. that they can *understand and produce novel expressions*.[5] It also seems widely believed that this can be explained somehow in terms of the principle of compositionality. But in spite of the intuitive connection, much needs to be filled in before the situation can be seen clearly. In particular, compositionality, which is formulated in terms of expressions *having certain meanings*, must be precisely connected with expressions *being understood*, and with expressions *being produced*. Two different connections are called for and I will address them – for clarity's sake – in sequence. Understanding will be addressed in this section, and production will be addressed in Section 3.2.

What is widely believed is not that speakers understand *all* novel expressions.[6] On the contrary, it is obvious that if a speaker is

---

[5] I will use 'speaker' in what I take to be an intuitive way to refer to someone who is able to participate normally both as a speaker and a hearer in various linguistic activities. A normal person is able to do this and I do not think I need to use 'speaker' to refer to someone very remote form a normal person for the things that I say in terms of 'speaker' to turn out true. I do not think that I need to build in linguistic competency for instance, but that sufficient competence falls into place by invoking 'normal' in a statistical sense. Except for when this is obvious from context, I will not mean by 'speaker' a person who is speaking, but someone who is able to speak and understand.

[6] Expressions can be novel to some speakers but not to others. So strictly speaking all claims pertaining to an expression' being novel have to be relativized to some speaker or group of speakers. However, I will sometimes omit the relativization for the sake of succinctness of expression. Also note that strictly speaking it is *expressions* and not *utterances* that are sometimes novel and sometimes not. All *utterances* are novel to a

unfamiliar with a certain simple expression ('catamorphism' for instance) it is not generally the case that she understands it or complex expressions containing it (such as 'Catamorphisms are nice'). The subclass of the novel expressions that speakers are generally assumed to be able to understand are the complex ones *that are built up from familiar expressions in familiar ways*. 'Jim loves Jill' is such an expression relative a speaker who has never come across this *particular sequence of words* but who is familiar with the expressions 'Jim', 'loves' and 'Jill' and who is familiar with expressions consisting of transitive verbs flanked by proper nouns.

That speakers understand this kind of novel expressions is evident from the fact that they understand utterances of these expressions, and that in the absence of evidence to the contrary, the best explanation of a speaker understanding an utterance of a certain expression involves that speaker understanding the uttered expression. We only hesitate to infer from a speaker's understanding an utterance of a certain expression to the speaker understanding that expression in a situation where we have reason to doubt that the speaker understands the expression being uttered (when we know that the speaker does not speak the relevant language for instance) or when another explanation of why she understood the utterance is salient (for instance when there is an interpreter present who whispers the meaning of the utterance to the speaker, or when we know that the speaker can read minds).

That speakers understand utterances of novel expressions is in turn evident from them generally being able to predict some of the behavior of their co-speakers on the basis of hearing them utter novel expressions. I take it that it is completely obvious that speakers are able to do this. It is consequently also beyond doubt that speakers understand utterances of novel expressions, and thus beyond doubt that they understand novel expressions.

In order to make the subsequent discussion as clear as possible, I will restrict attention to what it takes to explain why a single arbitrary

---

speaker since they are tokens, i.e. non-repeatable. But an expression being novel to a speaker is still a matter of her not having encountered *utterances* of that expression. If we claim that an utterance is novel for a certain speaker, this is informative only if we understand it as the claim that *the expression being uttered* is novel to the speaker.

English speaker $S$ understands novel expressions. The explanandum that will concern us in this section is thus the following.[7]

($LC^U$) $S$ understands novel expressions.[8]

Since we are concerned with accounts of why speakers have certain abilities, most claims about meaning – which strictly speaking have to be relatized to speakers or languages to makes sense – will be relativized to speakers. I will make use of the phrase 'a speaker $S$ assigns meaning $m$ to expression $e$'. By this I just mean that $e$ means $m$ to $S$. I will discuss more precisely what I take it that an expression meaning something to a speaker amounts to in Subsection 3.1.2.

  Since it is obvious that $LC^U$ is true, the two questions that remain to be answered in this section is whether compositional semantic theories can explain $LC^U$, and whether a semantic theory is *required* to be compositional in order to explain $LC^U$. I will go about answering these questions by first presenting a particular example of an account of $LC^U$ in terms of a compositional theory which at first glance seems explanatory. I will then, in the following sections, gradually work my way away from this particular example to a better general understanding of the explanatory situation and its various components in order to see more clearly what exactly is required in order to explain $LC^U$.

  The construal of the explanatory situation that I will ultimately propose is the following; $S$'s *understanding* an expression $e$ is a matter of $S$'s assigning a meaning to $e$ which is similar to those meanings assigned to $e$ by the members of some relevant speech community (a set of speakers); *explaining* $LC^U$ is a matter of giving a causal explanation – I here follow Gareth Evans – i.e. providing details about causal histories. It follows from these assumptions that explaining why $S$ understands novel expressions is really a matter of providing details that are common to the causal histories leading up to the converging meaning assignments of the speakers of $S$'s community. It will be argued that the particular challenge facing a

---

[7] To simplify the exposition, I here and elsewhere make claims of the form 'S is such and such'. If other readings are not clear from context these claims just mean that speakers in general are such and such.

[8] '$LC^U$' is short for 'Linguistic Creativity – Understanding'.

semantic theory that wants to explain the understanding of the relevant *novel* expressions is a matter of being *balanced*, i.e. the theory must be such that the causal histories it describes involves a dependence on enough experience with past linguistic transactions so that it does not entail that too many speakers are predicted to converge on the same meaning assignments, but not so much that it presupposes that utterances of all the expressions that speakers understand are part of their causal histories.

### 3.1.1 A Traditional Semantic Explanation

Richard Larson and Gabriel Segal's comprehensive treatise on Davidsonian truth-conditional semantics 'Knowledge of Meaning' contains a clear description of a by now familiar construal of how compositionality can enter into an explanation of $LC^{U}$. They describe their position in the following way.

> If knowledge of meaning is knowledge of a body of rules and principles, what sorts of rule and principles would these be? A very natural idea is that some of them would tell you about the meanings of individual words and morphemes and others would tell you how these meanings interact when their corresponding expressions are put together in a sentence [...] Semantic rules of this kind are said to be **compositional**. They give the semantic content of a sentence by specifying the semantic contributions of its parts and the semantic significance of putting those parts together according to a definite mode of syntactic combination.
>    The hypothesis that we know a set of compositional semantic rules and principles is a highly attractive one having a great deal of explanatory power[...]the hypothesis accounts [for instance] for the obvious but important fact that *we can understand new sentences*, sentences that we have never come across before. This[...]is easily explained if we have a body of rules that allow us to infer the meanings of new sentences from prior knowledge of the meanings of their parts and from knowledge of the semantic significance of their combination.
>                     (Larson and Segal 1995: 12, emphasis in original)

The following example theory can be used to illustrate how Larson and Segal think that a semantic theory can be used to explain $LC^{U}$.

$t_1$     *Syntax:*                         *Production rules:*

   S → 'Phil ponders',                *Substitution of Equivalents(SE):*
        'Chris agrees',                $F(\alpha)$
        'Jill knows Kate'            $\alpha$ iff $\beta$
   S → S·'and'·S                    --------
   S → S·'or'·S                     $F(\beta)$
   S → 'It is not the case that'·S

                                    *Universal Instantiation(UI):*
                                    For any S, $F(S)$
                                    ------------------
                                    $F(\alpha)$

*Semantics:*
1a.   'Phil ponders' is true iff Phil ponders.
1b.   'Chris agrees' is true iff Chris agrees
1c.   'Jill knows Kate' is true iff Jill knows Kate

For any sentences S, $S_1$ and $S_2$…
2a.   [$_S$ $S_1$ 'and' $S_2$] is true iff both $S_1$ is true and $S_2$ is true
2b.   [$_S$ $S_1$ 'or' $S_2$] is true iff either $S_1$ is true or $S_2$ is true
2c.   [$_S$ 'It is not the case that' S] is true iff it is not the case that
       S is true
2d.   [$_S$ $\alpha$] is true iff $\alpha$ is true (for any elementary sentence $\alpha$)

Let's assume that *S*, the speaker featured in $LC^U$, knows this theory
and that the novel expressions that *S* understands are among the
complex expressions covered by the theory. Given that *S* knows $t_1$ she
is in a position to infer the truth conditions not only of the
expressions that are familiar to her (say, the sentences corresponding
to 1a-1c and some sentences containing 'and', 'or' and 'it is not the
case that') but also for the expressions that are novel to her. Assume
that the sentence 'Phil ponders or Chris agrees' is one of these. Given
that *S* can assign an appropriate syntactic structure to the sentences,
i.e. given that she can assign the structure [$_S$ [$_S$ Phil ponders] or [$_S$
Chris agrees]]) to the sentence, she can infer its truth conditions in
the following way.

1.  [$_S$ [$_S$ Phil ponders] or [$_S$ Chris agrees]]
2.  [$_S$ [$_S$ Phil ponders] or [$_S$ Chris agrees]] is true iff either
    [$_S$ Phil ponders] is true or [$_S$ Chris agrees] is true.       2b, (UI)
3.  [$_S$ Phil ponders] is true iff 'Phil ponders' is true.       2d, (UI)
4.  [$_S$ Chris agrees] is true iff 'Chris agrees' is true.       2d, (UI)
5.  'Phil ponders' is true iff Phil ponders.       1a
6.  'Chris agrees' is true iff Chris agrees.       1b
7.  [$_S$ Phil ponders] is true iff Phil ponders.       3, 5, (SE)
8.  [$_S$ Chris agrees] is true iff Chris agrees.       4, 6, (SE)
9.  [$_S$ [$_S$ Phil ponders] or [$_S$ Chris agrees]] is true iff
    either Phil ponders or [$_S$ Chris agrees] is true.       2, 7, (SE)
10. [$_S$ [$_S$ Phil ponders] or [$_S$ Chris agrees]] is true iff
    either Phil ponders or Chris agrees.       8, 9, (SE)

If assigning truth conditions to a sentence is sufficient for understanding it, then given the previous assumptions (that $S$ knows $t_1$, and that the novel expressions $S$ understands are among the complex expressions of $t_1$) it seems that $t_1$ is able to explain $LC^U$. A crucial part of this explanation is that it describes a certain kind of mechanism whereby given only familiarity with some expressions the speakers competence is extended to cover novel expressions as well. $t_1$ is able to do this by containing, what Larson and Segal refers to as 'compositional rules'(2a-2d).

   If we had only been interested in providing a particular explanation of $LC^U$ we might have been satisfied with (an expanded version of) this theory. But since we are concerned with the general question of what is required to explain $LC^U$ we need to take a step back from $t_1$, and try to get a better grip of what the explanatory situation is like, in particular i) what kind explanatory standard that is at work, but also ii) what a reasonable explication of 'understanding' should amount to, and iii) how exactly compositionality must be construed in order for theories in accordance with it be able to provide explanations like the previous one. Let's start with the explication of compositionality.

### 3.1.2 A Stronger Compositionality Explication

It should be obvious from considering the previous example, that semantic theories in accordance with $C^{FE}$ – the explication of $C$ introduced in chapter two – will not in general be able to explain $LC^U$. Although $C^{FE}$ was suitable for the purposes of the previous chapter, it is, from the present explanatory perspective, objectionably weak; it ensures only that meanings *distribute over expressions* in a certain way. But in the context of providing explanations for $LC^U$ we seem to need something stronger; a requirement that there is some sort of *connection* between the meaning of a complex expression and the meaning of its parts and its mode of composition. That $C^{FE}$ guarantees no such connection makes it an unsuitable explication of $C$ in the present context.[9]

---

[9] That versions of compositionality explicated in terms of the existence of a certain function-in-extension are very weak and might not be adequate for some purposes is generally appreciated in the literature.

> Compositionality, as typically understood by semanticists, is the claim that the meaning of a complex expression is a *function* of the meanings of its constituents and the syntactic way these constituents are combined [...] Functions are cheap. Suppose we systematically assign natural numbers to all (disambiguated) English sentences, different numbers to different sentences. Then the meaning of any English sentence is a function of its associated number [...] Fodor and Lepore (quite sensibly) take compositionality to be a stronger claim than many formal semanticists do [...]
>
> (Szabó 2004: 341)

That one might have something stronger in mind with 'compositionality' than one gets from $C^{FE}$ is also illustrated by the following passage from Pagin (2004) who uses a version of compositionality similar to $C^{FE}$.

> It is not a violation of compositionality, in this standard sense, that a complex expression has a meaning that is much richer than than what you intuitively get out of the meanings of the parts. For instance, suppose we have expressions $e_1$ and $e_2$, a syntactic operation $\sigma$ and a meaning function m such that $\mu(e_1)$=George W Bush, $\mu(e_2)$=Silvio Berlusconi and $\mu(\sigma\ (e_1,e_2))$=*the proposition that George W Bush and Silvio Berlusconi will never have visited Bhutan together*. However odd, this wouldn't be a violation of [the principle of compositionality].
>
> (Pagin 2004, italics in original, sic.)

Before devising a new explication we can consider two other problems with explicating $C$ in terms of $C^{FE}$. These provide additional symptoms that will help suggest an appropriate treatment.

Consider first that it is sometimes held that the principle of compositionality is a congenial principle since it facilitates *lexical economy*.

> An approach which directs us to choose as lexical units the most compositionally basic units of the language has [...] simplicity in its favour. Wherever we can use composition, dictionary entries are avoided.
>
> (Katz and Fodor 1963: 192)

However, there is nothing about a theory being in accordance with $C^{FE}$ that guarantees lexical economy, in the sense of making use of fewer explicit semantic rules. $t_2$, for instance, accords with $C^{FE}$ even though it could not be more lexically prolific.

| $t_2$ | Syntax: | | Semantics: | |
|---|---|---|---|---|
| | N | → 'dog', | $\mu$('dog') | = {x : x is a dog} |
| | | 'cat', | $\mu$('cat') | = {x : x is a cat} |
| | | 'bird' | $\mu$('bird') | = {x : x is a bird} |
| | A | → 'brown', | $\mu$('brown') | = {x : x is brown} |
| | | 'angry', | $\mu$('angry') | = {x : x is angry} |
| | | 'loud' | $\mu$('loud') | = {x : x is loud} |
| | NP | → A·N | $\mu$('brown dog') | = {x : x is a brown dog} |
| | NP | → A·NP | $\mu$('angry dog') | = {x : x is an angry dog} |
| | | | $\mu$('loud dog') | = {x : x is a loud dog} |
| | | | $\mu$('brown cat') | = {x : x is a brown cat} |
| | | | $\mu$('angry cat') | = {x : x is an angry cat} |
| | | | $\mu$('loud cat') | = {x : x is a loud cat} |
| | | | $\mu$('brown bird') | = {x : x is a brown bird} |
| | | | $\mu$('angry bird') | = {x : x is an angry bird} |
| | | | $\mu$('loud bird') | = {x : x is a loud bird} |
| | | | … | |

Consider second that some functions-in-intension that clearly do not belong in a compositional theory can be a part of theories in accord with $C^{FE}$. Imagine for instance, a trivial semantic theory where

the only semantic rule is the identity function on expressions, '$\mu(E) = E$'. A semantic theory making use of this semantic rule should clearly not be counted as being compositional since it is, intuitively, the case that the meaning of a complex expression is not in this case determined by the meanings of its parts and its mode of composition. Instead, it is determined entirely by the exact *form* of the expression; if and only if two expressions have different forms will they be assigned different meanings. Despite this it is the case that a semantic theory that only makes use of this semantic rule will be in accordance with $C^{FE}$ since all expressions will have distinct meanings and the synonyms that are needed to set up a counterexample to $C^{FE}$ are not available.

The three problems considered so far have a common solution. What is needed is an explication of $C$ in terms of a restriction on the semantic rules which *describe* the meaning assignments of a semantic theory, not one in terms of a restriction directly on the meanings being assigned. Given that a compositional theory is required to *specify* the connection between the meaning of a complex expression and the meanings of its parts and its mode of composition, the theory will contain something which could be cited in an explanation of, for instance, why a speaker understands novel expressions. Additionally, if a theory is prohibited from assigning meanings to complex expressions one by one by way of *explicit* semantic rules and are forced to assign meanings to complex expressions collectively by way of general *implicit* rules, lexically prolific theories like $t_2$ would thereby be ruled out. Finally, by requiring that the implicit semantic rules of a theory are in a certain way, functions like the identity function which *makes reference* to the form of a complex expression and not only the meaning of the parts of a complex expression and its mode of composition will be excluded from being part of the semantic theory.

I thus propose that $C$ should be explicated in terms of $C^{FI}$.[10]

---

[10] We remember from chapter one that a semantic rule is an expression of the form '$\mu(x)=y$' which reads 'the meaning of $x$ is $y$' and that the rule is explicit if '$x$' is an expression, and implicit if '$x$' is a variable over expressions.

A semantic theory *t* accords with $C^{FI}$ iff

i) There are no explicit semantic rules for complex expressions in *t*.[11]

ii) Each implicit semantic rule in *t* is of the form '$\mu(x)=y$' where

    1) '*x*' is a mode of composition variable, and

    2) '*y*' is a *c-term*, where a *c-term* is defined in the following way:

        a) '$\mu(x_1)$' is a *c*-term if '$x_1$' is one of the variables from which '*x*' is built up.[12]

        b) '$g(x_1,\ldots, x_n)$' is a *c*-term if '$x_1$',…,'$x_n$' are *c*-terms and '*g*' is a functor corresponding to a humanly computable function defined on the set of meanings.[13, 14]

By requiring that a semantic theory accords with $C^{FI}$, we thereby require that the only way that meanings can be assigned to complex

---

[11] Without this requirement, $C^{FI}$ would not entail $C^{FE}$ and thus not be a proper compositionality explication. For without the requirement one complex expression could have been explicitly assigned a certain meaning even though another complex expression which has parts with the same meanings and has the same mode of composition as the first expression was not. $C^{FE}$ would then be false even though $C^{FI}$ could still be true. This requirement also solves the second problem discussed above.

[12] By 'mode of composition variable' is simply mean the kinds of expressions that have been enclosed by parenthesis on the left hand side of implicit semantic rules, i.e. a variable ranging over all the expression having been combined in a certain way. Thus, '$_{NP}$N P' is a mode of composition variable. We say that this variable is built up by the variables 'N' and 'P' but not from 'NP' (the variable which is to the far left of the expression) which is the name of the relevant mode of composition. The name of a mode of composition is included in the implicit semantic rules to keep track of which implicit syntactic rule it correspond to.

[13] Sometimes it is suggested that meanings should only be permitted to combine in more specific ways. For instance, Heim and Kratzer (1998) suggest the requirement that composition should always be a matter of applying one meaning to some other functional meaning. But I know of no reason to adopt this restriction and it would be question begging in the present context to require that some meanings must be functions.

By 'humanly computable' I simply mean computable with the resources available to humans.

[14] 'FI' in '$C^{FI}$' is short for 'Function-in-Intension'.

expressions is through mode of composition specific semantic rules at most sensitive to the meanings of the parts of expressions. We will refer to these implicit rules as *composition rules*. Given the kind of rules presupposed by $C^{FI}$ we can devise licensed substitutions similar to the production rules employed by Larson and Segal, by which we can obtain the meanings of complex expressions from the rules of the theory.

The new explication of $C$ can be illustrated with the following theory which is in accordance with $C^{FI}$.

$t_3$    Syntax:                          Semantics:

| | | | | |
|---|---|---|---|---|
| N | → | 'dog', | $\mu$('dog') | $= \{x : x \text{ is a dog}\}$ |
| | | 'cat', | $\mu$('cat') | $= \{x : x \text{ is a cat}\}$ |
| | | 'bird' | $\mu$('bird') | $= \{x : x \text{ is a bird}\}$ |
| A | → | 'brown', | $\mu$('brown') | $= \{x : x \text{ is brown}\}$ |
| | | 'angry', | $\mu$('angry') | $= \{x : x \text{ is angry}\}$ |
| | | 'loud' | $\mu$('loud') | $= \{x : x \text{ is loud}\}$ |
| NP | → | A·N | $\mu(_{NP} \text{ A N})$ | $= \mu(\text{A}) \cap \mu(\text{N})$ |
| NP | → | A·NP | $\mu(_{NP} \text{ A NP})$ | $= \mu(\text{A}) \cap \mu(\text{NP})$ |

Note first that $t_3$ is in accordance with $C^{FI}$ since i) no complex expressions are assigned meanings explicitly (i.e. directly in terms of those expressions), ii) the implicit semantic rules are of the form '$\mu(x)=y$' where '$x$' ('$_{NP}$ A N' respectively '$_{NP}$ A NP') is a mode of composition variable, and   '$y$' is a *c*-term ('$\cap$' is a functor corresponding to a humanly computable function, and the expressions '$\mu$(A)' and '$\mu$(N)', and '$\mu$(A)' and '$\mu$(NP)' respectively only make use of variables that that are combined into the relevant mode of composition variable). We can also note that $t_2$ is not in accordance with $C^{FI}$ since it contains explicit meaning assignments to complex expressions.

It is important to note that the previous formulation of $C^{FI}$ is just one example of how the general idea that compositionality could be understood in terms of a restriction on the semantic rules of a theory can be spelled out. The formulation contains numerous simplifying assumptions that one might not want to retain and that

are not necessary parts of the general idea. Hence, the inadequacies of this particular format need not be inadequacies of the general idea.[15]

One of the benefits with an explication of $C$ in terms of $C^{FI}$ is that it, like $C^{FE}$ makes no assumptions about the number of meanings an expression must have. A theory can be in accordance with $C^{FI}$ and still contain complex terms that have multiple meanings due to multiple implicit semantic rules corresponding to each mode of composition. So $C^{FI}$ by itself is a thin principle. But in contrast to the definition of $C^{FE}$, the definition of $C^{FI}$ does not need to contain explicit reference to semantic ranges. Instead, the semantic ranges are induced by the totality of the semantic rules (even though none of them by themselves makes reference to semantic ranges).

Another benefit of $C^{FI}$ is that it does not presuppose that meanings are understood in terms of truth-conditions, but is general enough to encompass meanings being any number of things.[16]

Let's now return to the problems that motivated the search for a new explication of the compositionality principle. If we explicate $C$ in terms of the aforementioned definition, a semantic theory utilizing the identity function on expressions (that is, '$\mu(E) = E$') as a semantic rule, will not be compositional. The identity function is not an admissible composition rule; its left hand side includes a general variable over expressions instead of a mode of combination variable, and, more importantly, its right hand side contains an expression which is not a c-term since it is neither an expression of the form '$\mu(E)$', nor something that can be built up by appropriate functors and such expressions.

In addition, a semantic theory in accord with $C^{FI}$ exhibits a certain lexical economy. A $C^{FI}$ theory encompassing meaningful complex expressions will always be more lexically parsimonious than another theory encompassing the same expressions but that contains only explicit semantic rules.

---

[15] Sometimes compositionality is spelled out in terms of something like $C^{FE}$ with an added computability requirement. Once we see the explanatory situation more clearly we can see why this does not provide a satisfactory explanation of $LC^U$. See Section 3.1.5 for a discussion.

[16] Since $C^{FI}$ restricts the form of the semantic rules of a theory in a certain way it turns out that Larson and Segal's particular theory turns out to be non-compositional. (It even turns out that their theory is not a semantic theory.) This is not very interesting though since very similar theories can be devised that are in accordance with $C^{FI}$.

So it seems that by adopting $C^{FI}$ at least two of the problems that beset $C^{FE}$ can be avoided. We will see in the following sections that it avoids the third as well. $C^{FI}$ thus seems to be a quite reasonable explication of compositionality. This will be substantiated in the rest of this chapter and it will be clear that a theory in accordance with $C^{FI}$ is able to perform the explanatory work that compositionality is expected to perform, and from this and the following two chapters, it will be clear that it avoids the problems that beset other conceivable explications.

Before moving on to consider $C^{FI}$'s explanatory utility, something about the way $C^{FI}$ induces lexical economy bears emphasis; even though $C^{FI}$ gives rise to the desired property, so does many other principles. Consider, for instance, the principle $F^{FI}$, which is just like $C^{FI}$ except that it allows reference, both on the left hand side and of the right hand side of an implicit semantical rule, to syntactic variables in. Hence, according to $F^{FI}$, '$\mu(e) = e$', is a permissible implicit semantic rule. Now this principle is in the same place as $C^{FI}$ with respect to lexical economy; a $F^{FI}$ theory encompassing meaningful complex expressions will always be more lexically parsimonious than another theory encompassing the same expressions but which contains only explicit semantic rules. It is not, strictly speaking, in virtue of $C^{FI}$ that being in accordance with $C^{FI}$ gives rise to lexical economy. The theories that conform to $C^{FI}$ also conform to a more general property, and it is *it*, rather than the specifics of $C^{FI}$ that is responsible for inducing lexical economy. What is essential for lexical economy is that meanings are assigned to complex expressions *qua* members of some category (by means of variables for instance). This is done in $F^{FI}$ theories, and a host of other theories. So even though theories in accordance with $C^{FI}$ exhibit lexical economy, this gives us no reason in general to prefer the compositional theories over non-compositional alternatives, since the alternatives can also instantiate the pattern that is really responsible for inducing the lexical economy of the $C^{FI}$ theories.

This last paragraph is an illustration of the disheartening moral for the compositionalist that will be the main conclusion of this chapter; it is clear that compositional theories have beneficial properties, but it is often not the case that they have these, strictly speaking, in virtue of being compositional.

### 3.1.3 Causal Explanations

The previous section explicates the principle of compositionality in such a way that theories in accordance with it should be able to provide explanations of $LC^U$. But in order to answer the question of what is required in order to explain $LC^U$ we must consider additional details about the explanatory situation, in particular: what general explanatory standard are theories supposed to meet in order to explain $LC^U$? This question will be answered in this subsection. The answer will be used so that a particular proposal concerning linguistic creativity (and the other phenomena that will be discussed in the next chapter) can be judged as explanatory, not by relying on more or less vague intuitions about whether particular accounts are proper explanations, but by seeing whether these accounts fulfill conditions of explanatory adequacy derived from a *general* account of explanation.

It is tempting to subsume explanations of $LC^U$ under the constraints of *causal explanations*. This is a well understood, and well explored, form of explanation that seems to fit the present context. By adopting considerations from this form of explanation, we will get some grip on how to think about and evaluate candidate explanations of $LC^U$. Lewis's account of causal explanation seems particularly suitable to this end due to its clarity and simplicity.

According to Lewis,

> …to explain an event is to provide some information about its causal history
>
> (Lewis 1986: 185)

where a *causal history* is a

> …relational structure. Its *relata* are *events*: local matters of particular fact, of the sort that may cause or be caused. I have in mind events in the most ordinary sense of the word: flashes, battles, conversations, impacts, strolls, deaths, touchdowns, falls, kisses[...] But also I mean to include events in a broader sense: a moving object's continuing to move, the retention of a trace, the presence of copper in a sample [...] The causal history of a particular event includes that event itself, and all events which are parts of it.
>
> (Ibid: 184-185)

On Lewis account there are several ways in which explanatory acts –
such as proposing a particular theory for instance – can be less than
satisfactory. Two of these, even though fairly obvious, should be
mentioned.[17]

> An act of explaining may be unsatisfactory because the
> explanatory information provided is unsatisfactory. In
> particular it might be misinformation: it might be a false
> proposition about the causal history of the explanandum.
>
> (Ibid: 193)

> The explanatory information provided may be correct, but
> there may not be very much of it. It might be a true but weak
> proposition; one that excludes few [...] of the alternative
> possible ways the causal history might be
>
> (Ibid: 193-194)

Even though Lewis's account is primarily an account of explanations
of particular events, his extension of the account to cover the
explanations of *kinds of events* will also be useful.

> All the events of a given kind have their causal histories, and
> these histories may to some extent be alike. Especially, the
> final parts of the histories may be much the same from one
> case to the next, however much the earlier parts may differ.
> Then information may be provided about what is common to
> all the parallel causal histories–call it *general explanatory
> information* about events of the given kind. To explain a kind

---

[17] Lewis also lists other more pragmatic ways in which an explanatory act can be
unsatisfactory. It might, for instance, provide the recipient with stale news, or it
might be packaged in such a way that the recipient cannot digest it. These pragmatic
aspects of explanatory acts will be less interesting in what follows. Although Lewis
suggests that explanatory acts might be unsatisfactory under the aforementioned and
other circumstances, he does not want to 'fuss' about how to draw the distinction
between something being a bad explanation and it not being an explanation at all.

> [W]e needn't make it clear [...] what to say about an unsatisfactory
> chunk of explanatory information, say one that is incorrect or one
> that is too small to suit us. We may call it a bad explanation, or no
> explanation at all.
>
> (Ibid: 186)

I will follow Lewis in not fussing about this.

of event is *to provide some general explanatory information about events of that kind*.

(Ibid: 192, emphasis added)

So to sum up; on the view espoused by Lewis, explaining *a particular event* is a matter of providing details about its causal history. Explaining *a kind of event* is to provide details that are common to the causal histories of these events. Providing explanatory information can be faulted for being incorrect or for not being ample enough (for some purpose). This gives us some grip on what to demand from an explanation of $LC^U$.

Consider now how well Larson and Segal's explanation of $LC^U$ fits this pattern. If we assume that a speaker understanding novel expressions, is just a matter of *a speaker assigning certain truth-conditions* to novel expressions, then it seems that $LC^U$ is a particular event in the same broad sense of 'event' according to which 'a moving object's continuing to move, the retention of a trace, the presence of copper in a sample' are particular events.[18] And even if Larson and Segal's explanation is not in terms of the *causal histories* of the relevant truth condition assignments, it is in terms of the *derivational histories* of these assignments. One might insist on this distinction – since inferences are usually held to be atemporal processes, while causal processes are not – but nothing from the perspective of the constraints suggested by Lewis on satisfactory acts of explanations seems to depend on it. An explanation along the lines of Larson and Segal's, will clearly be less than satisfactory if it provides *misinformation* about the derivational history, or if it provides *very little* information about the derivational history.

Moved by the aforementioned similarities, I will henceforth consider the explanations that compositionality has been assumed to be able to provide as causal explanations. If one feels compelled to resist this assumption, I do not think that any of my conclusions depend on anything other than it being appropriate to require that the constraints on explanations suggested by Lewis are met. And these constraints seem appropriate to apply to explanations such as that given by Larson and Segal as well. In addition, I do not think that this construal of the explanatory situation is objectionable from the

---

[18] I will henceforth assume that something being in a certain state is an event.

perspective of particular semantic theories. Several examples will be given throughout the thesis of the wide variety of semantic theories that can, at least, prima facie, provide causal explanations.

Since we are concerned with causal information, it should be spelled out in more detail what the pertinent states that are causally related are, i.e. what it is that semantic theories are expected to provide information about. I will assume that the relevant states are the *meaning assignments* of the speakers. These come in two varieties: meaning assignments to utterances and meaning assignments to expressions. Although there is surface similarity between the meaning assignments involving expressions and those involving utterances, the two kinds of states are importantly different.

The assignment by a speaker of a meaning to an *utterance* is a particular mental state at the end of a certain causal history. The state involves associating one part which is suitably related to an utterance – we might think of this part as a mental singular term referring to the utterance, or we might think of it as one of the grammatical terms corresponding to that utterance – and one part suitably related to a meaning – we might think of this part as a mental singular term referring to a meaning, but it might also be a mental particular instantiating the meaning (if meanings are mental types). In the normal case, a state of this kind has a causal history in which it is immediately preceded by an utterance of an expression, and the application of a series of dispositions of the speaker to get from the perception of the utterance to the assignment of a meaning to it.

The assignment by a speaker of a meaning to an *expression* is also a particular mental state at the end of a certain causal history. But this state is better thought of as a *dispositional* state consisting in the disposition to assign certain meanings to utterances of the relevant expression. In many cases the causal history of the assignment of a meaning to an utterance will feature the application of the disposition corresponding to the meaning assignment of a meaning to some appropriate expression. For instance, assume that a speaker hears an utterance of 'Wolf!' and goes through a series of perceptual and computational states that eventuates in her assigning a meaning to this utterance. Then in many cases one of the things that was part of this process was her having assigned a certain meaning to 'Wolf' (the expression) and that this part of the process corresponds to the exercise of a certain disposition. The causal histories of these

dispositions are different than those leading up to the assignments of meaning to utterances. The causal histories of the dispositions are not, for instance, normally such that they are immediately preceded by an utterance of the relevant expression. The processes are better characterized as *learning* processes, which feature, among other things, a speaker's ability to detect patterns over the linguistic transactions that she is involved in.

I was careful to hold that for 'many' utterances it was the case that their causal histories featured the application of a disposition corresponding to a meaning assignment involving the relevant expressions. This is very likely not the case for the causal histories of utterances of all expressions. For one of the points with compositional theories (and, as Section 3.1.5 will reveal, other theories besides) is that the meaning assignments involving complex expressions is determined by other factors about the speaker such as her assigning certain meanings to parts of the complex expressions for instance. Hence, it is natural to interpret the implicit semantic rules of (for instance) compositional semantic theories as suggesting that the speaker has a disposition to assign meanings to complex utterances based on, for instance, the meanings that are assigned to the parts of these utterances. But if the causal history leading up to the meaning assignment of a certain complex expression involves this kind of disposition – call it a 'construction-disposition' – then a separate disposition to assign meanings to utterance of the relevant expression would be redundant. Instead, having a disposition to assign a certain meaning to a complex expression seems to be a *mere* disposition, something that supervenes on the construction dispositions and the dispositions that it involves (such as those corresponding to the meaning assignments of simple expressions). This does not mean that meaning assignments corresponding to complex expressions are not dispositions, just that they are not full-blooded causally efficacious dispositions in the same sense that the dispositions corresponding to simple expressions are.

I will assume that semantic theories can provide causal information by providing, through the derivations the theory gives rise to, information of what states that are involved in assigning meanings to both utterances and expressions. In particular, the semantic theories

provide information about what states causal histories converge on.[19] Given a certain theory and the present interpretation of that theory, if the derivational histories corresponding to the meanings of, say, 'brown cow' and 'brown bird' converge, this amounts to that theory attributing a common cause to the involved events (assigning meanings to particular utterances, but also being disposed to assign certain meanings to these utterances, i.e. assigning certain meanings to the relevant expressions).

Before concluding this account of the kind of causal information that will be relevant in an explanation of $LC^U$ I want to discuss briefly the related problem of how we can say that one semantic theory correctly characterizes a speaker, while another theory that is extensionally equivalent to the first theory does not.[20] The solution to the problem is simply to maintain that the semantic rules of a semantic theory correspond to the aforementioned kind of causally efficacious dispositions and that manifestations of these dispositions can be used to evaluate the claim that one theory is correct while another extensionally equivalent theory is not.[21] Consider for instance the claim that $t_3$ correctly characterizes the semantic competence of a speaker but that another theory such as $t_2$ which is extensionally equivalent with $t_3$ does not.[22]

---

[19] This view and the view that semantic assignments to expressions consist in a speaker being disposed in a certain way is due to Evans ([1981] 1985). I depart from Evans in at least three ways; i) He is concerned with assigning truth conditions, and not meanings to expressions, ii) he only thinks that the explicit semantic rules correspond to dispositions, not the implicit ones, and iii) he felt compelled to define exactly what the dispositions amount to. I follow Wright (1986) and Davies (1987) in assuming that iii) is not necessary.

[20] Just saying that the speaker *knows* one theory but not the other does not seem to amount to very much since two of the characteristic features of knowledge, i) that a speaker who knows that $p$ assent to $p$, and ii) that things a speaker knows can be put to use in a wide range of situations, does not apply to the tacit knowledge that speakers are supposed to have of the axioms of semantic theories. The first difficulty has been raised by, for instance, Quine (1972) and Stich (1971), and the second by Evans ([1981] 1985).

[21] This proposal is also due to Evans ([1981] 1985).

[22] This theory also illustrates the point that not only *internal* semantic theories according to which meanings are something internal to the speaker such as mental states, but also *external* semantic theories according to which meanings are something external to the speaker such as references, can provide information about causal details. Although, for instance, '∩' does not denote a mental event, or even an event at all, a theory featuring such functors can provide details about causal histories nonetheless. The proponent of external semantic theories such as $t_3$ might for

| $t_3$ | Syntax: | | Semantics: | |
|---|---|---|---|---|
| | N | → 'dog', | $\mu$('dog') | = {x : x is a dog} |
| | | 'cat', | $\mu$('cat') | = {x : x is a cat} |
| | | 'bird' | $\mu$('bird') | = {x : x is a bird} |
| | A | → 'brown', | $\mu$('brown') | = {x : x is brown} |
| | | 'angry', | $\mu$('angry') | = {x : x is angry} |
| | | 'loud' | $\mu$('loud') | = {x : x is loud} |
| | NP | → A·N | $\mu(_{NP} A N)$ | = $\mu(A) \cap \mu(N)$ |
| | NP | → A·NP | $\mu(_{NP} A NP)$ | = $\mu(A) \cap \mu(NP)$ |

Given the present interpretation of a semantic theory, the aforementioned claim should entail the claim that certain things are true about the acquisition and decay of that speakers semantic competence.[23] For instance, we can understand how the speaker could have acquired $t_3$ without exposure to all the expressions it covers, but we cannot understand how the speaker could have acquired the alternative theory without such exposure. And if the speaker learns the meaning of a new adjective or noun, $t_3$ would predict that the speaker thereby learns the meanings of several other expressions as well. Similarly, if a speaker changes his view concerning what a particular word means, $t_3$ would thereby predict that he would change his view concerning several other expressions as well. Conversely, if a speaker's competence would decline, then $t_3$ would predict that we lose competence in larger chunks rather than incrementally expression by expression.

---

instance, argue that even though an application of the relevant semantic rule is not part of the etiology of the events that we want to explain, *a mental representation of it, operating on mental representations of its arguments* is part of this process. So, the theory provides relevant information about the causal history by characterizing the representations that are part of this history *in terms of what they are representations of*. On this interpretation a theory such as $t_3$ is able to provide details about the causal history of the meaning assignment of an expression. So it is, importantly, not part of the present construal of the explanatory situation that meanings have to be something mental in order for semantic theories to be able to explain phenomena such as $LC^5$. The distinction between internal and external semantic theories I make use of is roughly that of Davis and Gillon (2004).

[23] This proposal is also due to Evans. Larson and Segal (1995: 62) adopt a similar line.

### 3.1.4 The Understanding of Expressions

Now that we have a clearer view of the pertinent kind of explanation we can turn to the elucidation of $LC^U$. In particular we need to get clearer on what speakers' *understanding* expressions amount to.

On Larson and Segal's explanation of $LC^U$, understanding a sentence is construed as *assigning truth conditions* to that sentence. This is inappropriate in the context of the present inquiry. Not all semantic theories invoke truth conditions, so if assigning truth conditions is assumed to be a necessary part of understanding, and thus a necessary part of an explanation of $LC^U$, many candidate meanings would thereby be removed from consideration. So in order not to bias the present inquiry, we need an alternative explication of 'understanding'.

But understanding 'understanding' in terms of the more neutral *assigning a meaning* to an expression is not appropriate either. For even though assigning a meaning to an expression is *required* for understanding that expression, it is not sufficient. 'Understand' is an *achievement verb* like 'spell', 'see' and 'know' and as such it requires that some criterion of success is met in order for it to apply.[24] In the case of understanding expressions and utterances, it seems plausible that this criterion amounts to a restriction on permissible meaning assignments. Not any assignment of a meaning to an utterance or an expression counts as understanding.

For a speaker to understand *an utterance*, i.e. an expression token, it might seem plausible to require that the meaning she assigns to the utterance must match the one the producer of the utterance intended

---

[24] Cfr. Pagin (2003b). The expression 'achievement verb' is from Ryle (1949).

> We use the verb 'to observe' in two ways. In one use, to say that someone is observing something is to say that he is trying, with or without success, to find out something about it by doing at least some looking, listening, savoring, smelling or feeling. In another use, a person is said to have observed something when his exploration have been successful, i.e. that he has found something out by some such methods. Verbs of perception such as 'see', 'hear', 'detect', 'discriminate' and many others are generally used to record observational success, while verbs like 'watch' 'listen', 'probe', 'scan' and 'savour' are used to record observational undertakings, the success of which may be still in question.
>
> (Ryle 1949: 222)

it to have. For someone to understand *an expression*, i.e. an utterance type, on the other hand, some other success criterion has to be employed. In this case there is in general no speaker available who has produced the *expression* and whose intent we might cash out understanding in terms of, since speakers (who are not divine) cannot produce utterance-types.

Instead, I will assume – which seems appropriate in the present context – that for someone to understand an expression she must assign a meaning to that expression that *match what it means to the members of a certain speech community* (i.e. match what it means to a certain group of speakers).[25] It seems natural to say, for instance, that when a child approximates closely enough the meaning assigned to some expression by her parents, that she *understands* that expression. On this view, the claim '*S* understands *e*' is elliptical for '*S* understands *e* relative to speech community *c*'. This is welcome since claims about the understanding of expressions, just like claims about the meanings of expressions have to be relativized to make sense. Thus *S* will understand 'Banks bar smoking' if it means to her that banks prohibit smoking *relative a group of monolingual English speakers*, but even if this is the case she might not understand 'Banks bar smoking' relative to a group of monolingual Swedish speakers if it does not mean to her that someone named Banks was wearing a tuxedo.[26]

The main reasons in favor of this explication of 'understanding' is its simplicity, it being neutral with respect to what meanings are, and that it tends to make true (and make sense of) the things that one has maintained in terms of understanding in the debate on the underlying support for compositionality.

It is important to note that by holding that a speaker *S* must assign a meaning to an expression *e* that *matches* what it means to the members of a certain speech community in order for *S* to understand

---

[25] Horwich (1997; 2005) advance a similar view about understanding. However, since he thinks that the meaning properties of complex expressions are constituted by construction properties (see the previous chapter), he thinks that assigning matching construction properties are sufficient for understanding. This view of understanding presupposes compositionality so it would be inappropriate to adopt it here.

[26] If one wants to make more fine grained distinctions and wants to say that a speaker understands something in English, for instance, but only has access to bilingual speakers one can instead relativize to *a subset* of the meaning assignments of a group of speakers.

*e*, I mean that the relevant meanings must be *similar*, not that they must be identical. The evidence in favour of the understanding of expressions taking place is ultimately built on our ability to coordinate and predict behavior, and this form of data is not, it seems to me, powerful enough to warrant making the assumption that speakers assign exactly the same meanings to expressions. The predicted behavior does usually not exhaust all aspects of the meaning. Consider for instance the case where assigning a meaning to a general term is just a matter of assigning a certain extension to utterances of this term. I might accumulate evidence that several things that I take to be in the extensions of a certain term also belong to the extensions of that term according to you. But I'm almost never in a position to verify that we completely agree on the extensions.

   To simplify the rest of the discussion in this chapter, I'll assume that the speech community relative to which the understanding of our speaker *S* is understood consists only of a single English speaker *T*. This will make the exposition easier, and I do not think that anything substantial depends on this assumption. With this assumption in place our original explanandum is transformed into the following.

(LC$^U$) *S* and *T* assign similar meanings to novel expressions.

This transformation is explanatory relevant since it means that the explanandum that we hitherto have assumed to be *a particular event*, is more adequately described as *a kind of event*. For *S* and *T* assigning a similar meaning to a certain expression *e* is a matter of a speaker (of the community consisting of *S* and *T*) assigning a meaning of type *t* to *e*, where *t* is some fairly narrow type encompassing the meanings assigned to *e* by *S* and *T*. So explaining $LC^U$ is really explaining a type of event on a par with explaining why struck matches light for instance, or why certain molecules combine in a certain way. (This had been even clearer if we had required the meaning assignments to be identical, but it is true even if we do not.)

   Going from the earlier form of explanandum to the new one means a change in what is explanatory required. Previously we only needed details from causal histories, now we need details *that are common* to several causal histories. A *prima facie* difficulty arises from this fact. We have assumed that semantic theories are about single speakers,

but this is problematic if we are required by our explanatory standard to provide information about the causal histories of several speakers. But note that what the standard requires is rather that we *give information* that is common to several causal histories not that we *explicitly describe* several causal histories. So given that we assume that the speaker that a semantic theory describes is *a typical speaker of her community* (in our case, of the community consisting of $S$ and $T$), we thereby provide information that is common to several speakers. And it is thus sufficient to correctly characterize a single typical speaker in order to explain $LC^U$.[27]

### 3.1.5 Possible Explanations of why Speakers Understand Novel Expressions

We can now turn to the question of what we need to assume about $S$ and $T$ in order to explain why they assign similar meanings to novel expressions. We noted above (at the beginning of Section 3.1) that the novel expressions to which $S$ and $T$ assign similar meanings are a subset of the complex expressions. So what we are interested in is why $S$ and $T$ are similarly disposed to assign meanings to utterances of novel *complex* expressions.

For purposes of exposition let's assume that (1) is one of the novel expressions that $S$ and $T$ assigns similar meanings to and let's refer to the novel expressions that $S$ and $T$ assign similar meanings to as *the N's*. So, (1) is one of the N's.

(1)     Dogs bark.

Since $LC^U$ is a kind of event, our preferred explanatory standard mandates that we need to provide details that are *common* to the causal histories of $S$ and $T$'s meaning assignments. What does $S$ and $T$ have in common that can be cited in an explanation of $LC^U$?

Had $S$ and $T$'s meaning assignments been common to all members of their species, $LC^U$ might have been explained in terms of $S$ and $T$'s common genetic endowment. But not all speakers understand (1) relative to $T$ (roughly: the speakers not competent in English do not) so any explanation assuming that the meaning assignment

---

[27] But see below for a *prima facie* problem that this leads to in conjunction with $K^{#1}$ theories.

corresponding to (1) is due only to genetic hardwiring, would clearly provide misinformation about $S$'s and $T$'s causal histories and thus (according to the adopted explanatory standard) constitute an unsatisfactory explanation. Such an explanation would predict that understanding is much more frequent than it actually is: all human speakers would understand each other (with respect to the relevant expressions).

Had the N's been familiar expressions, $LC^U$ could have been explained in terms of $S$ and $T$'s common exposure to utterances of the N's and by way of common inductive inferences to the meanings of the N's. But the N's are novel, so any explanation of $LC^U$ by invoking prior exposure of $S$ and $T$ to utterances of the N's would provide misinformation about $S$ and $T$'s causal histories and thus (according to the adopted explanatory standard) constitute an unsatisfactory explanation.

But it is certain (since we know that $S$ and $T$ assign similar meanings to novel complex expressions) that the utterances that $S$ and $T$ have been exposed to are not without commonalities. If $S$ and $T$ assign similar meanings to (1) for instance, it seems safe to say that they must have been exposed to utterances of 'dogs' and 'bark' and utterances with the same mode of composition as that of (1). More generally it seems safe to say that $S$ and $T$ must have been exposed to utterances of the atomic parts of the N's and to utterances having the modes of compositions of the N's.

We cannot expect a semantic theory to provide details about what expressions a speaker has been exposed to – so these commonalities cannot be used in an explanation. But it is a very small step from recruiting, for an explanation of $LC^U$, the commonalities among the utterances that $S$ and $T$ have been exposed to, to recruiting commonalities in how $S$ and $T$ assign meanings to atomic expressions and they assign modes of compositions to complex expressions. And this is something that a semantic theory can provide details about. The explanation of why $S$ and $T$ assign similar meanings to the N's would thus be in terms of $S$ and $T$ assigning similar meanings to the parts of the N's and they assigning similar modes of composition to the N's, *and* them assigning the meanings they assign to the N's *causally depending* on $S$ and $T$ assigning the meanings they do to the parts of the N's and $S$ and $T$ assigning the modes of composition that they do to the N's.

Reflecting on the first two unsatisfactory explanations, we can say that this explanation is *balanced*: the causal histories it describes involves a dependence (of *S* and *T*'s meaning assignments to novel expressions) on enough experience with past linguistic transactions so that it does not entail that too many speakers have *S*'s and *T*'s particular linguistic abilities (the ability to understand the particular novel expressions that they are able to understand), but not so much that it presupposes that utterances of all expressions that *S* and *T* assign similar meanings to are part of their causal histories. In short, a balanced explanation entails that *S* and *T* assign similar expressions to the N's without entailing that all speakers will do so. What is built into the explanation is that *S* and *T* assigning similar meanings to novel expressions is conditional on them having been exposed to similar utterances. Conditionalizing in this way is crucial in order to have a causal account that does not overgeneralize (and thus is unsatisfactory). It seems to me that giving a balanced account is the central challenge with causally explaining $LC^U$.

Considerations of a different sort also point in the direction of the previous explanation.

We know that the meaning *S* and *T* assign to a complex expression *depend* on the meaning they assign to its parts, since the meaning they assign to a complex expression often change when the meaning they assign to one of its parts change – (1) would mean something different to *S* and *T* if 'bark' meant to them what 'meow' means to us.

We also know that the meaning *S* and *T* assign to a complex expression *depend* on the mode of composition they assign to it since the meaning they assign to a complex expression often change when the mode of composition they assign to it is changed – 'happy men and women' means one thing to *S* and *T* if assigned a mode of composition along the lines of [[happy [men and women]] and another if assigned a mode of composition along the lines of [[happy men] and [women]].

Since the N's are complex, we know that the meanings *S* and *T* assign to the N's depend on the meanings that *S* and *T* assign to the parts of the N's and what modes of composition that *S* and *T* assign to the N's. This suggests, again, that $LC^U$ should be explained in terms of *S* and *T* assigning similar meanings to the parts of the N's and they assigning similar modes of composition to the N's, *and* them assigning the meanings they do to the N's *causally depending* on

*S* and *T* assigning the meanings they do to the parts of the N's and *S* and *T* assigning the modes of composition that they do to the N's. It thus seems beyond doubt that explanations of $LC^U$ must *at least* make reference to meanings of parts and modes of composition. With this in place, we can turn to explanations in terms of semantic theories.


*Explanations in terms of Compositional theories*

A theory in accordance with $C^{FI}$ such as the following can provide an explanation of the aforementioned kind.

$t_4$    *Syntax:*                    *Semantics:*

| | | | |
|---|---|---|---|
| N | → | 'dogs', | $\mu$('dogs') | = {x : x is a dog} |
| | | 'cats' | $\mu$('cats') | = {x : x is a cat} |
| V | → | 'bark', | $\mu$('bark') | = {x : x barks} |
| | | 'meow' | $\mu$('meow') | = {x : x meows} |
| NP | → | N | $\mu(_{NP}$ N) | = $\mu$(N) |
| VP | → | V | $\mu(_{VP}$ V) | = $\mu$(V) |
| S | → | NP·VP | $\mu(_{S}$ NP VP) | = $f(\mu$(NP), $\mu$(VP)) |

where…
$f(x, y) = 1$ if $x \subseteq y$ else $0$


This theory explains $LC^U$ on the following assumptions; i) the theory is interpreted to be about *S*, ii) the theory is such that the information it provides about *S*, at least to some extent, applies to *T* as well, iii) the N's are a subset of the complex expressions covered by the theory, iv) the theory gets the structure and meaning assignments it attributes to *S* and *T* right, v) the implicit semantic rules are interpreted as providing information about causal dependencies, vi) the causal dependencies that the implicit semantic rules provide information about are correct (e.g. *S*'s meaning assignments to 'Dogs bark' and 'Cats bark' really have a common partial cause, 'Cats meow' and 'Cats bark' really have a common partial cause etc.)

Given these assumptions, $t_4$ explains $LC^U$ by giving details that are common to the causal histories leading up to *S*'s and *T*'s assigning similar meanings to novel complex expressions. Importantly, the theory provides causal details (through the mode of composition specific implicit semantic rules in terms of the meanings of parts of

expressions) that indicate a dependence of *S* and *T*'s meaning assignments on common experience (having been exposed to certain kinds of expressions) which makes the explanation balanced: it explains why *S* and *T* assign similar meanings to novel expressions although not all speakers do this. More specifically, a theory like $t_4$ does this by attributing mechanisms to *S* (and indirectly to *T*) such that what meanings she assigns to complex expressions depends on what meanings she assigns to atomic expressions and what modes of composition she assigns to complex expressions.[28]

Not only does $t_4$ explain $LC^U$ on the previous six assumptions, *all* theories in accordance with $C^{FI}$ explain $LC^U$ on these assumptions. It will be useful to stipulate that a *constraint X* explains a phenomena *Y* iff all theories in accordance with *X* explain *Y* (under a given set of assumptions). Given this definition we have made good on the promise given in section 3.1.2: $C^{FI}$ is an explication of compositionality that is able to explain things that compositionality is expected to explain.

We can also spell out the claim made in that section that $C^{FE}$ is 'objectionally weak' from an explanatory perspective. For given the aforementioned assumptions it is not the case that all theories in accordance with $C^{FE}$ explain $LC^U$. Hence $C^{FE}$ does not explain $LC^U$.

Consider $t_5$ for instance, a theory in accordance with $C^{FE}$.

---

[28] It is compelling to see a speaker's assigning certain meanings to complex expressions, i.e. her being disposed to assign certain meanings to utterances of these expressions as just *being* her having this mechanism, her assigning certain meanings to atomic expressions and her assigning certain modes of composition to complex expressions. It should be noted that this is not objectionable from the perspective of the present explanatory standard. Since the 'causal history of a particular event includes that event itself, and all events which are parts of it' (Lewis 1986: 185), we can causally explain a speaker assigning certain meanings to complex expressions in terms of him having this mechanism and him making these other assignments even if the explanadum just is the explanans. This does not make the relevant explanation trivial since it breaks down the explanandum into several parts on which it causally depend.

$t_5$     *Syntax:*                    *Semantics:*
      N → 'cow'             $\mu$('cow')          $= m_1$
      A → 'brown',          $\mu$('brown')        $= m_2$
          'black'           $\mu$('black')        $= m_3$
      NP → A·N              $\mu$('brown cow')    $= m_4$
                            $\mu$('black cow')    $= m_5$

Even given the six assumptions above $t_5$ does not explain $LC^U$ (note that v) and vi) are vacuously satisfied). Without the implicit semantic rules, the theory does not provide any details about the causal histories of $S$'s (and $T$'s) assignment of meanings to novel expressions. Or, at least, since the causal history of an event $e$ includes itself, the information provided here is not very ample and the explanation is thus unsatisfactory (according to the present explanatory standard). In particular, the explanation (and the theory) is not balanced, since it gives no causal account of why $S$ understands novel expressions relative $T$ but why not all speakers do this.

   It should also be noted that strengthening $C^{FE}$ by adding a computability requirement to it does not increase its explanatory utility in the present context. A function $f$ is normally assumed to be *effectively computable* if a finite list of instructions can be given that in principle makes it possible to determine the value of $f(n)$ for any argument $n$.[29] But since $t_5$, for instance, only covers a finite number of complex expressions, it satisfies this stronger form of $C^{FE}$ as well.[30] Hence, this stronger form of $C^{FE}$ is not able to explain $LC^U$ either (since theories like $t_5$ which satisfy this principle cannot). What is important for providing the causal connections is not that such a list of instructions relative some theory *can* be given, but that that list *is* given. What is required is thus something closer to $C^{FI}$ than to $C^{FE}$ conjoined with a computability requirement.[31]

---

[29] See Boolos and Burgess et al. (2002).
[30] In particular there exist the following finite list of instructions that makes it possible to determine the value of $f(x)$ for any argument $x$.

$$'r(\sigma(e_0,\ldots,e_n)) = \{m_4\} \text{ if } n=1, \mu'(e_1) = \{m_2\} \text{ and } \mu'(e_0) = \{m_1\}$$
$$\{m_5\} \text{ if } n=1, \mu'(e_1) = \{m_3\} \text{ and } \mu'(e_0) = \{m_1\}'$$

[31] However, a theory in accordance with $C^{FE}$ coupled with the requisite list of instructions is not equivalent to a theory in accordance with $C^{FI}$. The other problems with spelling out compositionality in terms of $C^{FE}$ remain. See Section 3.1.2.

*Explanations in terms of Non-Compositional theories*

Prior to the discussion of theories in accordance with $C^{FI}$ we arrived at the conclusion that $LC^U$ should be explained in terms of $S$'s and $T$'s assigning similar meanings to the parts of the N's and their assigning similar modes of composition to the N's, *and* their assigning the meanings they do to the N's *causally depending* on $S$'s and $T$'s assigning the meanings they do to the parts of the N's and $S$'s and $T$'s assigning the modes of composition that they do to the N's. This was motivated by the fact that we need a balanced explanation, something explaining why $S$ and $T$ assign similar meanings to the N's even though not all speakers do so. By rooting the speaker's understanding of complex expression in his understanding of parts and his ability to assign modes of compositions, we root this understanding in something which is plausible to think is common for $S$ and $T$ but which is not common to all speakers. Additional support for an explanation along the aforementioned lines was had through fairly direct evidence for the fact that the meaning-assignments of $S$ pertaining to complex expressions depended on $S$'s meaning-assignments involving the parts of these expressions, and $S$'s mode of composition assignments.

But note that the preferred explanation, and the reasons in favor of it, only point towards an explanation in terms of a *dependence* of $S$'s meaning assignments involving complex expressions on her meaning assignments to the parts of these expressions and her assigning certain modes of composition to the complex expressions. The compositional explanation given in the previous subsection is one example of one such explanation – one where the meanings assigned to complex expressions *only depend* on the modes of compositions assigned to them and the meanings assigned to their parts – but there are many others. We can imagine alternative explanations that will feature, alongside the meanings assigned to parts and modes of composition, additional determinants.[32] So far no reasons have been given against such explanations.

---

[32] Jeff Pelletier has proposed the idea that much more might enter into the determination of the meaning of complex expressions than the meanings of their parts and their modes of composition. He has argued that the requirement that a semantics should be compositional should be replaced with the requirement that it is grounded, i.e. that it is (if I understand him correctly) describable in terms of non-inductive, non circular recursive functions. He suggests that a

The following three sub-sections will give three alternative determination principles, which all have the same explanatory utility (with respect to $LC^U$ and the present explanatory standard) as do $C^{Fl}_i$. The determination principles that will be considered here are far from exhaustive and might not even be completely disjoint. I bring attention to these particular principles since they will prove to be especially interesting in Chapter Five and Chapter Six.

*Adding Form as a Determinant*

The first class of non-compositional explanations that we will consider contain those explanations where the meanings a speaker assigns to complex expressions depend on the meanings she assigns to parts of those expressions, the modes of composition she assigns to those expressions *and* the form she perceives them to have. Although there is really a range of alternatives here, we can, for simplicity, use 'form' in the sense of physical shape. Since expressions just are types of shapes, the account amounts to a determination of meaning where the expressions themselves contribute to the meaning of complex expressions. On this account, (formally) distinct complex expressions that have the same modes of composition, and parts with the same meanings can be assigned different meanings. So an account of this type will allow for a semantic difference between, for instance, 'seeks dogs' and 'seeks canines' even though the two expressions have parts with the same meanings and they have the same mode of composition (they are both simple verb-phrases consisting of a transitive verb and a common noun) since 'seeks dogs' and 'seeks canines' are different expressions i.e. are different physical shapes. An account like this might not only be interesting from the point of view of intensional

---

> [S]emantic evaluation, in general, can bring it to play all kinds of facts, all kinds of information, it could bring in context, it could bring in inferences, it could bring in world knowledge to evaluate an expression, where these facts, etc. are not part of the meanings of the parts of the expression (and they are furthermore not dictated by the "method of combination" used to construct the expression).
>
> (Pelletier 2004: 149)

The following sub-sections spell this out in more detail.

verbs such as 'seeks' but might also, as will be illustrated below, be used to treat constructions involving quotation marks.

The general form of determination that is involved here can be captured in terms of the following determination principle (which was considered briefly at the end of Section 3.1.2) which is similar to, but importantly more permissive than, $C^{FI}$.

> A semantic theory $t$ accords with $F^{FI}$ iff

> i)   There are no explicit semantic rules for complex expressions in $t$.
> ii)  Each implicit semantic rule in $t$ is of the form '$\mu(x)=y$' where

>> 1)   '$x$' is a variable over expressions, and
>> 2)   '$y$' is a *f-term*, where a *f-term* is defined in the following way:

>>> a)   '$\mu(x_1)$' is a *f*-term if '$x_1$' is one of the variables from which '$x$' is built up.
>>> b)   '$x_1$' is a *f*-term if '$x_1$' is one of the variables from which '$x$' is built up.
>>> c)   '$g(x_1,\ldots, x_n)$' is a *f*-term if '$x_1$',…,'$x_n$' are *f*-terms and '$g$' is a functor corresponding to a humanly computable function defined on meanings and/or expressions.[33]

The main difference between $F^{FI}$ and $C^{FI}$ is that $F^{FI}$ permits but $C^{FI}$ forbids that variables over expressions feature on the right hand side of implicit semantic rules outside the scope of an interpretation function.

To illustrate how a non-compositional (in the sense of $C^{FI}$) theory in accordance with $F^{FI}$ can explain $L^{CU}$ we can consider the following theory.

---

[33] The initial '$F$' of '$F^{FI}$' is short for 'Form'.

$t_6$    *Syntax:*              *Semantics:*

| | | | |
|---|---|---|---|
| N | → 'dogs', | $\mu$('dogs') | = {x: x is a dog} |
| | 'canines' | $\mu$('canines') | = {x : x is a dog} |
| $P_1$ | → 'is an expression' | $\mu$('is an expr…') | = {x: x is an expr…} |
| $P_2$ | → 'are nice' | $\mu$('are nice') | = {x : x is nice} |
| E | → N, $P_1$, $P_2$, Q, S | $\mu(_E X)$ | = $\mu(X)$ [34] |
| Q | → '''·E·''' | $\mu(_Q$ ''' E ''') | = $g(E)$ [35] |
| S | → Q·$P_1$ | $\mu(_S Q P_1)$ | = $f(\mu(Q), \mu(P_1))$ |
| S | → N·$P_2$ | $\mu(_S N P_2)$ | = $f(\mu(N), \mu(P_2))$ |

where…
$g(x)$    = {y : y=x}
$f(x, y)$ = 1 if $x \subseteq y$ else 0

Since $t_6$ contains the implicit semantic rule '$\mu(_Q$ ''' E ''')= $g(E)$' the right hand side of which is not a *c*-term, $t_6$ does not accord with $C^{FI}$.[36] In fact, a theory like $t_6$ does not even accord with $C^{FE}$. Yet a theory like $t_6$ can explain $LC^U$ on the same assumptions that a theory like $t_4$ can do so; i) the theory is interpreted to be about *S*, ii) the theory is such that the information it provides about *S*, at least to some extent, applies to *T* as well, iii) the N's are a subset of the complex expressions covered by the theory, iv) the theory gets the structure and meaning assignments right, v) the implicit semantic rules are interpreted as providing information about causal dependencies, vi) the causal dependencies that the implicit semantic rules provide information about are correct. Given these assumptions, $t_6$ and all the other theories in accordance with $F^{FI}$, explains $LC^U$ by giving details that are common to the causal histories of *S* and *T* (in this case them assigning certain meanings to simple expressions and certain modes of compositions to complex expressions, them perceiving expressions in

---

[34] The semantic rule corresponding to the syntactic rule 'E → N, $P_1$, $P_2$, Q, S' is just an identity function on meanings regardless of what syntactic category that is involved. '''

[35] The quotation marks '' and ''' are enclosed in quotation marks in order not to 'trap' the concatenation functors and the syntactic variable.

[36] There are theories of quotation that are in accordance with compositionality (see Davidson ([1979] 2001) for instance). The point of this section is not that non-compositional theories in accordance with $F^{FI}$ are better than theories in accordance with $C^{FI}$, but that both can provide explanations of $LC^U$.

the same way, and them employing certain implicit semantic rules) that lead to their converging meaning assignments.

Since it is likely that two speakers that assign the same mode of composition to an expression and the same meanings to the parts of that expression also perceive the same expression in the same way, adding that perceived physical shape of an expression might be part of the determination of the meanings of complex expressions does not seem to lead to the prediction that understanding is any less scarce than a theory in terms of $C^{FI}$ would predict that it is.

In addition, non–$C^{FI}$ theories in accordance with $F^{FI}$, just like theories in accordance with $C^{FI}$, provides causal details (through the implicit semantic rules) that indicate a dependence of $S$'s and $T$'s meaning assignments on common experience (having been exposed to certain kinds of expressions). It is thereby indicated why even though $S$ and $T$ assign similar meanings to novel expressions, not all speakers do this.

### Adding Ectal Dispositions as Determinants

Since the semantic theories we are considering provide explanations of $LC^U$ when they are interpreted to be about speakers, other aspects of the minds of speakers might also be recruited in these explanations. We noted in Subsection 3.1.3 that a speaker assigning a meaning to an expression seems to be a matter of her being disposed to assign certain meanings to utterances of that expression. One way to characterize dispositions like these is in terms of functions from contexts to (utterance) meanings. Speakers are disposed to react in a range of different ways in conversational contexts and not all of these dispositions correspond to that speaker assigning a meaning to an expression. Yet, there does not seem to be anything that rules out that a semantic theory cites these kinds of dispositions, what we can call *ectal dispositions*, in an explanation of $LC^U$. A theory citing such dispositions might be interesting since it might be well-equipped to deal with what has been referred to as *unarticulated constituents*.[37]

John Perry has drawn attention to the fact that the utterances of certain expressions seems to express propositions that contain parts that do not correspond to any part of the expression being uttered.

---

[37] See (Perry [1986] 2000).

For instance, utterances of simple weather reports such as 'It rains' seems to express the proposition that it rains at a certain time *at a certain place*. But although the relevant time is signified by the tense of the verb, and the relevant kind of event is signified by the verb, there is nothing encoding the place. The relevant place seems to be an *unarticulated constituent* of the proposition expressed. Since English speakers understand expressions like 'It rains' they must have some way of systematically arriving at the propositions expressed by utterances of these expressions. One way to account for these particular forms of unarticulated constituency is by attributing to the speaker a disposition to attend to certain places depending on which context she is in. This disposition can then influence the meaning the speaker assigns to expressions like 'it rains', i.e. her disposition to assign meanings to utterances of 'It rains'.[38] These dispositions can formally be modeled as contextual functions; functions from contexts to extensions. These functions bear affinities to Kaplanian characters, but need not be the meaning of any expression. The disposition relevant to the example would then be represented as contextual function from contexts to places. Before illustrating how these functions can be used in a semantic theory we can detail a determination claim that is permissive enough to allow for contextual functions to play a role in semantic theories.

A semantic theory $t$ accords with $X^{FI}$ iff

i) There are no explicit semantic rules for complex expressions in $t$.

ii) Each implicit semantic rule in $t$ is of the form '$\mu(x)=y$' where

1) '$x$' is a mode of composition variable, and
2) '$y$' is a *x-term*, where a *x-term* is defined in the following way:

    a) '$\mu(x_1)$' is a x-term if '$x_1$' is one of the variables from which '$x$' is built up.

---

[38] The issue of unarticulated constituency is discussed in more detail in Chapter Five.

b) '$g$' is a $x$-term if '$g$' refers to a contextual function.

c) '$g(x_1,\ldots,\ x_n)$' is a $x$-term if '$x_1$',…,'$x_n$' are $x$-terms and '$g$' is a functor corresponding to a humanly computable function defined on the set of meanings and/or a set of contextual functions.[39]

The difference between $X^{FI}$ and $C^{FI}$ resides in the former allowing for reference to contextual function while the latter does not (unless they are the meanings of expressions).

   The following theory, which is in accordance with $X^{FI}$ but not with $C^{FI}$, illustrates how contextual functions can play a role in explaining $LC^U$.[40]

$t_7$   *Syntax:*                    *Semantics:*

V $\rightarrow$ 'rain', 'snow'    $\mu$('rain')    $= \lambda c \in C.$ the rain property

E $\rightarrow$ 'It'              $\mu$('snow')   $= \lambda c \in C.$ the snow property

t $\rightarrow$ '$t_1$',           $\mu$('$t_1$')   $= \lambda c \in C.$ the time prior to $c$

  '$t_2$',                       $\mu$('$t_2$')   $= \lambda c \in C.$ the time of $c$

  '$t_3$'                        $\mu$('$t_3$')   $= \lambda c \in C.$ the time after $c$

S' $\rightarrow$ E·V            $\mu(_{S'}\text{E V})$   $= \mu(V)$

S $\rightarrow$ S'·t           $\mu(_S \text{S' t})$   $= f(\mu(S),\ \mu(t),\ g)$[41]

where…

$C$ is a set of contexts

$g = \lambda c \in C.$ a salient location in c
    [a *contextual function*]

$f(x,\ y,\ z) = \lambda c \in C.\ <x(c),\ y(c),\ z(c)>$

Since $t_7$ contains the implicit semantic rule '$\mu(_S \text{S' t}) = f(\mu(S),\ \mu(t),\ g)$' where the right-hand side is not a $c$-term (since it contains the functor

---

[39] The initial '$X$' of '$X^{FI}$' is short for 'conteXt'.

[40] Weiskopf (2007) proposes an analysis of the semantics of noun-noun compounds that would also be an example of this kind of theory. Weiskopf himself argues that his view is compositional, but he does not precisely state a version of compositionality which his view is in accordance with.

[41] For simplicity I treat tense inflection as concatenation (where the t's represent tenses)

'g' corresponding to a contextual function), $t_7$ does not accord with $C^{FI}$.[42] Yet a non-compositional theory like $t_7$ can, just like the previous two theories, explain $LC^U$ given that the appropriate assumptions are in place. To repeat; i) the theory is interpreted to be about $S$, ii) the theory is such that the information it provides about $S$, at least to some extent, applies to $T$ as well, iii) the N's are a subset of the complex expressions covered by the theory, iv) the theory gets the structure and meaning assignments, as well as the ectal disposition-assignments right, v) the implicit semantic rules are interpreted as providing information about causal dependencies, vi) the causal dependencies that the implicit semantic rules provide information about are correct. Like the previous two kinds of theories, theories in accordance with $X^{FI}$ explain $LC^U$ by attributing mechanisms to $S$ (and indirectly to $T$) such that what meanings she assigns to complex expressions depends on what meanings she assigns to atomic expressions and what modes of composition she assigns to complex expressions. In addition, some $X^{FI}$ theories invoke non-linguistic dispositions and suggest a dependence on these as well.

Like non-$C^{FI}$ theories in accordance with $F^{FI}$, non-$C^{FI}$ theories in accordance with $X^{FI}$, are balanced to the same extent that $C^{FI}$ theories are. The non-linguistic functions might themselves provide balance because the speaker having these might also be a matter of generalizations across past utterances. So these theories will *qua* featuring these functions be able to explain why $S$ and $T$ assign similar meanings to the N's even though not all speakers do this.[43] So $X^{FI}$ theories give balanced explanations.

---

[42] Again, there are compositional accounts of the constructions involving unartic-ulated constituents (see Pagin (2004) for instance). But the idea is not that the non-compositional $X^{FI}$ theories are better than the alternative $C^{FI}$ theories, only that both kinds of theories can explain $LC^U$.

[43] Note that it is ok, *ceteris paribus*, to also assume that the dispositions corresponding to the contextual functions are genetically hardwired in cases where the contextual functions are only a part of the right hand side of the implicit semantic rules and these also featuring a dependence on additional linguistic competence (such as, for instance, knowing what 'rains' mean). We do not thereby overgeneralize a linguistic competence (such as understanding 'It rains') since it in part depends on the speaker having been exposed to certain things.

## Adding General Knowledge as a Determinant

Having gone as far as an explanation of $LC^U$ in terms of a theory invoking certain features of the speaker's attention, we can easily imagine other determination claims that allow other features of a speaker's cognition to be cited. It will be particularly worthwhile to look closer at the extent to which the *general knowledge* of a speaker can influence her meaning assignments, where by 'general knowledge', I mean the not necessarily true beliefs that the speaker has about general matters of fact (brown cows being dangerous, black cats bringing bad luck etc).[44]

We can formulate a criterion which is permissive enough to allow reference to general knowledge within a semantic theory in the following way.

A semantic theory *t* accords with $K^{FI}$ iff

i) There are no explicit semantic rules for complex expressions in *t*.

ii) Each implicit semantic rule in *t* is of the form '$\mu(x)=y$' where

1) '$x$' is a mode of composition variable, and
2) '$y$' is a *k-term*, where a *k-term* is defined in the following way:

a) '$\mu(x_1)$' is a *k*-term if '$x_1$' is one of the variables from which '$x$' is built up.

b) '$b$' is a *k*-term if '$b$' refers to a state of general knowledge.

c) '$g(x_1,\ldots, x_n)$' is a *k*-term if '$x_1$',…,'$x_n$' are *k*-terms and '$g$' is a functor corresponding to a humanly computable function defined on the set of meanings and/or a set of general knowledge.[45]

---

[44] Using 'knowledge' in this way is somewhat misleading from a philosopher's perspective, but I'm borrowing the term from cognitive psychology where 'knows that p' is frequently used without any commitment to 'p'.

[45] The initial '$K$' of '$K^{FI}$' is short for 'Knowledge'.

I'm going to restrict myself in the sequel, when exemplifying non-compositional explanations in terms of $K^{FI}$ to a subclass of a speakers' general knowledge. In particular, I will employ what is sometimes referred to as *extensional feedback* in the literature on cognitive psychology, a particular form of general knowledge.[46] This is a form of knowledge regarding the objects denoted by complex expressions which is obtained by interaction with these entities. One might for instance, have extensional feedback from interacting with black cats to the effect that black cats generally bring bad luck.

Given this class of general knowledge we can exemplify a non-compositional theory in accordance with $K^{FI}$ in terms of inferential roles. According to this theory the meanings of expressions are sets of inferences. The inferences are described by statements of the form "'A'→ 'B'" which denote the speakers disposition to believe B if she believes A.

$t_8$    *Syntax:*
     N    →  'cow', 'cat'
     A    →  'brown', 'black'
     NP  →  A·N

   *Semantics:*
   $\mu$('cow')     = {  'x is a cow' → 'x is an animal',
                       'x is a cow' → 'x produces milk',
                       …                           }
   $\mu$('cat')      = {  'x is a cat' → 'x is an animal',
                       'x is a cat' → 'x meows',
                       …                           }
   $\mu$('brown')   = {  'x is brown' → 'x is colored',
                       'x is brown'→ 'x is non-black',
                       …                           }
   $\mu$('black')    = {  'x is black' → 'x is colored',
                       'x is black'→ 'x is non-brown',
                       …                           }
   $\mu(_{NP} A N)$     = $\mu(A) \cup \mu(N) \cup f(\mu(A) \cup \mu(N), b)$

---

[46] See Hampton (1988).

where…

 *b* is a set of extensional feedback

*f*(*x*, *y*)=     the set of inferences corresponding to the subset of *y* that is the   extensional   feedback which is about the things that the inferences in *x* pertain to.

*f* can be illustrated in the following way: if we assume that the speaker has, for instance, come across animals of a certain color that produce milk and learnt about them that they are dangerous, *f* will yield something like ' → "x is dangerous"' as value when fed the corresponding inferences as arguments.

$t_8$ does not accord with $C^{FI}$ since it contains the implicit semantic rule '$\mu(_{NP}$ A N)= $\mu$(A)$\cup\mu$(N)$\cup f(\mu$(A)$\cup\mu$(N), *b*)' the right hand side of which is not a *c*-term since it contains the term '*b*'. Yet like the previous three theories $t_8$ seems to be able to explain $LC^U$ given something similar to the now familiar assumptions; i) the theory is interpreted to be about *S*, ii) the theory is such that the information it provides about *S*, at least to some extent, applies to *T* as well, iii) the N's are a subset of the complex expressions covered by the theory, iv) the theory gets the structure  assignments, the meaning assignments *and the belief set* right, v) the implicit semantic rules are interpreted as providing information about causal dependencies, vi) the causal dependencies that the implicit semantic rules provide information about are correct. Extensional feedback could thus feature in explanations of $LC^U$.

In contrast with the previous three kinds of theories it is not obvious that a theory in accordance with $K^{FI}$ can provide plausible explanations of  $LC^U$. There are at least two different worries, both having to do with the fact that speakers often differ greatly in what they believe.

First, there is the worry that since the explanatory standard we have adopted requires that we give details that are *common* to the relevant causal histories of *S* and *T*, a theory invoking the belief set of *S* will not be able to provide an explanation if *S*'s and *T*'s belief sets are very different. Since *S*'s and *T*'s belief sets often will be very different, the worry goes, $K^{FI}$ theories will often fail to explain $LC^U$. However, even though *S* and *T* differ in respect to their belief sets, this does not compromise the fact that a $K^{FI}$ theory can provide details that are

common to the causal histories leading up to $S$ and $T$'s meaning assignments. $S$ and $T$ can still be similar with respect to the meanings they assign to atomic expressions, the modes of composition they assign to complex expressions, the semantic rules they employ and thus, importantly, *how their meaning derivations are sensitive to general knowledge*. What the explanatory standard requires is that common details are provided, not that *only* common details are provided. And even if some details do not apply to $T$, correct $K^{FI}$ theories still provide ample details that are common to the causal histories of the two speakers. So the $K^{FI}$ theories do not seem to be objectionable in this respect.

Second, there is the worry that if understanding is defined in terms of meaning and meaning is dependent on general knowledge, then $K^{FI}$ theories will predict that there is much less understanding then there actually is, since speakers vary significantly in what they believe.

Two things should be noted here. First, the evidence in favor of speakers generally understanding novel expressions is not strong enough to dismiss out of hand that misunderstanding due to differences in general knowledge actually takes place. For even if speakers are often successful in coordinating their behavior with other similar speakers by means of novel expressions, it does seem like large differences in general knowledge, like that between people of different social, cultural or professional background, often leads to less successful coordination. Consider for instance the communicative problems that can ensue when researchers from different but related disciplines attempt to collaborate, or the difficulties inherent in reading something written by someone with a radically different understanding of the world (someone from a completely different time or place for instance). Difficulties like these does not establish that something like $K^{FI}$ is correct directly, since *prima facie* differences in the understanding of complex expressions might be traceable to differences in meaning assignments to simple expressions, deviating implicatures or different codes of conduct. But they do suggest that there might be a case to be made for differences in general knowledge sometimes leading to genuine misunderstanding. In addition, the evidence that is available to us to determine whether understanding or misunderstanding obtains is far from conclusive. Not all aspects of the meaning assigned to an expression are manifest in all utterances of that expression, and those that are not can hide misunderstanding.

Say that meanings are sets of properties, and that the meaning of 'cat' is just the property of being a cat, and the meaning of 'black' is just the property of being 'black' but the meaning of 'black cat' is the property of being black, a cat, and something which brings bad luck. In situations where communication takes place in terms of 'black cat' but luck is not an issue, potential deviation in meaning assignments would be hidden from view.

   The second thing to note, and this is even more important, is that the amount of influence general knowledge has on understanding depends on the specifics of particular $K^{FI}$ theories. It is not necessary that a non-compositional $K^{FI}$ theory predicts more misunderstanding than a $C^{FI}$ theory. The proponent of a $K^{FI}$ theory does not need to hold that all differences in general knowledge will result in differences in the meaning assignments to complex expressions, just like the proponent of $C$ is not committed to say that all composition rules are structure specific. She is free to hold that the meaning of both, say, adjective-noun combinations and adverb-verb combinations are derived in the same way (through function application for instance). The amount of influence general knowledge has on the meaning of complex expressions is dependent on the specifics of the $K^{FI}$ theories.

   More importantly, and the main reason I believe, for thinking that there definitely can be adequate non-compositional $K^{FI}$ theories is that in order for $S$ and $T$ to understand a certain expression $e$ the meanings they assign to $e$ must be *similar*, but not necessarily *identical*. So given that $S$ and $T$ differ with respects to their beliefs about the denotation of some complex expression $e$ covered by $t_7$, $t_7$ is still compatible with $S$ and $T$ understanding $e$ since we are, I think, at liberty to say in general that they still assign *similar* meanings to that expression. $t_7$ would then explain this understanding since $t_7$ would give information about common causes of the two speakers meaning assignments (in particular, meaning assignments of parts, modes of composition used, and semantic rule employed). Indeed, how similar two meanings must be in order for understanding to result is largely underdetermined. In the extreme case, the worry that a theory like, for instance, $t_7$ predict that enough understanding takes place can be removed by simply supplementing it with the claim that the meanings of complex expressions of the theory are similar enough for understanding when they are identical with respect to the inferential roles conferred to them by their parts. So given $t_7$, it can be

maintained that $S$ and $T$ understand an arbitrary complex expression $e=e_1 \cdot e_2$ not only when $\mu_S(e)=\mu_T(e)$, (this has already been concluded to be too strong) but when $\mu_S(e)\approx\mu_T(e)$ which can, at least for the purposes of a theory like $t_7$, be precisely defined in the following way: $\mu_S(e)\approx\mu_T(e)$ iff $\mu_S(e_1)\cup\mu_S(e_2)=\mu_T(e_1)\cup\mu_T(e_2)$.[47] Given this additional assumption, $t_7$ gives balanced explanations if and only if a compositional variant of it does. So despite initial worries, it is safe to assume that a non-compositional $K^{FI}$ theory like $t_7$ can give adequate explanations of $LC^U$.

### General Remarks on Attempts to Appoint a Best Explanation

I have exemplified four different kinds of explanations of $LC^U$. It follows that a theory is not required to be compositional in order to explain why speakers understand novel expressions. Considerations of what is required in order to explain $LC^U$ thus gives us no reason to require that a semantic theory must be compositional, rather than accord with one of the other principles.

The compositionalist could argue, at this point, that even though theories in accordance with either of these principles *could* explain $LC^U$, the compositional theories explain $LC^U$ *better* than the alternatives. If this is true we might be justified in requiring that a semantic theory must be compositional in virtue of this. I will offer remarks on this line of thought in this subsection.

I can see roughly three different kinds of reasons for why one would prefer one explanation over another. First, an explanation might be preferred over an alternative due to methodological considerations such as agreed upon protocol for how one is supposed to conduct the relevant form of theorizing. Second, an explanation might be preferred over an alternative due to it, but not the alternative, being compatible with known empirical data. Third, one explanation might be preferable over another due to it being better than the alternative from the perspective of general theoretical properties such as simplicity or logical strength.

Let's consider first methodological assumptions. The only such assumption that I'm familiar with is one that conflicts with $X^{FI}$. By

---

[47] If we want we can define $\approx$ in terms of the meanings of the *ultimate* parts of the expressions thereby ensuring that general knowledge never influence the conditions for when understanding obtains.

considering $X^{FI}$ theories we take a step away from the conception of semantic theorizing advocated by Larson and Segal. On their view, a speaker's semantic competence consists in certain *semantic knowledge*, and according to them, this knowledge only contains information about the *context independent meanings* of expressions. Hence, given an utterance of 'She is here', your semantic knowledge will, according to Larson and Segal, not tell you, for instance, who 'she' refers to or what place is referred to by 'here'. It will only tell you, roughly, that the utterance is true only if some independently identified female is near the speaker.[48] Commitments like these – commitments that pertain to the semantics-pragmatics distinction – relate to what in the entirety of the significance, use and utterance of an expression, should be of concern to semantic theories and what should be left outside (to be dealt with by pragmatic theories). Larson and Segal take a fairly conservative stance here, when they hold that their semantic theory is concerned only with non-relativized meanings, i.e. meanings that are not relativized to something like contexts of utterance.

It seems inappropriate to make this assumption in the present context. By assuming that a semantic theory should not concern itself with contexts of utterance we rule out certain accounts of meaning of expressions. For instance, on their view meanings could not be identified with *characters*, functions from contexts to contents, which are defined explicitly in terms of contexts.[49] Additionally, by assuming that a semantic theory should not concern itself with contexts of utterance we remove possible determinants (such as ectal dispositions) of what meanings speakers assign to complex expressions. The meaning of an expression pertains to a pattern across the meanings of actual and potential utterances of that expression. The pattern corresponding to a complex expression might well turn out to be underdetermined by the patterns corresponding to parts of those expressions (and what mode of composition that is assigned the complex expressions) and depend in addition on other ways in which the speaker is sensitive to context. If we draw the semantics-pragmatics distinction too narrowly we remove this possible form of determination. Removing this possibility *a priori* when it might

---

[48] See Larson and Segal (1995: 21)
[49] See Kaplan ([1977] 1989) and Schiffer (2003) for semantic accounts in terms of characters.

potentially feature in a correct explanation of $LC^U$ is illegitimate and we should thus resist drawing the semantics-pragmatics distinction in the way that Larson and Segal suggest. Doing otherwise would compromise the present part of the inquiry. We are not concerned with the question of whether compositionality is true *on a certain restrictive construal of the semantics-pragmatics distinction*, but with the question of whether compositionality *is true*. It follows that we should be careful not to introduce, as part of the explanatory set-up, methodological constraints on the possible determinants of speakers assigning meanings to complex expressions. If we are not careful we run the risk of not being able to answer the question that we are ultimately interested in. For if we make additional constraining assumptions we would be in a position not to answer what compositionality can tell us about meanings, but only what compositionality, given these additional constraints, can tell us about meanings. Hence, by making use of such assumptions we would potentially beg the question against semantic theories that can provide the requisite explanations, but that do so in violation of the additional assumptions.

Let's consider next empirical assumptions. Again, I'm only familiar with one such assumption that pertains to the foregoing. This is *the modularity thesis*. This assumption conflicts with $K^{FI}$ (and probably $X^{FI}$ as well). By not making it, we will take yet another step away from the conception of semantic theorizing advocated by Larson and Segal.

According to the classical Chomskyan view, a theory of competence describes a certain faculty of the mind (*the language faculty*) which is responsible for the things the speaker is linguistically able to do. Given – what seems to be assumed by Larson and Segal – that we assume that the language faculty is isolated from other mental faculties and from general cognition, only certain kinds of information can be cited by an adequate semantic theory.[50] If for

---

[50] Larson and Segal write the following.

> Semantic theory as we have sketched it here is a component of the larger enterprise of cognitive linguistics [...] Like cognitive linguistics as a whole, it assumes that linguistic competence consists of an unconscious body of knowledge, what Chomsky has termed the "language faculty." Furthermore, it assumes that the language faculty contains a specifically semantic module: a particular isolable domain

instance, the computations taking place in the language faculty are assumed to be cut off from the general knowledge we have of the non-linguistic world, then a correct theory of competence cannot invoke knowledge of this kind.

Although this is not the place for a detailed evaluation of the modularity thesis, I think that there are good reasons to resist the thesis in the present context. The meanings that we are concerned with in this thesis are those invoked when making, for instance, synonymy and ambiguity judgments. But it seems clear that judgments like these are influenced by general knowledge. Consider, for instance, (2).

(2)     The boy left his straw in his drink.

This sentence is ambiguous between at least two readings depending on whether 'straw' is interpreted in terms of 'a tube for sucking up a beverage from a container' or in terms of 'a stalk of threshed grain'. It is also the case that the first reading is the default reading of (2) and that the second is only available after some reflection. It is obvious that the relative accessibility of the two readings is due to our knowledge of the world, in particular about how drinks are commonly consumed. But given that this is true, our ambiguity judgments are influenced by general knowledge. So with respect to the kind of meaning that we have been concerned with in this essay, the modularity thesis is false.[51]

---

of linguistic knowledge beyond phonology, morphology, syntax, etc.,
that is concerned with meaning.

(Larson and Segal 1995: 22)

See Fodor (1983) and Borg (2004) for other proponents of the modularity thesis. We will return briefly to the modularity thesis in Chapter Seven.

[51] This is of course to say nothing about whether there are other semantic notions with respect to which modularity is true, only that it is false with respect to the notions that we are concerned with here.

Note that if the modularity thesis is construed not as an empirical assumption but as a methodological one (see Chapter Seven for a discussion of this distinction) the same kind of considerations that were leveled against the drawing the semantic/pragmatic distinction conservatively, can be used against the modularity thesis: assuming it as part of the inquiry would obscure the goal that we are ultimately concerned with.

Let's finally consider general theoretical properties such as simplicity or logical strength that one might recruit in order to argue that one explanation of $LC^U$ is better than another.

Consider first simplicity. If we have two competing explanations of a particular phenomenon, it could be argued that we should *ceteris paribus* choose the simpler theory.[52] The compositionalist could argue that explanations in terms of $C^{FI}$ are *simpler* than the alternative non-compositional explanations, since the other principles makes reference to more determinants than $C^{FI}$ does.

However, since it is really the particular theories that provide the explanations, it is them rather than the determination theses that should be compared in terms of simplicity. But when we turn to comparisons between theories, which determination principles the theories accord with is just one determinant among many of how simple a particular theory is. This will also depend on, for instance, the number of lexemes of each theory, the number of rules, and the complexity of each rule (which can vary independently of the number of determinants). So it is not true that the theories in accordance with one of the determination principles is generally simpler then the theories in accordance with another. So compositional explanations have no general advantage over alternative explanations when it comes to simplicity. In addition, when we turn to theory comparisons rather than principle comparison there is an important shift in priority. What we are ultimately concerned with is whether or not compositionality can be used to remove certain theories from consideration (and thereby inform us about meanings). But if we turn to simplicity judgments among theories, then the more complex theory would be rejected *qua its complexity* and not due to whether or not it accords with this or that determination principle. So it seems that simplicity considerations has nothing to offer the compositionalist who advances the position that compositionality can be used in order to remove certain theories from consideration.

Consider next logical strength. If we have two competing explanations of a particular phenomenon, then it could be argued

---

[52] I do not want to go along with this argument in the abscence of an argument for why simplicity is truth-conducive. But for the sake of showing that this venue of compositionalist defence is closed anyway, I will assume that some such argument has been given.

that we should *ceteris paribus* choose the one embodying weaker assumptions (since it is more likely to be true).

The determination principles we have been concerned with in this chapter can be partially ordered in terms of logical strength in the following way.

$$F^{FI}$$

$$\uparrow$$

$$C^{FI}$$

$$X^{FI} \qquad\qquad K^{FI}$$
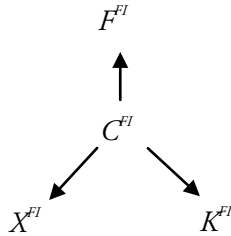
As is demonstrated, if we want to make the weakest possible assumption about a semantic theory which is compatible with it providing an explanation of $LC^{U}$ we are tied between deploying $F^{FI}$, $X^{FI}$ or $K^{FI}$ (or any other form of explanation along similar lines). So considerations of logical strength do not favor $C^{FI}$.

To conclude this section; the requirement that a semantic theory should explain why speakers understand novel expressions does not give us a reason to think that a semantic theory must be compositional. Instead, it gives us a reason to think that one out of a set of determination theses must be true. Since it also seems difficult to argue that the compositional explanations are clearly better than the alternatives, I conclude that no justification for imposing compositionality as a constraint on a semantic theories stems from our desire to explain why speakers understand novel expressions.

## 3.2 Explaining the Production of Novel Expressions

Statements of linguistic creativity frequently involve speakers being able to understand *and* produce novel expressions.[53] But what exactly the second half of this ability is supposed to amount to, and what role, if any, compositionality is supposed to play in explaining it is not clear. This section will attempt an explication of the ability, and

---

[53] See, for instance, Katz and Fodor (1963: 181) and Partee (1970: 61)

once this is done argue that compositionality is neither necessary nor sufficient for a theory to explain why speakers have it.

Speakers are not, strictly speaking, able to produce *expressions*. Since expressions are types, ordinary speakers cannot produce them.[54] In addition, that speakers are able to *articulate certain utterances* is not very interesting from the present perspective. No syntactic-semantic principle such as *C* is needed in order to provide an explanation this.

These considerations lead to the idea that what is interesting from the present perspective is connected to speakers' ability to *assign utterances to utterance- meanings*, i.e. the ability to find an expression which can be tokened in order to express a certain meaning. This ability is not *in itself* very interesting from the present perspective either. If the assignment of utterances to meanings is supposed to be completely unconstrained, i.e. if no additional success criterion is adopted, then there is no explanatory need for a principle ensuring that the assignment is systematic in some way.[55]

What is more interesting is that speakers are disposed to assign utterances to utterance meanings *in a similar way*. If the relevant expressions are familiar and species general, speakers being so disposed, could be explained either in terms of induction or in terms of hardwired dispositions.[56] But if the expressions are novel and not common to all speakers, we need something else. It turns out that we do since it is obvious that i) speakers are similarly disposed to assign utterances of *novel* expressions to *familiar* meanings (e.g. English speakers who assign utterances of 'John killed James' the meaning $m_1$ and are familiar with the relationship between actives and passives, are disposed to occasionally assign utterances to 'James was killed by John' to $m_1$), that ii) speakers are similarly disposed to assign utterances of *familiar* expressions to *novel* meanings (e.g. English speakers who have understood utterances of 'You killed James', are disposed to occasionally assign utterances of 'You killed James' to related but distinct novel meanings), and that iii) speakers are similarly disposed to assign utterances of *novel* expressions to *novel*

---

[54] This is of course unless on thinks that a type can be created by instantiating it for the first time which is true on some theories of types. Even if one thinks this, production is obviously not thought about in this way when one maintains that something like compositionality can explain it.

[55] Cfr. Section 3.3.4.

[56] Cfr. Section 3.2.5.

meanings (e.g. English speakers not familiar with 'James killed John' or the meanings of its utterances nonetheless converge in assigning utterances of 'James killed John' to certain meanings).[57] i)-iii) seem indubitable, and are in as much need of explaining as $LC^U$. Since I assume that i) and ii) will be straightforward to explain once an explanation of iii) is in place I will frame the explanandum of present concern in the following way.

($LC^P$) *S* and *T* are similarly disposed to assign utterances of
novel expressions to novel utterance meanings.[58]

Given no extra assumptions, it seems at first glance that a compositional theory could provide the relevant details. Assume that the following theory correctly characterizes *S* and *T*.

$t_9$  *Syntax:*            *Semantics:*

| | | | |
|---|---|---|---|
| N | → 'dog', | $\mu$('dog') | = {x : x is a dog} |
| | 'cat', | $\mu$('cat') | = {x : x is a cat} |
| | 'bird' | $\mu$('bird') | = {x : x is a bird} |
| A | → 'brown', | $\mu$('brown') | = {x : x is brown} |
| | 'angry', | $\mu$('angry') | = {x : x is angry} |
| | 'loud' | $\mu$('loud') | = {x : x is loud} |
| NP | → A·N | $\mu(_{NP}$ A N) | = $\mu$(A) $\cap$ $\mu$(N) |
| NP | → A·NP | $\mu(_{NP}$ A NP) | = $\mu$(A) $\cap$ $\mu$(NP |

Assume that the meaning assigned to an expression in the theory is also the meanings assigned to utterances of these expressions. Then we can imagine that $LC^P$ is true of *S* and *T* since we can imagine that they when they want to convey a certain meaning (e.g. {x : x is a brown angry bird}) construct the possible expressions one-by-one and using the semantic rules of the theory in order to determine what each expression mean until they find an expression meaning what they want to convey.

However, in spite of the aforementioned account, $t_9$ is not able to provide a satisfactory explanation of $LC^P$. That speakers actually go

---

[57] Note that for some speech communities we have to abstract away from strong personal preferences regarding what kind of expressions to use in which situations.
[58] '$LC^U$' is short for 'Linguistic Creativity – Production'.

through the previously described procedure is very unlikely since it would make production i) incredibly inefficient, and ii) very varied in terms of how much time that is needed (which depends on whether one gets lucky while searching).[59] So $t_9$ is not satisfactory, since it is likely that it provides misinformation about the causal histories of the utterance-assignments it concerns. This means that a theory being compositional (in the sense of $LC^P$) is not sufficient for it being able to explain $LC^P$.

We might be tempted to adopt, instead, something like the inverse of *C* in order to be able to account for efficient, stable production. *IC*, the inverse of *C* is a first approximation of such a principle.

(IC)   The expression of a complex meaning is determined by
        the expression of its parts and its mode of composition.

*IC* is not really what we are after though since 'meaning' as it occurs in *C* pertains to a pattern across meanings of utterances. But it is not meanings of this kind (cfr. Kaplanian characters) that we want to convey, it is the meanings of utterances (cfr. Kaplanian contents).[60] We need instead something like a requirement that the dispositions of speakers to assign certain utterances to certain meanings depend on, for instance, them being disposed to assign certain utterances to other meanings.[61] If we assume that the speakers represent the relevant meanings as structured, we can assume, similarly to the explanations of $LC^U$, that speakers are disposed to assign utterances to certain meanings based on assigning a certain mode of composition to these meanings and assigning certain utterances to their parts. The following theory (a revision of $t_7$) can be used in order to provide an explanation of this kind.

---

[59] The first of these points is one of the main points of Pagin (2003a)
[60] I'm not here assuming that meanings are either characters or contents. I just use Kaplans notions to clarify the threat of an equivocation.
[61] This is very close to a principle suggested by Pagin (2003) for similar reasons.

$t_{10}$ *Syntax:*            *Semantics:*

    V → 'It rains',    $\mu$('It ra..')   = $\lambda c \in C$. the rain property

           'It snows'    $\mu$('It sn..')   = $\lambda c \in C$. the snow property

    t → 't$_1$',        $\mu$('t$_1$')      = $\lambda c \in C$. the time prior to c

         't$_2$',        $\mu$('t$_2$')      = $\lambda c \in C$. the time of c

         't$_3$'        $\mu$('t$_3$')      = $\lambda c \in C$. the time after c

    S → V·t       $\mu(_S\,V\,t)$    = $\lambda c \in C$. $<\mu(V)(c), \mu(t)(c), g(c)>$

*IC Rule:*

$\mu^*(< M, N, Q>)$ in $c = e_1 \cdot e_2$, where $e_1$ is an utterance of an expression $E_1$ s.t. $\mu(E_1)(c) = M$ and $e_2$ is an utterance of an expression $E_2$ s.t. $\mu(E_2)(c) = N$ [62]

where…

$g = \lambda c \in O$. a salient location in $c$

By having recourse to the disposition characterized by the IC-rule we thereby have a balanced explanation of $LC^P$. We can imagine $S$ and $T$ arriving at their utterances by entertaining structured representations of their meanings (such as '$<M, N, Q>$') and then based on explicit meaning assignments arrive at utterances that can be concatenated in order to convey the original meaning.

    But note that $t_{10}$ like its predecessor $t_7$ is non-compositional in the sense of it not being compatible with $C^{FI}$. But this means that compositionality is not necessary to explain $LC^P$. So whether a theory can explain $LC^P$ seems to be independent of whether or not it is compositional.

## 3.3 Concluding Remarks

I have argued in this chapter that compositionality, as explicated along the lines of $C^{FI}$ is sufficient to causally explain why speakers are able to understand expressions. I have also argued that $C^{FI}$ is not required in order to do this. One can imagine several other principles that are such that the theories in accordance with them have the same explanatory potential as theories in accordance with $C^{FI}$ (even though

---

[62] $\mu^*$ is a function from utterance meaning (represented here as n-tuples) to utterances

these theories are not in accordance with $C^{FI}$). I've substantiated this in terms of the principles $F^{FI}$, $X^{FI}$, and $K^{FI}$. In addition, it turned out that a theory's being in accordance with $C^{FI}$ is neither sufficient nor necessary in order for that theory to explain why speakers are similarly disposed when it comes to the assignment of novel utterances to novel utterance meanings.

    This leaves us with no reason to think that a theory must conform to $C^{FI}$, specifically, in order to be correct. Considerations of explanations of linguistic creativity thus do not provide us with any justification for requiring that a semantic theory must be compositional. As it turns out, this will also be my conclusion after having considered productivity and systematicity in the next chapter.

# 4

# The Explanatory Utility of Compositionality; Productivity and Systematicity

The previous chapter revealed that a semantic theory could explain *linguistic creativity* without being in accord with *C*. This means that linguistic creativity alone gives us no reason to think *C* is true, and thus gives us no justification for requiring a semantic theory to accord with *C*. This chapter will focus on additional attempts to find reasons to think that *C* is true. It will be concerned with *Productivity* – the supposed fact that speakers understand *an infinite number* of expressions – and *Systematicity* – the supposed fact that speakers are such that the set of expressions that they understand *constitutes clusters of syntactically and semantically related expressions*.

  The chapter will be structured around the same questions as the previous one. It will discuss, for each phenomenon, i) whether that phenomenon *is likely to be real* (as opposed to merely apparent), and if so ii) whether semantic theories according with *C* can explain it, and iii) whether semantic theories are *required* to accord with *C* in

order to do so. The motivation for asking these questions is the same as in the last chapter. If theories according with *C* are able to explain why speakers have at least one of the relevant abilities, then, given that i) we have good reasons for thinking that speakers actually have this ability, and ii) it turns out that other, competing, determination claims are unable to explain why speakers have this ability, we shall have good reason to think that *C* is true (i.e. that correct semantic theories accord with *C*), and thus have a good reason to impose on semantic theories the requirement that they accord with *C*. However, if *C* is able to provide one of the relevant explanations, and there are good reasons to believe in the relevant explanandum, *but it turns out that other equally reasonable principles can also provide the explanation*, there will be no reason to think that *C*, specifically, is true. We will only have reason to think that *one out of a set of principles is true*. And we can then justifiably require at most that a semantic theory accords with one or the other of these principles. Similarly, if *C* is able to provide the relevant explanation, and it turns out that other competing determination claims are unable to explain why speakers have this ability but that *there are no good reasons to believe in the relevant explanandum*, the explanation will give us no reason to think that *C* is true and thus *no* justification for requiring semantic theories to accord with *C*.

In addressing these questions I will draw on the discussion presented in the previous chapter. In particular, I will assume that a speaker *S understanding* an expression *e* is a matter of *S* assigning a meaning to *e* which is similar to the meanings assigned to *e* by the members of some relevant speech community (a set of speakers); that *explaining* the relevant phenomena is a matter of giving *causal* explanations; and that what is really interesting from the perspective of compositionality is *understanding*, not *production*. (I will thus not be concerned with production in this chapter.)

I begin (in Section 4.1) by fleshing out a new explication of compositionality that will, at least *prima facie*, be needed if we are to provide some of the explanations of productivity and systematicity that will concern us in this chapter. Although it will ultimately be argued that we do not need it in order to explain the phenomena that actually obtain, this explication will serve as a useful contrast to the other explications of *C* that have already been introduced. This will be demonstrated in this chapter as well as in Chapter Five and

Chapter Six. In the next section (Section 4.2) I discuss productivity and in the section following that (Section 4.3) I examine three different forms of systematicity. I conclude, (in Section 4.4) that i) even though *C* (as explicated in the previous chapter) *can* be recruited to explain the forms of productivity and systematicity that there is reason to think actually obtain, ii) *C* is not *required* in order to explain productivity and systematicity. Hence, productivity and systematicity do not give us any reason to think that *C*, specifically, is true, and thus they give us no justification for requiring a semantic theory to accord with compositionality. The reasons to think that speakers really are productive and systematic are actually less compelling than the reasons to think that speakers are linguistically creative, so, if anything, productivity and systematicity give us less reason than linguistic creativity to think that semantic theories must be compositional.

## *4.1 An Even Stronger Compositionality Explication*

A version of the principle of compositionality is sometimes given that assumes that the meanings of complex expressions are themselves complex and are *built up* from the meanings of parts of those expressions.[1] An explication along these lines is importantly different from those we have considered so far. I introduce it here in order to discuss whether we need something like it to explain productivity and systematicity, and in order to contrast it with the previous explications of compositionality in Chapter Five (where we shall discuss potential problems with compositionality) and Chapter Six (where we shall examine the informative potential of compositionality).

---

[1] In addition to Frege's possible commitment to a principle such as this (which will be discussed below), Pagin (2003) and Fodor and Lepore ([2001] 2002) have also held that, at least at some level and at least in some sense, meanings are complex entities. Here is an excerpt illustrating Fodor and Lepore's view.

> [T]he meaning of "dogs bark" supervenes on the meanings of "dogs" and "bark" because the meanings of "dogs" and "barks" are *parts* of the meaning of "dogs bark"[…]complex meanings (don't just supervene on, but) actually *contain* the constituent meanings[…]
> (Fodor and Lepore [2001] 2002: 60 Emphasis added)

Frege makes the following remarks in the opening passage of his 'Compound Thought' and thereby suggests an explication of compositionality in terms of complex meanings.

> It is astonishing what language can do. With a few syllables it can express an incalculable number of thoughts, so that even a thought grasped by a terrestrial being for the very first time can be put into a form of words which will be understood by somebody to whom the thought is entirely new. This would be impossible, were we not able to distinguish *parts in the thought corresponding to the parts of a sentence, so that the structure of the sentence serves as an image of the structure of the thought.*
>
> (Frege [1923] 1977: 55, emphasis added)[2]

Generalizing away from Frege's particular syntactic and semantic notions, we obtain the idea that *the meaning of a complex expression is a complex whole which contains as parts the meanings of parts of the expression.*[3] This is an intriguing idea, but something which has to go through several refinements before it amounts to a precise explication of compositionality.

First, there is a trivial sense in which we can say that the meanings are complex, since they are meanings of complex expressions. What we are concerned with is an additional claim, not encompassed by the

---

[2] Many passages in Frege echo this one. E.g. 'We can regard a sentence as a mapping of a thought: corresponding to the whole-part relation of a thought and its parts, we have, by and large, the same relation for the sentence and its parts' (Frege [1919] 1979: 255); and '…thoughts have parts out of which they are built up. And these parts, these building blocks, correspond to groups of sounds, out of which the sentence expressing the thought is built up, so that the construction of the sentence out of parts of a sentence corresponds to the construction of thought out of the parts of a thought. And as we take a thought to be the sense of a sentence, so we may call a part of a thought the sense of that part of the sentence which corresponds to it' (Frege [1914] 1979, p. 225).

In spite of these passages it is not certain that Frege really took thoughts to be complex in the sense offered in the main text. Various disclaimers can be found in his work: e.g. the passage quoted in the main text continues 'To be sure, we really talk figuratively when we transfer the relation of whole and part to thoughts; yet the analogy is so ready to hand and so generally appropriate that we are hardly ever bothered by the hitches which occur from time to time'. Hence I shall refrain from actually attributing this view to Frege.

[3] According to Frege the thought corresponding to a sentence is the sense, or meaning, of that sentence.

previous explications of compositionality – something that can function as an additional constraint on semantic theories: that meanings have their complexity independently of them being meanings of the expressions they are in fact meanings of.[4]

Second, it seems reasonable to strengthen the intuitive idea somewhat. Meanings of complex expressions should be required not only to be complex (i.e. to have parts), but to have *structure*.[5] There are things – mereological sums for instance – that have parts but no structure, so the demand that something be complex *and* structured is stronger than the demand that the thing in question be complex. The stronger demand seems reasonable, though, since the meanings of complex expressions can differ even though they have parts with the same meanings: 'dog party' and 'party dog', for instance, have different meanings even though they have parts with the same meaning. The semantics of phrases like these could be accounted for in terms of the thesis that although their meanings are complex entities with the same parts, they are complex entities that are structured from these parts in different ways.

Third, turning to the question of how *C* can be explicated in terms of complex meanings, we can note that if we simply replace 'is determined by' in *C* by 'is built up from' we end up with the counter-intuitive principle that the meanings of complex expressions are built up from the meanings of their parts *and their modes of composition*. Since the occurrence of 'mode of composition' in *C* refers to a way in which grammatical terms are combined, it seems nonsensical to hold that modes of composition are parts of the meanings. Much more plausible is the idea that the meanings of complex expressions are built up from the meanings of the parts of these expressions *in accordance* with their modes of combination, where 'in accordance with' is interpreted in the following weak sense: if two complex expressions have the same mode of combination, their meanings are built up in the same way (i.e. have the same structure); and if two complex expressions have different modes of composition, their

---

[4] Pagin (2003a) explicates Frege's suggestion in terms of the requirement that the semantic algebra is free. The form of structure that flows from this requirement is extrinsic to the meanings and depends on the system they are embedded in. I have something more intrinsic in mind.

[5] Note that Frege seems to assume that thoughts are structured as well: '…so that the structure of the sentence serves as an image of *the structure of the thought*.'

meanings *might* be built up in the same way (i.e. might have the same structure) or might be built up in different ways.

Fourth, we want to avoid building into our explication of compositionality the requirement that expressions and meanings *have exactly the same complexity*.[6] So far we have said nothing that is incompatible with simple expressions having *complex* meanings; we have said only that complex expressions cannot have simple meanings. It seems important to resist building this requirement into an explication of $C$, since its adoption rules out the notion that expressions of different syntactical complexity can be synonymous. This is strikingly implausible, in view of expression pairs such as 'brother' and 'male sibling' or 'doe' and 'female deer'.[7]

Fifth, in order to capture the intuition behind $C$ that the meaning of a complex expression *is determined by* the meanings of that expression and its mode of composition (and thus obtain something strictly stronger than $C^{FE}$) we need to build in some general requirement governing *which* of the meanings of the parts of a complex expression that must be parts of the meaning of that expression. $t_1$ is an illustration of a theory in which the meaning of a complex expression contains *some* of the meanings of its parts, as parts, but not all of them.

| $t_1$ | Syntax: | | Semantics: | |
|---|---|---|---|---|
| | N | $\rightarrow$ 'dogs', | $\mu$('dogs') | $= m_1$ |
| | | 'canines' | $\mu$('canines') | $= m_1$ |
| | A | $\rightarrow$ 'brown' | $\mu$('brown') | $= m_2$ |
| | NP | $\rightarrow$ A·N | $\mu$('brown dogs') | $= <m_1, m_2>$ |
| | | | $\mu$('brown canines') | $= <m_1>$[8] |

---

[6] Notice that Frege seems to suggest that something like this is true for thoughts: '…so that the structure of the sentence *serves as an image* of the structure of the thought.' This mirror requirement is something that has occasionally been required of artificial languages. Cf. Russell ([1918] 1956): '[I]n a logical correct symbolism there will always be a certain fundamental identity of structure between a fact and the symbol for it…'

[7] Although this point is important it is sometimes neglected. Szabo (2000: 488-489) argues that what he calls 'the building principle' – that the meaning of a complex expression is built up from the meanings of its constituents – entails that 'sameness of meaning entails sameness of structure'. But this does not follow in the general case.

[8] I am here assuming that the elements of an ordered *n*-tuple are parts of that *n*-tuple.

This theory does not accord with $C^{FE}$, since it contains two complex expressions with parts with the same meaning, the same mode of composition, but different meanings. However, all meanings of complex expressions contain meanings of the parts of these expressions as parts. The natural way to avoid this kind of divergence from $C^{FE}$ is to strengthen the new explication of $C$ so that the meaning of a complex expression is required to have as parts *all* and *only* the meanings of its parts.

   Sixth, to the extent that this is possible, we would like our new explication of $C$ to be neutral with respect to the amount of meanings terms have. Both $C^{FE}$ and $C^{FI}$ were *thin* versions of compositionality and assumed nothing about the number of meanings a grammatical term has. The claim that the meaning of a complex expression is built up from the meanings of its parts entails that there can be no complex expressions without meanings, but it seems perfectly compatible with the possibility that an expression has several meanings and the possibility that some of the parts of a complex expression have no meanings (as long as these parts are not themselves complex). The following two conditions encompass these considerations.

> For each complex grammatical term $e=\sigma(e_0,\ldots,e_{n-1})$ and each meaning $m \in \mu'(e)$ it is the case that $m$ is a structure built up from the parts $m_0,\ldots,m_{n-1}$, where $m_0 \in \mu'(e_0),\ldots,m_{n-1} \in \mu'(e_{n-1})$, and for no two meanings $m$, $m' \in \mu'(e)$ it is the case that $m$ and $m'$ is built up from the same parts.

> For each syntactic operation $\sigma$, and for each sequence of expressions $e_0,\ldots,\ e_{n-1}$ for which $\sigma$ is defined, given that $\sigma(e_0,\ldots,e_{n-1})=e$ then for each $n$-tuple $m_0 \in \mu'(e_0),\ldots,m_{n-1} \in \mu'(e_{n-1})$ it is the case that there is a meaning $m \in \mu'(e)$ which is a structure built up from the parts $m_0,\ldots,m_{n-1}$.

If our explication makes use of conditions like these, it will be fairly thin.[9]

   Seventh, since we noted in the previous chapter that $C^{FI}$ had a certain explanatory utility (i.e. theories in accordance with it could

---

[9] That being thin is an attractive feature will be shown in the next chapter.

explain linguistic creativity), we should build $C^{FI}$ into our new explication of $C$ to ensure that it has this explanatory utility as well.

The explication of compositionality that follows from these seven considerations – what I shall dub $C^{BU}$ – amounts to the following. [10]

A theory $t$ accords with $C^{BU}$ iff

i) There are no explicit semantic rules for complex expressions in $t$.

ii) Each implicit semantic rule in $t$ is of the form '$\mu(x)=y$' where

1) '$x$' is a mode of composition variable, and
2) '$y$' is a *c-term*, where a *c-term* is defined in the following way:

a) '$\mu(x_1)$' is a *c*-term if '$x_1$' is one of the variables from which '$x$' is built up.
b) '$f(x_1,\ldots,x_n)$' is a *c*-term if '$x_1$',…,'$x_n$' are *c*-terms and '$f$' is a functor corresponding to a humanly computable function defined on the set of meanings.

iii) For each complex grammatical term $e=\sigma(e_0,\ldots,e_{n-1})$ and each meaning $m\in\mu'(e)$ it is the case that $m$ is a structure built up from the parts $m_0,\ldots,m_{n-1}$, where $m_0\in\mu'(e_0),\ldots,m_{n-1}\in\mu'(e_{n-1})$, and for no two meanings $m$, $m'\in\mu'(e)$ it is the case that $m$ and $m'$ is built up from the same parts.

iv) For each syntactic operation $\sigma$, and for each sequence of expressions $e_0,\ldots,e_{n-1}$ for which $\sigma$ is defined, given that $\sigma(e_0,\ldots,e_{n-1})=e$ then for each $n$-tuple $m_0\in\mu'(e_0),\ldots,m_{n-1}\in\mu'(e_{n-1})$ it is the case that there is a meaning $m\in\mu'(e)$ which is a structure built up from the parts $m_0,\ldots,m_{n-1}$.

---

[10] 'BU' in '$C^{BU}$' is short for 'built up'.

With this new explication in place we can turn to the first of the two phenomena that will concern us in this chapter – productivity.

## *4.2 Productivity*

In the previous chapter we scrutinized a commonly cited reason for believing that *C* is true – the fact that speakers understand novel expressions. It is also common to cite *productivity* as a reason for believing that C is true. This supposed phenomenon can be captured, with reference to an arbitrary speaker *S*, in the following way.[11]

(P[U])     *S* understands an *infinite* number of expressions.[12]

---

[11] It is sometimes speakers and sometimes languages that are held to be productive (see, Larson and Segal (1995: 12) for an example of the first kind of formulation, and Fodor and Lepore (2002: 1) for an example of the second). I will be concerned only with explanations of phenomena described by the first formulation. The meaning assignments of speakers seem more basic, and thus more important, than the meaning assignments of languages. In addition, description of the properties of speaker's languages presupposes an idealization that will obscure the riches that can determine the meaning of a complex expression. The two formulations both describe non-evident phenomena in need of additional support.

   Davidson's learnability argument is very similar to a productivity argument set out in terms of the second of these two productivity formulations.

> When we can regard the meaning of each sentence as a function of a finite number of features of the sentence, we have an insight not only into what there is to be learned; we also understand how an infinite aptitude can be encompassed by finite accomplishments. For suppose that a language lacks this feature; then no matter how many sentences a would-be speaker learns to produce and understand, there will remain others whose meanings are not given by the rules already mastered. It is natural to say such a *language* is unlearnable.
> (Davidson 1965: 8 Emphasis added)

With 'learnability' we do not have the previous ambiguity, since it is nonsensical to hold that speakers are learnable or unlearnable; only languages are learnable or unlearnable.

[12] As should be evident from *P[U]*, productivity is not limited to speech *production*.

   Productivity is sometimes captured by the claim that 'Speakers are *able* to understand an infinite number of expressions'. Note the ambiguity in the way in which 'understanding an expression' is understood. In the version of productivity given in the main text, 'understanding an expression' pertains to assigning a meaning to an expression, i.e. being disposed to assign certain meanings to utterances of that expression. But in the version given a few sentences ago, 'understanding an expression' pertains to the assignment of meanings to utterances. If we add 'is able to'

A capacity to understand *novel* expressions does not automatically give the speaker the ability to understand an *infinite* number of expressions. The number of novel expressions she is able to understand might very well be finite. But the conviction is common, in the literature, that speakers *as a matter fact* understand an infinite number of expressions; and if they do, then, obviously, since they are familiar only with a finite number of expressions, they will understand novel expressions.

We can plug the explication of 'understanding an expression' of the previous chapter into $P^U$ in order to get an explanandum in terms of meanings, i.e. something more readily connected to the principle of compositionality.

($P^U$)   *S* and *T* assign similar meanings to an infinite number of expressions.[13]

Being able to assign meanings (at all) to an infinite number of expressions calls for an explanation in it self. But it seems that this explication captures the situation better than the one couched exclusively in terms of unconstrained meaning assignments. If we are able to assign meanings to an infinite number of expressions, what demands explanation is not just that we are able to assign them all meanings, but also that we are able to assign them similar meanings. Given this explication we can turn to the question i) whether speakers are productive and, if this is the case, the further question ii) what is required of a semantic theory if it is to explain why they are productive.

*Are Speakers Productive?*

It is less evident that speakers are productive than that they are linguistically creative. This difference can be put in the following way: whereas $LC^U$ can (in the absence of conflicting evidence) be safely inferred from i), which is beyond doubt, $P^U$ can correspondingly be

---

to the first reading we run the risk of getting too many layers of dispositions, i.e. of a speaker's being disposed *to be disposed* to understand utterances.
[13] '$P^U$' is short for 'Productivity – Understanding'.

inferred from ii); but the latter is of no real interest since ii) is obviously false.

i)      $S$ has understood novel utterances.
($LC^U$) $S$ understands novel expressions.

ii)     $S$ has understood an infinite number of utterances.
($P^U$)   $S$ understands an infinite number of expressions.

Whereas $LC^U$ follows readily from an assumption concerning what a speaker has in fact done that is beyond doubt, this is not the case with $P^U$, so we have to find a different justification for believing in $P^U$. I am familiar with three lines of reasoning that might lend credence to something like $P^U$. The third of these seems better than the others, but none is completely satisfactory.[14]

First, if one is already confident that speakers are correctly characterized by compositional semantic theories (embodying recursive syntactical rules), it will follow that speakers are correctly described by $P^U$ to the extent that they converge on what meanings they assign to simple expressions, and what modes of composition they assign to complex expressions, and that they are correctly described by the same implicit semantic rules. So an antecedent commitment to compositionality could support $P^U$.

However, this form of support leads to unwanted circularity in the present context. What is needed here is support for productivity that is independent of a belief in compositionality, since we are looking to productivity to support compositionality.

Second, by modifying the statement of productivity slightly one could obtain something very plausible. One could, for instance, understand productivity, not in terms of 'an infinite number of expressions', but in terms of 'a hugely large number of expressions'.[15] This form of productivity – call it $P^{U}*$ – is better supported than the one with which we began. Consider the following two assumptions.

---

[14] I leave to one side here arguments that invoke the novelty of the relevant expressions and essentially collapse $P^U$ it into $LC^U$.
[15] See, e.g., Grandy (1990) and Pelletier (2004:142 fn. 7)

iii)   $S$ has understood $n$ utterances (a hugely large number of utterances).

($P^{U}*$)   $S$ understands $n$ expressions (a hugely large number of expressions).

No matter what number is chosen for $n$, $P^{U}*$ can safely be inferred from iii). In addition, for many large values of $n$, iii) is beyond doubt. So $P^{U}*$ (for at least some $n$) is well supported.

However, $P^{U}*$ provides significantly poorer support for compositionality than does $P^{U}$. The argument from $P^{U}$ to $C$ is something like the following.

($P^{U}$)   $S$ and $T$ assign similar meanings to an *infinite* number of expressions.

($P_2$)   Speakers can only assign a finite number of meanings to expressions explicitly (since minds are finite)

(C)   Correct theories of speakers' semantic competence must embody something like compositionality where implicit semantic rules are employed.

Now even if it is the case, as I shall argue in the next sub-section, that $P^{U}$ can be explained, and that $P_2$ can be respected, even under the assumption that semantic theory is not compositional, the problem with this argument is not $P_2$. That minds are finite seems to be a fairly weak, reasonable assumption to make. Contrast the previous argument with the following one, set with $P^{U}*$.

($P^{U}*$)   $S$ and $T$ assign similar meanings to $n$ expressions (a hugely large number of expressions).

($P_2*$)   Speakers can only assign $m$ (where $m<n$) meanings to expressions explicitly.

(C)   Correct theories of speakers' semantic competence must embody something like compositionality where implicit semantic rules are employed.

This argument, whatever its other faults, embodies a fairly strong empirical assumption. Whereas it is clear that minds can only contain a finite number of explicit meaning assignments (i.e. a finite number of distinct mental representations), it is significantly less clear that

there is a number $m$ (lower than $n$) such that speakers can only assign $m$ meanings explicitly. While it seems safe to assume $P_2$ in the absence of empirical inquiry, the same probably cannot be said of $P_2{}^*$. So in the absence of such results, it seems that $P^{U}*$, though more plausible than $P^{U}*$, does not provide better support for $C$ than does $P^{U}*$. There is a trade-off here. What is wrong with the first argument is that the first premise is not well supported. But when this premise is weakened, the second premise needs to be strengthened, and this is the problem with the second argument.

   Third, perhaps the fact that we can understand an infinite number of expressions can be established inductively, in the following way. We note first that we understand each of the expressions in a sequence like the following, i.e. that we converge on what they mean.

(1)    We destroyed the anti-missiles.
(2)    We destroyed the anti-anti-missiles.
(3)    We destroyed the anti-anti-anti-missiles.
(4)    We destroyed the anti-anti-anti-anti-missiles.

We note secondly that (2) results from (1) by prefixing 'anti' to 'missiles', and that we understand (2). We see that (3) results from (2), and (4) results from (3) in the same way as (2) results from (1), and that we understand all the resulting expressions. We thus have inductive evidence for thinking that the prefixing of 'anti' to 'missiles' will *always* result in an understandable expression. But if this is true, we understand an infinite number of expressions. $P^{U}$ is thus supported.

   The problem with this inductive generalization is that it might not be without exceptions. We have trouble understanding (5), for instance, even though it results from repeated application of the aforementioned procedure.

(5)    We destroyed the anti-anti-anti-anti-anti-anti-anti-
       anti-anti-anti-anti-anti-anti-anti-anti-anti-anti-
       anti-anti-anti-anti-anti-anti-anti-anti-anti-missiles.

We certainly have a somewhat vague idea of what (5) is about, but it is not certain that we would actually converge in our meaning assignments to it: we might disagree over whether the description of a

certain situation would entail (5); similarly we might disagree over whether (5) is true in a certain situation we have encountered.

Fodor and Pylyshyn give the following argument in response to inductive arguments like the aforementioned one breaking down.

> [T]here are a number of considerations which suggests that, despite de facto constraints on performance, one's knowledge of ones language supports an unbounded productive capacity in much the same way that one's knowledge of addition supports an unbounded number of sums. Among these considerations are, for example, the fact that a speaker/hearer's 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 speaker's memory and attention rather that what he knows about – or how he represents – his language.
>
> (Fodor and Pylyshyn 1988: 34)

Their response is compelling; there is a feeling that we know how to go on with a case such as (5), and thus that we would under more ideal circumstances converge on its meaning. If this is correct, (5) might not really be a counterexample to the induction under discussion, and $P^U$ might retain its inductive support in spite of cases like (5). Although I think that this kind of argument offers the most compelling support for $P^U$ (in the context of supporting C), it seems somewhat arbitrary to say whether one would want to maintain that $P^U$ is true after being exposed to it or whether one would want to deny this. It seems equally reasonable either to say that $P^U$ is true if we allow ourselves more ideal conditions under which understanding can take place, or to refuse to idealize in this way because the data supporting understanding in normal circumstances is not available for the problem cases and to conclude that $P^U$ is not supported. Since I will ultimately argue that all the phenomena standardly cited to support belief in $C$ (that actually obtains) can be explained by non-compositional theories, I will charitably assume that $P^U$ is true, and thus demands explanation, in order to show that it does not constitute an exception to my general thesis.

*What is required to Explain Productivity?*

What do we need to assume about $S$ and $T$ in order to explain, causally, why they assign similar meanings to an infinite number of expressions, i.e. why they are similarly disposed in their assignments of meaning to utterances of an infinite number of expressions? Given that minds are finite, $S$ and $T$ could not have an infinite number of causally real, distinct, dispositions corresponding to each of their meaning assignments. They could have causally real, distinct, dispositions corresponding to their meaning assignments involving simple expressions, since they assign meanings only to a finite number of simple expressions. But this is not possible for the infinitely many meaning assignments involving complex expressions. We have to assume instead that among $S$'s and $T$'s causally real mental states giving rise to meaning assignments involving complex expressions, some are *general* in the sense that they each give rise to several of the relevant meaning assignments. Given that *the number of meanings* assigned to complex expressions by $S$ and $T$ is also infinite, these states cannot all be *constant* in the sense that they are dispositions to assign *the same meaning* to each of the expressions they give rise to.[16] In other words, since, for instance, $\mu$('anti-missile') $\neq$ $\mu$('anti-anti-missile') $\neq$ $\mu$('anti-anti-anti-missile') and so on, at least some of the causally real states of $S$ and $T$ that give rise to their meaning assignments involving complex expressions cannot be constant dispositions. Instead, in many cases we have to assume that $S$ and $T$ have causally real *variable* dispositions that give rise to the meaning assignments involving complex expressions. These dispositions are variable in the sense that they are dispositions to assign different meanings to the expressions they are responsible for. We thus arrive at the question of what the output of these dispositions varies with, i.e. the question what their outputs depend on. We know that the meaning $S$ and $T$ assign to a complex expression *depends* on the meaning they assign to its parts, since the meaning they assign to a complex expression often changes when the meaning they assign to one of its parts changes – 'anti-missile' would mean something different to $S$ and $T$ if 'missile' meant the same thing as 'bear' to

---

[16] That both the number of expressions and the number of meanings are supposed to be infinite is emphasized by, e.g., Fodor and Lepore (2002) and Fodor and Pylyshyn (1988).

them. We also know that the meaning $S$ and $T$ assign to a complex expression *depends* on the mode of composition they assign to it, since the meaning they assign to a complex expression often changes when the mode of composition they assign to it is changed: 'happy men and women' means one thing to $S$ and $T$ if assigned a mode of composition along the lines of [[happy [men and women]] and another if assigned a mode of composition along the lines of [[happy men] and [women]]. So we know that the variable dispositions that give rise to the meaning assignments involving complex expressions that we need to attribute to $S$ and $T$ must be such that the meanings they give rise to depends on what modes of composition are assigned to the complex expressions and what meanings are assigned to their parts.

At this point is seems that we can devise an adequate explanation of why $S$ and $T$ assign similar meanings to an infinite number of expressions in the following way: attribute to $S$ and $T$ a finite number of causally real dispositions giving rise to i) their meaning assignments involving simple expressions and ii) their mode of composition assignments; and attribute to them a finite number of causally real general variable dispositions to assign meanings to complex expressions based on the meanings assigned to their parts and the modes of composition assigned to them. That $S$ and $T$ coincide in their possession of these dispositions is plausibly explained by the fact that they have been exposed to similar linguistic transactions. If the details of the attributions are correct, those attributions will explain why $S$ and $T$ assign similar meanings to an infinite number of expressions. They will do so in virtue of giving details that are common to the causal histories of $S$'s and $T$'s meaning assignments, and because they are not incompatible with what we know of speakers.

But this is just one possible explanation. So far, we have no reason for thinking that the causally real general variable dispositions to assign meanings to complex expressions that are attributed to $S$ and $T$ are not such that the meaning assignments they give rise to *also* depend on other states of $S$ and $T$ (and not only the meaning assignments to parts and mode of composition assignments). To be more specific, we have no reason to think that other states of $S$ and $T$ also influence the meanings that they assign to complex expression as long as this dependence does not entail that the meanings $S$ and $T$

assign to complex expressions are insufficiently similar. Among the possible extra determinants we find all the factors discussed in the previous chapter: $S$'s and $T$'s assignment of particular forms to complex expressions (or their parts), ectal dispositions to respond to context in various ways, and $S$'s and $T$'s beliefs about the world. Nothing about these factors prevents an account like the one we are considering from involving them.

These reflections translate into talk about semantic theories in the following way. A theory in accordance with, for instance, $C^{FI}$, $F^{FI}$, $X^{FI}$, or $K^{FI}$, can explain $P^U$ granted the following assumptions: i) the explicit semantic rules of the theory are interpreted as characterizing causally real states of $S$ that give rise to some of her meaning assignments; ii) the syntax of the theory, which must be recursive to cover an infinite number of expressions, is taken to characterize those of $S$'s states that give rise to her mode of composition assignments; iii) the implicit semantic rules of the theory characterize $S$'s general variable dispositions to assign meanings to complex expressions on the basis of certain other factors; and iv) we know that $S$ and $T$ are similar in these respects, so that the theory provides information not only about $S$ but also about $T$. With these assumptions, something like the following compositional (in the sense of $C^{FI}$) theory might explain $P^U$.

$t_2$    *Syntax:*
        S → 'Phil ponders',
            'Chris agrees',
            'Jill knows Kate'
        S → S·'and'·S
        S → S·'or'·S
        S → 'It's not the case that'·S

       *Semantics:*

$$\mu(\text{'Phil ponders'}) = \textit{True}$$
$$\mu(\text{'Chris agrees'}) = \textit{True}$$
$$\mu(\text{'Jill knows Kate'}) = \textit{False}$$
$$\mu(_S\, S_1 \text{ 'and' } S_2) = \wedge(\mu(S_1), \mu(S_2))$$
$$\mu(_S\, S_1 \text{ 'or' } S_2) = \vee(\mu(S_1), \mu(S_2))$$
$$\mu(_S\, \text{'It's not the case that' } S) = \neg\mu(S)$$

This theory is accords with $C^{FI}$, since i) no complex expressions are assigned meanings explicitly (i.e. directly in terms of those expressions), and ii) the implicit semantic rule are of the form '$\mu(x)=y$' where '$x$' is a mode of composition variable, and '$y$' is a *c*-term.

But given the previous assumptions something like the following non-compositional $F^{FI}$ theory might also explain $P^{U}$.[17]

$t_3$     *Syntax:*                              *Semantics:*

N → 'dogs',                $\mu$('dogs')           = {x: x is a dog}
      'canines'              $\mu$('canines')        = {x : x is a dog}
$P_1$ → 'is an expression' $\mu$('is an expr…')   = {x: x is an expr…}
$P_2$ → 'are nice'          $\mu$('are nice')       = {x : x is nice}
E → N, $P_1$, $P_2$, Q, S   $\mu(_E X)$              = $\mu(X)$[18]
Q → '''·E·'''               $\mu(_Q$ '' E '')        = $g(E)$[19]
S → Q·$P_1$                 $\mu(_S Q P_1)$          = $f(\mu(Q), \mu(P_1))$
S → N·$P_2$                 $\mu(_S N P_2)$          = $f(\mu(N), \mu(P_2))$


      where…
      $g(x)$    = {y : y=x}
      $f(x, y)$  = 1 if $x \subseteq y$ else 0

This theory fails to accord with compositionality (on any explication). Yet it seems to be in the same position to explain $P^{U}$ as the previous theory. It gives rise to an infinite number of meaning assignments, and it is equally compatible with what we know about speakers. So, to sum up, it seems that compositional (in the sense of $C^{FI}$) theories *can* explain why speakers are productive, but that theories are *not required* to be compositional in order to do this.

The situation is similar with the $X^{FI}$ and $K^{FI}$ theories. The simplest way to illustrate this is to take a non-compositional theory that accords with $X^{FI}$ or $K^{FI}$ (e.g. one of the theories given in Chapter Three) and combine it with $t_3$. The resulting theory will not be compositional (since it has a part which is non-compositional), but it

---

[17] Cfr. Werning (2005).
[18] The semantic rule corresponding to the syntactic rule 'E → N, $P_1$, $P_2$, Q, S' is just an identity function on meanings regardless of what syntactic category that is involved. ''
[19] The quotation marks '' and ''' are enclosed in quotation marks in order not to 'trap' the concatenation functors and the syntactic variable.

will assign meanings to an infinite number of expressions (since it has another part – $t_3$ – which assigns meanings to an infinite number of expressions). On the other hand, it will accord with $X^{FI}$ or $K^{FI}$ (depending on what kind of theory that was combined with $t_3$). This illustrates an important point. $X^{FI}$ theories and $K^{FI}$ theories need contain neither an infinite number of ectal dispositions nor an infinite number of beliefs (which would be unreasonable) in order to account for $P^{U}$. I therefore conclude that not only $C^{FI}$ and $F^{FI}$ theories, but also $X^{FI}$ and $K^{FI}$ theories, can account for $P^{U}$.[20]

Before moving on we should consider an objection to the conclusion that $C^{FI}$ can explain $P^{U}$. It might be argued that a theory according with, for instance, $C^{FI}$ does not in general explain why speakers are productive in the strong sense in which they really are productive, since the speaker such a theory describes will (if it encompasses a recursive syntax) only attribute the speaker with the ability to assign meanings to an infinite number of expressions, not necessarily the ability to *assign an infinite number of meanings to expressions*. But, the argument goes, speakers really are productive in the sense that the set of all their meaning assignments encompasses an infinite number of meanings. This purported shortcoming of some of the theories according with $C^{FI}$ is illustrated by $t_2$, which is compatible with $C^{FI}$ but only encompasses two different meanings (the true and the false).

One might conclude from this argument that compositionality is best explicated not by $C^{FI}$ but by $C^{BU}$, since the latter avoids the previous problem. For a theory according with $C^{BU}$ that embodies a recursive syntax always embodies an infinite number of meanings. The following theory illustrates this.

$t_4$    *Syntax:*          *Semantics:*

     A    → 'anti'      $\mu$('anti')     = the property of being anti

     NP   → 'missile'    $\mu$('missile') = the property of being a missile

     NP   → A·NP      $\mu(_{NP} \text{ A NP}) = f(\mu(\text{A}), \mu(\text{NP}))$

     where…

---

[20] Note that the arguments considered in Chapter Three against trying to appoint one kind of explanation as superior to the others apply in this context as well. These arguments also apply in the context of giving explanations of systematicity.

$$f(x, y) = <x, y>$$

According to theories, like this one, which accord with $C^{BU}$ the meaning of each complex expression is a structure built up from the meanings of the parts of that expression, and this entails that the meaning of a complex expression will always have a meaning distinct from the meanings of its parts. Hence, it follows, where a semantic theory embodies a recursive syntax (thereby ensuring an infinite number of complex expressions) and accords with $C^{BU}$, that that theory embodies an infinite number of meanings.

   However, the previous argument, favoring the explication of compositionality in terms of $C^{BU}$, is flawed. Even though theories that accord with $C^{FI}$ will not in general encompass an infinite number of meanings, the *correct* such theories (if there are any) that cover cases like the 'anti-missile' combinations will encompass an infinite number of meanings if the speaker the theory is about actually assigns an infinite number of meanings to these expressions. It is only the subclass of theories according with $C^{FI}$ that contains the correct theories about which we make any explanatory claims. Obviously, if a theory that accords with $C^{FI}$ encompassed only erroneous meaning assignments, this theory would not provide an adequate explanation of anything. Since we are concerned only with correct theories, we do not need a compositionality criterion to ensure that the number of meanings is right: this will sort itself out when the theory encompasses appropriate meaning assignments for 'anti-missile', 'anti-anti-missile', 'anti-anti-anti-missile', and so on. In fact if we adopt $C^{BU}$ we get more meanings than we want, for the result, following from $C^{BU}$, that a complex expression cannot mean the same as one of its parts, is quite unintuitive. For example, this result runs counter to synonymy judgments such as 'The sun is shining or the sun is shining' means the same as 'the sun is shining'. It also makes pleonastic expressions impossible (e.g. 'the female doe' will not mean the same thing as 'the doe'). So explicating compositionality in terms of $C^{BU}$ might not be a good idea. $C^{FI}$ is sufficient if the aim is to explain productivity.

## 4.3 Systematicity

The question whether speakers are *systematic* in the sense discussed in the literature on compositionality, and if so what it would take to explain this, is less straightforward than the corresponding questions for productivity.[21] For although Jerry Fodor, who first drew attention to this way of supporting compositionality, originally presented 'systematicity' as a name for a single phenomenon, it has since transformed into a general term covering several different phenomena.[22] We thus need to address several distinct phenomena and the following subsections will each address three of these in turn.[23]

---

[21] Systematicity, like productivity, is sometimes assumed to be a property of speakers, sometimes a property of languages. Again, see Larson and Segal (1995: 11-12) for an example of the first kind of formulation, and Fodor and Lepore (2002: 2) for an example of the second. As has already been indicated in the main text, I will frame the discussion of the different systematicity phenomena as one concerning certain properties of speakers.

[22] Fodor (1987) was first to draw attention to systematicity. Fodor and Pylyshyn (1988) distinguished between *the systematicity of cognitive representations* – which is probably the phenomenon that Fodor (1987) had in mind – and *the co-occurrence of systematicity of cognitive representations and semantic relatedness, the parallelism between syntactical structure of sentences and their entailments* and *the systematicity of inference*. Following Fodor and Pylyshyn's subdivision of the systematicity phenomena, Kenneth Aizawa (see Aizawa (1997a; 1997c; 1997b; 2003b; 2003a)) added *diachronic systematicity* and *the co-occurrence of systematicity of cognitive representations and diachronic systematicity* to their list of phenomena. In addition to these explicitly marked distinctions, different authors often seem to mean different things by 'systematicity', and this gives rise to additional distinctions. See Chalmers (1990a; 1990b), Cummins (1996), Fodor and McLaughlin (1990), Hadley (1997a; 1997b), Johnson (2004), McLaughlin (1993a; 1993b), Niklasson and van Gelder (1994), and van Gelder and Niklasson (1994) for some of the views on systematicity. In addition to all these differences, there are, in the literature on compositionality, several phenomena which are clearly related to systematicity, but are discussed under alternative names. E.g. the discussion of systematicity is anticipated in Gareth Evans' (1982) discussion of *the generality constraint* (something bearing more than a passing resemblance to *the systematicity of cognitive representations*); and what we might refer to as *the closure of understanding under subterms* – a phenomenon that has been assumed to support *reverse compositionality* (e.g. see Fodor and Lepore ([2001] 2002) and Fodor (1998b; 1998c)) – also closely resembles systematicity.

[23] I will be concerned with the systematicity of language, not the systematicity of thought. Fodor (1987) and Fodor and Pylyshyn maintained that both thought and language were systematic, and they discussed the systematicity of language merely as a precursor to their discussion of the systematicity of thought.

### 4.3.1 The Closure of Understanding under Syntactic Operations

Fodor originally drew attention to systematicity in the following way.

> The property of linguistic capacities that I have in mind is one that inheres in the ability to understand and produce sentences. That ability is –as I shall say– systematic: by which I mean that the ability to produce/understand some of the sentences is intrinsically connected to the ability to produce/understand many others[…]You don't, for example, find native speakers who know how to say in English that John loves Mary but don't know how to say in English that Mary loves John. If you did find someone in such a fix, you'd take that as presumptive evidence that he's not a native English speaker but some sort of tourist.
>
> (Fodor 1987: 149)

From this passage it is clear that systematic understanding is a matter of the speaker understanding certain sentences (what we can call the *projection set*) when she understands certain other sentences (the *base set*). What is not clear from the passage, is exactly which sentences a speaker is supposed to be able to understand given that she is able to understand certain other sentences, i.e. what the general connection between the base set and the projection set is supposed to be. Fodor does not elaborate this, but the range of examples he makes use of in the article (and others) can be used to surmise that the phenomenon is supposed to amount to something like the following.[24]

---

[24] This is based on the following examples that have been used to illustrate systematicity.

|      | Base Set                               |       | Projection Set                        |
|------|----------------------------------------|-------|---------------------------------------|
|      | *Base Set*                             |       | *Projection Set*                      |
| (6)  | 'John loves Mary'                      | →     | 'Mary loves John'                     |
| (7)  | 'John loves the girl'                  | →     | 'The girl loves john'                 |
| (8)  | 'P → Q'                                | →     | 'Q → P'                               |
| (9)  | '¬(P&Q)'                               | →     | '¬P'                                  |
| (10) | 'John snores', 'Flounders swim'        | →     | 'Flounders snore', 'John swims'       |
| (11) | 'John loves Mary                       | →⊢    | 'Peter loves Mary'                    |

Given only (6), (7), and (8) one might have hypothesized that the phenomenon is supposed to be that speakers who understand some sentences also understand the sentences that result from syntactically permissible permutations of the constituents of the base sentences. But (9) suggests something stronger. The projection set is not made up wholly of the results of syntactically permissible permutations of the

($S^{CO}$)   If $S$ understands the sentences of a set of sentences $\varphi$, then $S$ understands all sentences that have the same mode of combination as at least one of the sentences in $\varphi$ and that are built up only from expressions that are parts of the sentences in $\varphi$. [25]

---

sentences in the base set. It also contains sentences of the same mode of combination as sentences in the base set but with fewer constituents. From (10) we obtain the result that the base set and the projection set can contain several sentences, and from (11) we obtain the result that no constituents not present in the base set can be contained in the sentences in the projection set. These considerations suggest that systematicity is supposed to amount to something like $S^{CO}$. The examples are collected from several different texts written by Fodor. (6) is from Fodor (1987: 149), (7) is from Fodor and Pylyshyn (1988: 37), (8), (9) and (11) are from Fodor (1998a: 97) and (10) is from Fodor and Lepore ([2001] 2002: 58).

[25] 'CO' in '$S^{CO}$' is short for 'Closure under syntactic Operations'.

It is unclear whether this is really what Fodor and Pylyshyn are after. Some passages suggest that the phenomenon they discuss in the relevant section is purely syntactical – something along the lines of the idea that if speakers find certain sentences grammatical they will also find certain others grammatical. This would make the explanation they offer for this kind of systematicity more sensible: 'If you assume that sentences are constructed out of words and phrases, and that many different sequences of words can be phrases of the same type, the very fact that one formula is a sentence of the language will often imply that other formulas must be too: in effect, systematicity follows from the postulation of constituent structure' (Fodor and Pylyshyn 1988: 38). But compositionality has no place in this kind of explanation, since the phenomenon is purely syntactic. Hence, the existence of this kind of systematicity does not support compositionality. In addition, their discussion of systematicity makes use of 'understanding' and 'knows how to say that', which are semantic notions. This also suggests that they do not have a purely syntactic phenomenon in mind. In addition they claim, but do not substantiate their belief, that semantic structure is also required for explaining this kind of systematicity. This is hard to make sense of on the purely syntactic interpretation of systematicity. My interpretation makes sense of these statements and harmonizes with the illustrations on offer. In later works Fodor does offer a more precise general formulation of systematicity which is in conflict with the one I have proposed: 'Systematicity is the property that a system of representations has (whether or not it is productive) if each of the symbols it contains occurs with the same semantic value as a constituent of many different hosts' (Fodor 2001: 6). But this seems odd, since the property described here is the context-independence of meaning, the very same property that Fodor and Pylyshyn use to *explain* certain forms of systematicity. (These will be discussed in the next two sections). So explanandum and explanans are run together. The form of systematicity that Larson and Segal (1995: 11-12) commit to seems to very similar to ($SY^{CO}$).

Put differently, it seems that this form of systematicity amounts to understanding being closed under syntactic operations.

## Is Understanding Closed under Syntactic Operations?

Empirical research is seldom cited to support systematicity claims. When it was first introduced, systematicity was put forward as a more or less self evident property and characterized only anecdotally. However, it is not obvious that it actually obtains, at least in the general form given by $S^{CO}$. As was the case with productivity, $S^{CO}$ pertains to something that goes beyond what speakers have actually done. We can again compare the situation with linguistic creativity and rehearse the observation made in the previous section: whereas $LC^{U}$ can (in the absence of conflicting evidence) be safely inferred from i) which is beyond doubt, $S^{CO}$ can, correspondingly, be inferred from ii), but this is of no real interest since ii) is obviously false.

i)       *S* has understood novel utterances.
($LC^{U}$) *S* understands novel expressions.

ii)      If *S* has understood utterances of the sentences of a set
         of sentences $\varphi$, then *S* has also understood utterances of
         all sentences that have the same mode of combination
         as at least one of the sentences in $\varphi$ and that are
         built up only from expressions that are parts of the
         sentences in $\varphi$.
($S^{CO}$) If *S* understands the sentences of a set of sentences $\varphi$,
         then *S* understands all sentences that have the same
         mode of combination as at least one of the sentences in
         $\varphi$ and that are built up only from expressions that are
         parts of the sentences in $\varphi$.

Few speakers are such that the utterances that they have been exposed to and understood satisfy ii). Not everyone that has understood utterances of 'Mary loves John' has understood an utterance of 'John

---

Some of Fodor and Pylyshyn's remarks suggest that they do not see systematicity as completely general phenomenon. So what I will later refer to as $S^{CO}*$ might fit what they wanted to maintain better than does $S^{CO}$.

loves Mary'. No speaker satisfies ii) if syntax is recursive. Given that speakers' understanding of utterances is not systematic, the justificatory status of systematicity is the same as that of productivity. That is, that speakers really are productive or systematic has to be assumed on the basis of data that do not exhibit, strictly speaking, that they are either. So systematicity is also a weaker justificatory basis for believing in compositionality than is linguistic creativity.

And just as there were putative counterexamples to $P^U$ there are putative counterexamples to $S^{CO}$. This can be seen clearly when we plug an analysis of 'understanding' in terms of meaning into the statement of $S^{CO}$ – which we need to do in order to get an explanandum that is more readily connectable to the principle of compositionality.

(S$^{CO}$)   If $S$ and $T$ assign similar meanings to the sentences of a set of sentences $\varphi$, then $S$ and $T$ assign similar meanings to all sentences that have the same mode of combination as at least one of the sentences in $\varphi$ and that are built up only from expressions that are parts of the sentences in $\varphi$.

This is problematic. It seems attractive to hold, for instance, that (12) is meaningless, but at the same time hold that sentences (13)–(15) are meaningful.[26]

---

[26] See Chomsky (1957: 15). Other famous examples include 'The vertebrate silence worries the legal sail', from Tesnierè (1953) and 'Quadruplicity drinks procrastination', from Russell (1940: 166). Quine ([1975] 1981: 110) disagrees that sentences like these really are meaningless. However, he does not, it seems to me, deny that they are *intuitively* meaningless, asserting only that it is better to *treat* them as trivially false because this simplifies the grammar. I will instead take sentences like these at face value.

The examples are contrived and peculiar, and this might suggest that they can be ignored. However, there are many simpler examples along the same lines. The problem with (12) is that it contains a series of category mistakes. Ideas are not the kind of thing that can be said to sleep or be colorless or green, and sleeping is not something one can be said to do furiously. But I take it that the reason why (12) is meaningless is that it contains *at least one* category mistake, not that it contains several. So if (12) is meaningless so are 'Ideas sleep', 'I sleep furiously', 'Ideas are green'. Examples like these are legion. See Chapter Five for further discussion of meaninglessness.

(12)    Colorless green ideas sleep furiously.
(13)    I want a vase that is either colorless or green.
(14)    I get my best ideas while I'm sleeping.
(15)    Huge endangered bears attack furiously.

Given that $S$ and $T$ assign (similar) meanings to (13)–(15), $S^{CO}$ now entails that (12) is meaningful for them. But this seems incorrect. Thus doubt is cast on $S^{CO}$.

Since something like $S^{CO}$ seems to be true for some portions of language, one might maintain that although $S^{CO}$ is false, something close to it, like $S^{CO}*$, is true.

($S^{CO}*$)    If $S$ and $T$ assign similar meanings to the sentences of a
          set of sentences $\varphi$, then *in many cases $S$ and $T$ assign
          similar* meanings to all sentences that have the same
          mode of combination as at least one of the sentences in
          $\varphi$ and that are built up only from expressions that are
          parts of the sentences in $\varphi$.

Though $S^{CO}*$ is vague, it seems that, in those homogenous regions of language where categorical mismatches do not show up, there are base-set/projection-set pairs of the kind that fit $S^{CO}*$. This calls for an explanation.[27]

### What Explains why Understanding is Closed under Syntactic Operations?

Explaining $S^{CO}*$ is very similar to explaining $LC^{U}$. $S^{CO}*$ is the phenomenon that whenever $S$ and $T$ assign similar meanings to the sentences in some set (the base set), they also assign similar meanings to the sentences in a second set (the projection set). The latter are sentences with the same mode of composition as at least one of the sentences in the base set that only have parts that are featured by the

---

[27] There are other ways to salvage $S^{CO}$. One could frame it in terms of assigning similar semantic ranges instead of in terms of assigning similar meanings and thus avoid the commitment to meaningfulness. Another approach would be to argue that the category mismatches are ruled out syntactically. So even though 'John likes ice-cream' is meaningful, 'Ice-cream likes John' might not be. But by reference to a syntactical restriction, such as the rule that transitive verbs like 'loves' only take animate arguments as subjects, these kinds of example might not counter-exemplify $S^{CO}$.

sentences in the base set. Explaining this seems to place a restriction on the possible explanations of why $S$ and $T$ assign similar meanings to the sentences of the base set – namely, that the explanation invoked for this must entail that $S$ and $T$ assign similar meanings to the sentences in the projection set.

A base set can include any sentences, and thus many sentences that are not familiar to the speaker (typically she will not have encountered utterances of these expressions). So to explain why $S$ and $T$ assign similar meanings to sentences of this set is in effect to explain linguistic creativity (vis-à-vis a small set of sentences). Hence, we have the explanations of the previous chapter as possible explanations of why speakers assign similar meanings to the sentences of their base sets. To find an explanation of $S^{CO}$, we now have to see whether any of the explanations of $LC^U$ corresponding to all the relevant base sets are such that they entail that $S$ and $T$ will assign similar meanings to the sentences in the corresponding projection sets. All the explanations of $LC^U$ considered in the previous chapter fit this mold.

Consider, first, a compositional explanation of linguistic creativity in terms of a $C^{FI}$ theory of, say, a speaker understanding 'John loves Mary'. Theories of this kind will explain the fact that $S$ and $T$ assign similar meanings to 'John loves Mary', in spite of its being novel to them, by assuming three things: that $S$ and $T$ assign similar meanings to the parts of this expression (described by the explicit semantic rules of the theory); that they assign similar modes of composition to this expression (described by the syntax of the theory); and that they are disposed in a certain way to assign meanings to complex expressions based on the meanings they assign to parts of these expressions and their modes of composition (described by the implicit semantic rules). Given these assumptions, and given that $S$ and $T$ have experiences that lend themselves to the formation of the relevant hypotheses about the meanings of atomic expressions and the modes of composition to assign to complex expressions, and given lastly that the theory describing $S$ also gives details of $T$, we have a causal explanation of this limited form of linguistic creativity. We have provided details which are common to the causal histories of the meaning assignments of $S$ and $T$. And this theory is compatible with what we know about the meaning assignments relating to complex expressions (i.e. that they depend on modes of composition and meanings of parts).

This explanation also explains why $S$ accords with $S^{CO}*$ when it comes to the base-set/projection-set pair {'John loves Mary'} – {'Mary loves Mary', 'Mary loves John', 'John loves John'}. For if the syntactic structure of 'John loves Mary' is something like [$_{NP}$ John][$_{VP}$ loves][$_{NP}$ Mary], then the facts that the parts of 'John loves Mary' are assigned similar meanings by $S$ and $T$, that $S$ and $T$ are disposed to assign meanings to complex expressions in a similar way, and that the sentences of the projection set make use of the same mode of composition and the same parts as those in the base set (something that follows from the definition of the 'base set' and 'projection set' given $S^{CO}*$) will lead $S$ and $T$ to assign similar meanings to the sentences in the projection set. Hence the explanation of $LC^U$ in terms of $C^{FI}$ is also an explanation of $S^{CO}$.

This conclusion follows for the alternative explanations of $LC^U$ as well. The difference between these and the $C^{FI}$ explanations lies in their inclusion of theories that might make reference to additional determinants of the meanings of complex expressions. But on the assumption that there are implicit semantical rules, corresponding to modes of composition, which ensure that the meanings of all complex expressions of a certain mode of composition are computed in similar ways, we have a general guarantee that the fact that $S$ and $T$ assign a similar meaning to a certain complex expression will lead to them assigning similar meanings to all other expressions of the same mode of composition and featuring the same parts as long as they possess similar determinants.

For instance, consider an explanation of $LC^U$ in terms of $F^{FI}$. Given an explanation of, say, the understanding of 'John likes "Lisa"' couched in terms of an implicit semantic rule featuring a variable that ranges over forms of expression, it follows, just as if the rule had been a composition rule, that $S$ and $T$ will assign similar meanings to other complexes that can be formed in the same way (e.g. {'Lisa likes "Lisa"', 'Lisa likes "John"', 'John likes "John"'}) – always assuming, of course, that $S$ and $T$ assign similar meanings to parts, the same modes of composition, and the same forms to expressions. Similarly for $X^{FI}$ and $K^{FI}$ explanations of $LC^U$.

The very same resources that were shown to be sufficient for explanations of $LC^U$, then, are also sufficient for explanations of $S^{CO}*$. So even though a theory's being in accord with compositionality (explicated in terms of $C^{FI}$) can explain $S^{CO}*$, theories are not *required*

to be compositional to do this. This means that $S^{CO*}$ gives us no basis on which to require that a semantic theory should accord with *C*.

### 4.3.2 Semantical Relatedness

When Fodor first drew attention to it, systematicity did not seem to involve anything over and above something like $S^{CO*}$.[28] But in subsequent work, Fodor has emphasized additional ways in which the understanding of speakers is systematic.

> We now add that which sentences are systematically related is not arbitrary from a semantic point of view. For example, being able to understand "John loves the girl" goes along with being able to understand "the girl loves John", and there are correspondingly close semantic relations between these sentences: in order for the first to be true, John must bear to the girl the very same relation that the truth of the second that the girl to bear to John. By contrast, there is no intrinsic connection between understanding either of the John/girl sentences and understanding semantically unrelated formulas like "quarks are made of gluons" or "the cat is on the mat" or "2+2=4"; it looks as though semantical relatedness and systematicity keep quite close company.
>
> (Fodor and Pylyshyn 1988: 41-42)

It seems that what Fodor and Pylyshyn are after is something like the following principle.

($S^{SR}$)    The sentences of each of *S*'s base sets are semantically related to the sentences of their projections sets.[29]

Assume that the previous characterization of base sets and projection sets is correct and that (16) is thus one base-set/projection-set pair.

|  | *Base Set* |  | *Projection Set* |
|---|---|---|---|
| (16) | 'John loves Mary' | → | 'Mary loves John' |
|  |  |  | 'John loves John' |
|  |  |  | 'Mary loves Mary' |

---

[28] See Fodor (1987: 149ff).
[29] 'SR' in '$S^{SR}$' is short for 'Semantic Relatedness'.

If this is the case, $S^{SR}$ entails that 'John loves Mary' is semantically related to 'Mary loves John', 'John loves John', and 'Mary loves Mary'. What exactly this involves is somewhat obscure, however, since it is unclear that what 'being semantically related' amounts to. I can imagine two different interpretations.

First, it could simply be the statement that the expressions that build up the sentences in the base set mean the same *in these sentences* as they mean *in the sentences* in the projection set. On this interpretation 'John loves Mary' and 'Mary loves Mary' are semantically related if, for instance, 'loves' mean the same thing in both sentences. Call this interpretation $S^{SR1}$.[30]

Second, one can interpret $S^{SR}$ as a claim about the involved expressions being semantically similar, in the sense of being semantically partially identical.[31] On this interpretation 'John loves Mary' and 'Mary loves Mary' are semantically related if their meanings are partially identical. Call this interpretation $S^{SR2}$.[32]

*Is $S^{SR}$ True?*

It is doubtful that $S^{SR}$ is true on either interpretation. Let us consider $S^{SR1}$ first. On many occasions it seems that what an expression means *in* another expression depends on that latter expression. Consider the variation exhibited by 'lion' and 'drop' in the following sentences for instance.[33]

(17a)  Four lions occupy the corners of Trafalgar Square.
(17b)  Four stone lions occupy the corners of Trafalgar Square.
(18a)  Most students here drop geography in their final year
(18b)  Most students here drop geography lectures in their final year.

In (17a) 'lions' seems to mean 'lions', but in (17b) it seems to mean 'representations of lions'; and in (18a) 'drop' seems to mean 'drop

---

[30] This interpretation is supported by the fact that Fodor and Pylyshyn seem to think that what would explain it is that 'a lexical item [makes] approximately the same semantic contribution to each expression in which it occurs' (42).

[31] See Armstrong (1978: 95ff) for advocacy of the notion that similarity is to be understood in terms of partial identity.

[32] This interpretation is supported by the fact that Fodor and Pylyshyn exemplify the semantic relatedness in terms of the conditions for truth of two sentences *involving the same things* (the girl, john, and love).

[33] The examples are from Cohen (1987).

studying', but in (18b) it seems to mean 'drop attending'.[34] Yet this is inconsistent with $S^{SR1}$ given the existence of the following base-set/projection-set pairs.

> *Base Set*
>
> (17*) 'Four lions occupy the corners of Trafalgar Square'
> 'I love stone sculptures'                    →
>
> *Projection Set*
> 'Four stone lions occupy the corners of Trafalgar Square'
> …
>
> *Base Set*
>
> (18*) 'Most students here drop geography in their final year'
> 'I love math lectures'                    →
>
> *Projection Set*
> 'Most students here drop geography lectures in their final year'
> …

So it seems we have reason to think that $S^{SR1}$ is false.

Consider now $S^{SR2}$. Although I am unaware of any data contradicting it, there do not seem to be any clear reasons to think it is true either. The intuitions that could support it are muddled by irrelevant information.

Note, first, that it can be hard to tell whether two sentences have similar *meanings*, because the circumstances in which utterances of these sentences are true might be similar. The case that Fodor and Pylyshyn drew attention to points only to similarities between the circumstances in which utterances of 'John loves the girl' and 'the girl loves John' will be true. Circumstances of both kinds will contain John, the girl and the love one of them feels for the other. Now if

---

[34] It seems implausible to invoke lexical ambiguity in order to explain these cases, since they are near enough productive. The meaning of 'drop', e.g., co-varies with the meaning of the adjoining noun-phrase, so explaining it in terms of ambiguity would mean that 'drop' has as many meanings as its potentially adjoining noun-phrases. Even if these are not infinitely many, the concession that 'drop' corresponds to a couple of thousand different homonyms with different meanings is unattractive. See Chapter Five for more discussion of cases like these.

meanings are identified with such circumstances, it will follow that they, the meanings, have the same parts and are thus partially identical. But since, in the present context, we are not assuming anything about what meanings are, these data give us no reason to think that the meanings of the sentences are partially identical. Indeed the fact that the data are available obscures any other pertinent intuitions we might have.

Note, second, that whether two sentences have similar *meanings* is obscured by the fact that the relevant sentences will always have similar syntax (since we are concerned with similarities in the meanings of sentences in the base set and the meanings of the sentences in the projection set, and it follows from $S^{CO*}$ and the definitions of base set and projection set that these sentences will always be syntactically similar). For instance, there is obviously something very similar about (19) and (20).

(19)   I love candy.
(20)   I went to David's house.

But is this similarity due only to (19) and (20) featuring 'I' or their meanings featuring a common part? I am inclined to think the latter.

When we try to disregard extensional and syntactical information it seems that not much remains that would support a belief in $S^{SR2}$. It is therefore doubtful that $S^{SR}$ is true on either of the offered interpretations. Hence, consideration of $S^{SR}$ gives us no reason to think that $C$ is true.

### What is required to explain $S^{SR}$?

Since we have no reason to think that $S^{SR}$ actually obtains, it is really immaterial, from the perspective of the truth of $C$, what is required in order to explain $S^{SR}$. But, out of curiosity, it is perhaps worth mentioning that theories according with $C^{FI}$ are in general insufficient to explain $S^{SR}$ on either interpretation. There is no guarantee, in a semantic theory that accords with $C^{FI}$, that the meanings of parts of complex expressions will mean the same in all expressions in which they feature: $C^{FI}$ has nothing at all to say about the notion of 'what $x$ means *in y*'. Nor is there any guarantee that the meanings of two syntactically related expressions will have similar meaning. However,

theories according with $C^{BU}$ do generally suffice to explain $S^{SR}$ on both interpretations. According to $C^{BU}$ the meaning of a complex expression contains the meanings of its parts as parts. This means that if $e_1$ and $e_2$ are both expressions having $e'$ as an immediate part, $\mu(e_1)$ and $\mu(e_2)$ have a common part (i.e. $\mu(e')$). It seems very natural to identify this part with what $e'$ means *in $e_1$* and *in $e_2$*. So $S^{SR1}$ follows from the fact that a theory is in accord with $C^{BU}$. In addition, since $\mu(e_1)$ and $\mu(e_2)$ have a common part it follows that they are partly identical, and hence similar. So $S^{SR2}$ also follows from a theory's being in accord with $C^{BU}$.

### 4.3.3 The Closure of Understanding under Sub-Terms

The final phenomenon I will discuss has not been placed under the heading of 'systematicity', but rather under 'reverse compositionality'. However, since it is bears a close resemblance to the systematicity phenomena, and in particular $S^{CO}$, and since it might support a belief in $C^{BU}$ if it is true, I will attend to it as well.

According to Fodor and Lepore something like the following, which is analogous to what is sometimes called *closure under subterms*, is true of speakers of natural languages.[35]

($S^{CS}$)   If $S$ is able to understand a complex expression, then she is able to understand each of the parts of that expression.[36]

---

[35] See Fodor and Lepore ([2001] 2002). In the context of formal semantics a meaning function is closed under subterms if it is such that each part of a meaningful expression is also meaningful. See, e.g., Westerståhl (2004: 564).

[36] 'CS' in '$S^{CS}$' is short for 'Closure under Subterms'.

$S^{CS}$ is a slightly more general claim than that explicitly made in Fodor and Lepore ([2001] 2002). They claim that what needs explaining is that 'you practically never find people who understand "dogs bark" but don't understand "dogs" or "bark"', and that 'people who understand ["dogs bark and cats purr"] generally understand both conjuncts' (Fodor and Lepore [2001] 2002: 60). However, they clearly need a fairly general phenomenon to support their conclusions. Thus they state that 'the reverse compositionality of complex expressions relative to their lexical constituents is just a special case of the reverse compositionality of complex expressions with respect to their constituents tout court, lexical or otherwise. Since, in natural languages, every constituent expression has infinitely many hosts, this amounts to an infinite amount of reverse compositionality' (Ibid: 60). Apparently, then, Fodor and Lepore think the phenomenon is wide enough to cover every constituent. Their hedging ('practically never' and 'generally') seems to serve only to remove idioms from consideration.

So according to $S^{CS}$, if a speaker understands 'dogs bark', she will understand 'dogs' and 'bark' as well.

*Is Understanding Closed under Subterms?*
It seems that $S^{CS}$, as stated, is false. A speaker might understand a complex expression without understanding its parts. Kenneth Johnson has offered a counterexample in terms of the following expressions.[37]

(21)  building the house
(22)  build

Johnson points out that 'build' is semantically speaking a *telic* verb i.e. such that it specifies a temporal endpoint in constructions like (23).

(23)  Mary built the house.

Its being telic makes 'build' different from, for instance, 'watch' which is an *atelic* verb. Hence, 'watch' does not specify a temporal endpoint in (24).

(24)  Mary watched the house.

Johnson's point is that constructions resulting from the attachment of the progressive morpheme 'ing' to a verb conceal its telicity. Hence, 'build' does not specify a temporal endpoint in constructions like (25).

(25)  Mary was building the house.

But then, the argument goes, it seems reasonable to assume that a speaker could understand (25) (or, more specifically, (21)) without

---

They say: 'It is not, of course, a necessary truth that if you understand a syntactically conjunctive sentence you understand each syntactic conjunct; the sentence might be an idiom' (ibid, n.10).
[37] See Johnson (2006).

knowing whether 'build' is telic or atelic, i.e. without understanding 'build'.

Much less complicated counterexamples can easily be imagined. Many speakers understand 'baseball bat' prior to understanding 'baseball'; when learning first order logic, many speakers understand the notion of a 'universal quantifier' before understanding 'quantifier', and so on. One might perhaps argue that these counterexamples are not really complex (at the time at which the speaker understands only them, not their parts). But this seems highly implausible with counterexamples that can be devised involving pleonastic expressions. To see this, assume that *S* understands (26), below, relative to a speech community *G*, but does not assign (27) any meaning (and thus do not *understand* it relative to any community).

(26) Spinster
(27) Female

Now, it seems perfectly coherent to say that if *S* was told that, relative to *G*, (28) means the same thing as (26), and if *S* adjusted his meaning assignments accordingly, *S* would then understand (28) but not one of its parts.

(28) Female spinster

Since (28) is pleonastic, from *S's* perspective (27) could have any meaning – e.g. FEMALE or UNMARRIED – which can be absorbed by the meaning of (26). To sum up, in numerous situations, plausibly, it is not always the case that if someone understands a complex expression she will also understand its parts.

Besides these counterexamples, $S^{CS}$ seems more generally problematic, since it takes speakers always to have solved what we can call 'the mapping problem'. Even if the speaker is certain about what collection of meanings she should assign to parts of an expression, she might be at a loss to know which meanings to assign to which parts.[38] For instance, even if *S* knows that 'Red Triangle' denotes the set of red triangles, and that it is an adjective-noun combination, *and* that

---

[38] This, it seems to me, is what Robbins (2005) takes to be the core problem with any principle similar to $S^{CS}$.

the denotations of the parts of 'red triangle' are the set of red things and the set of triangles, $S$ might not know that it is 'red' that denotes the red things and not the triangles and vice versa.

The previous considerations suggest that $S^{CS}$ is false. What seems more likely to be true is something like the following.

($S^{CS}*$)  If $S$ is able to understand a complex expression then she
is *often* able to understand the parts of that expression.


*What explains the closure of understanding under sub-terms?*

Had $S^{CS}$ been true we would have needed something like $C^{BU}$ in order to explain it. For if a speaker is characterized correctly by a $C^{BU}$ theory, the meaning of each complex expression is a structure built up from the meanings of the its parts. In this case, if the structure of the meaning of a complex expression indicated which meanings belong to which parts of that expression, we know the meaning of the parts of that expression if we know its meaning. Theories according only with $C^{FI}$ do not facilitate these explanations; there is no general guarantee that the meanings of the parts of a complex expression can be recovered from the meaning of a complex expression. However, since we have reason to think that $S^{CS}$ is false, it is more interesting to consider what is required in order to explain $S^{CS}*$.

Robbins offers the following insight concerning this issue. $S^{CS}*$ is explained by only assuming a principle such as $C^{FI}$.[39] For given that we, speakers, often employ composition rules (or $F^{FI}$ rules, and so on, depending on which kind of determination is going on) when we understand utterances, we are in fact in a situation where we know the meaning of a complex expression only if we know the meanings of the parts of that expression. Where the utterances we encounter are utterances of novel expressions we must employ the said rules, and thus must often understand the parts of the expressions. So $S^{CS}*$ is explained by the determination principles that we have previously employed in order to explain linguistic creativity, productivity and systematicity.

---

[39] See Robbins (2005). Any of the other determination principles (such as $F^{FI}$, $X^{FI}$, or $K^{FI}$) will work as well.

## 4.4 Concluding Remarks

The conclusions of this chapter should be disappointing for the compositionalist. The phenomena we have considered that are likely to be real – productivity, $S^{CO}*$, and $S^{CS}*$ – *can* be explained by compositionality (in the sense of $C^{FI}$), but they *can also* be explained by other principles. In addition, both productivity and systematicity (in the sense given by $S^{CO}*$), lack the kind of immediate grounding in speakers' understanding of utterances that linguistic creativity has, and to this extent they give us less compelling reasons to believe in the principles that can be recruited to explain them.

Linguistic creativity, productivity and systematicity are the three most commonly cited reasons for believing in compositionality, and we can thus conclude at this point that, from the three most commonly cited reasons for believing in compositionality, there does not seem to stem any reason to think that the principle of compositionality is true. We are therefore not justified in imposing the compositionality constraint (or any of the other particular determination principles) on semantic theories. What we are allowed to impose is the constraint that a semantic theory must accord with *some* determination principle such as $C^{FI}$, $F^{FI}$, $X^{FI}$, or $K^{FI}$. In Chapter Six I explore what this means from the perspective of what we can learn about meaning from compositionality. Chapter Five will defend the agnostic position reached here by showing that none of the phenomena usually cited as problems for compositionality are sufficiently problematic to make a case for rejecting compositionality. At any rate, they are not problematic for some versions of $C$ (in particular thin versions of $C^{FI}$).

Before moving on to discuss potential problems for compositionality I want, briefly, to comment on $C^{BU}$. It is commonly assumed that propositions (i.e. the entities normally taken to be the meanings of utterances) are structured. In parts of their discussion on systematicity, it seems that Fodor and Pylyshyn do not clearly distinguish two levels of meaning, for some of their results, if they are to pertain to the meaning of expressions, seem to assume that something like propositions are also the meanings of expressions. What is interesting is that, even if this is denied, there seem to be plenty of things that one can use the hypothesis that meanings of expressions are structured complexes in order to explain. In

explaining productivity, the hypothesis can ensure that the semantic theory encompasses an infinitude of meanings; in explaining $S^{SR}$, it can be used to account for semantic similarity; and in explaining $S^{CS}$, it can guarantee the availability of certain meanings if one has access to certain others. If we had reason to think that the semantic competence of a speaker involves these phenomena, the phenomena could be called upon to support the idea that the meanings of expressions (i.e. the meanings of utterance types) are structured. However, since the data do not give us the reasons indicated, it seems that the assumption that some meanings are structured complexes is too strong. Certainly a principle, such as $C^{BU}$, entailing that the meanings of *all* complex expressions are complex structures is much too strong. Nonetheless, one appreciates the prima facie attraction of $C^{BU}$, since the assumption that meanings are structured can be used for so many different things.

*…in which I myself do not believe, but which regularly visit at night and in the weary twilight with the illusory force of axioms.*

– Jorge Luis Borges

# 5

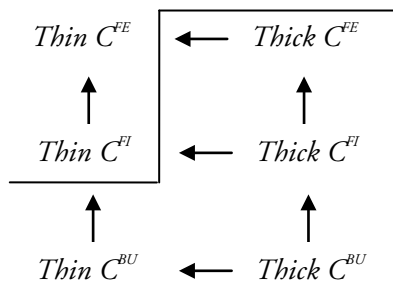# Supposed Problems with Compositionality

The goal of inquiry in this essay is to ascertain the extent to which the principle of compositionality (*C*) can be justifiably imposed as a constraint on semantic theories and thereby provide information about what meanings are.

(C)    The meaning of a complex expression is determined by the meaning of its parts and its mode of composition.

It cannot be justifiably required that a semantic theory should accord with *C* if it turns out that there are reasons to think that *C* is false, i.e. if it turns out that there are reasons to think that theories in accordance with *C* are incorrect. This prompts us to look closer at reasons to think that *C* is false, and this is what we will do in this chapter.

The subsequent discussion will be broken down into seven parts. Each section will address a particular potential problem. The chapter will discuss the problems that have been assumed to stem from the existence of idiomatic expressions such as 'It is raining cats and dogs' (Section 5.1.1), the existence of ambiguous expressions such as 'The philosophers lifted the piano' (Section 5.1.2), the existence of meaningless expressions such as expletives like 'it', and category mismatches such as 'green ideas' (Section 5.1.3), the existence of interaction effects such as 'lion' meaning 'representation of lion' in 'stone lion' (Section 5.1.4), the supposed existence of a particular form of context sensitivity in expressions such 'It rains' (Section 5.1.5), attempting to provide an adequate semantics for propositional attitude reports such as 'Kim believes that dogs are dangerous' (Section 5.1.6) and attempting to provide an adequate semantics for quantified conditionals such as 'Every student will succeed unless he goofs off'(Section 5.1.7).

It will be concluded (in Section 5.2) that none of these phenomena are problems for all the explications of compositionality that we have considered so far. But some of them are problems for some of the explications. In particular, (thin) $C^{FE}$ and (thin) $C^{FI}$ seem defensible, and the remaining explications problematic. This can be illustrated in the following way where the entailment relations between the principles have been made explicit and the acceptable explications are placed above the line.

$$
\begin{array}{ccc}
\textit{Thin } C^{FE} & \longleftarrow & \textit{Thick } C^{FE} \\
\uparrow & & \uparrow \\
\textit{Thin } C^{FI} & \longleftarrow & \textit{Thick } C^{FI} \\
\uparrow & & \uparrow \\
\textit{Thin } C^{BU} & \longleftarrow & \textit{Thick } C^{BU}
\end{array}
$$

An important feature of the dialectic should be emphasized before the various problems are engaged. At several points in the discussion it will be suggested that various phenomena and various constructions *could* reasonably be understood in a compositional way. I do not want to argue for the stronger claim that the compositional alternatives are

to be preferred over the non-compositional alternatives. Such judgments will have to be suspended until particular theories covering several different phenomena can be compared. In addition, making such judgments also depend on a much more in-depth discussion of the topics covered in this chapter than the one that can be offered here. So my main conclusion following this chapter and the previous two is thus the modest one that, *when it comes to justifiably imposing compositionality as a constraint on semantic theories, the problem is that we do not have any conclusive reasons to think that compositionality is true, not that there exists conclusive evidence to think that it is false.*

## 5.1 Supposed Problems

### 5.1.1 Idiomaticity

The *idiomatic expressions* of natural languages are perhaps the expressions that, at first glance, most immediately seem to be in conflict with the principle of compositionality. Consider, for instance, the following four examples.

(1)    kick the bucket
(2)    spill the beans
(3)    jump the gun
(4)    It's raining cats and dogs.

The common-sense conflict between these expressions and compositionality is that it is hard to believe that a speaker would be able figure out, for instance, what 'kick the bucket' means (idiomatically) from her knowledge of the meanings of 'kick', 'the' and 'bucket'.[1]

Notice first that the perceived tension does not pertain directly to how compositionality explains, for instance, speakers' ability to understand novel expressions. The idiomatic expressions are not normally among those that we understand when they are novel to us. Instead, the problem, if it is a problem, seems to be that the compositionality thesis is stated in an overly general way. It quantifies over *all* complex terms and conditions their meanings to be in a

---

[1] I will not attempt a general definition of 'idiomatic expression' but rely instead on examples such (1) – (4) and an intuitive understanding of this notion.

certain way even though what, strictly speaking, is needed for explaining speakers ability to understand novel expression is just that all complex, *novel*, terms are conditioned in this way. So if there is a problem, it seems that it is not directly connected with providing an explanation in terms of compositionality, but with the generality of the imposition of the compositional mold on the terms covered by explaining theory. We might thus avoid the problem completely by a simple restriction of the generality of the compositionality principle.

Notice second that the complexity that $C$ makes reference to on all explications we have so far considered, pertained to the complexity of *grammatical terms* and not to the complexity of *expressions*. Whether the grammatical terms underlying expressions such as (1) - (4) really are complex cannot be straightforwardly read off their surface structure, and given that it turns out that the grammatical terms are simple, there is no conflict at all between $C$ and the idiomatic expressions, for the latter is concerned only with the meaning determination of complex grammatical terms. [2]

It seems very likely that the grammatical terms corresponding to a wide range of idiomatic expression really are simple.[3] One of the most compelling reasons for thinking that certain terms are complex is that if we treat them as complex we can thereby explain how we can understand them without prior exposure. But we *cannot* understand expressions like (1) – (4) without prior exposure, so there is no such reason.[4]

---

[2] The assumption that all *expressions* that contain meaningful parts correspond to complex grammatical terms is clearly incorrect. On this assumption, expressions like, for instance, 'fortune', 'caraway', 'margin', 'winnow', 'pillage' and 'capsize' would correspond to complex grammatical terms.

[3] Most idiomatic expressions have literal meanings as well. I will ignore this and the grammatical terms that carry these meanings since it seems that they can be treated normally. E.g. the syntax and semantics corresponding to the literal reading of 'John kicked the bucket' seems unproblematic.

[4] In some discussions of idiomatic expressions it is pointed out that saying of an expression that it is an idiom is fine from the perspective of providing the explanations that compositionality is aimed to provide, as long as this is only done *a finite number of times*. The emphasis is on the *number* of times this is permitted if one is concerned with the *productivity* of thought and language. If it is assumed that more than a finite number of expressions are idioms (and they are understood as syntactically simple) then we have no explanation of how speakers can understand them all. However, if one like myself, thinks that compositionality should first and foremost account for speakers understanding of *novel* expressions then this requirement is too weak. It is only the expressions that we cannot understand when

That idiomatic expressions such as (1) – (4) really correspond to simple grammatical terms is supported by some plausible syntactic tests for semantically relevant syntactic constituency.[5]

First, what seem to be noun phrases in (1) - (3) cannot be the antecedents of anaphoric expressions in such a way that the idiomatic interpretation is transferred to the expression containing the anaphor. Compare the normal behavior of (5) and (6) with the behavior exhibited by (7) - (9).

(5)    John struck the match.       James struck it too.
(6)    John antagonized the fans.   James antagonized them too.

(7)    John kicked the bucket.      ?James kicked it too.
(8)    John spilled the beans.      ?James spilled them too.
(9)    John jumped the gun.         *James jumped it too.

Second, expressions such (1) – (4) cannot be sub-divided into question-answer pairs corresponding to the supposed constituency breaks in such a way that the answer carries the semantic interpretation of the original idiom. Compare the normal behavior of (10) and (11) with the behavior exhibited by (12) - (15).

(10)   What did john strike?        The match
(11)   Who did John antagonize?     The fans

(12)   What did John kick?          The bucket
(13)   What did John spill?         The beans
(14)   *What did John jump?         The gun
(15)   *What did it rain?           Cats and dogs

Third, idiomatic expressions do not support the general entailment relations – existential generalization and simplification of conjunction for instance – that we should expect from expressions corresponding

---

they are novel to us that can be idioms (given that idioms are understood to be syntactically simple).

[5] These tests are versions of the standard syntactic tests on syntactic constituency. See Haegeman (1995) and Carnie (2002). Modifications have been made only to avoid obscuring intuitions pertaining to the literal interpretations of the idiomatic expressions.

to the complex terms that seems closest at hand for expressions having the surface structure of expressions such as (1) – (4). Compare the entailments of (16) and (17) with the absence of entailments of (18) - (21).

(16)   John struck the match.          → John struck something.
(17)   John killed Jane and Joe.          → John killed Jane.

(18)   John kicked the bucket.          ↛ John kicked something.
(19)   John spilled the beans.          ↛ John spilled something.
(20)   John jumped the gun.          ↛ *John jumped something.
(21)   It is raining cats and dogs.          ↛ It is raining cats.

The application of these three tests together with the lack for any explanatory reason to think that they correspond to complex grammatical terms suggests that idiomatic expressions such as (1) – (4) do not correspond to complex grammatical terms. Since the complexity relevant for the three explications of compositionality ($C^{FE}$, $C^{FI}$, and $C^{BU}$) that we have primarily been concerned with is that of grammatical terms, I conclude that idiomaticity and compositionality are generally not in conflict.[6]

---

[6] The literature on idiomaticity (e.g. the overview by Titone and Connine (1999)) suggests that there are exceptions to the pattern described in the main text. Some idioms such as 'pull strings' and 'leave no stone unturned' exhibit behavior that might suggest that they are not syntactically simple. Although a complete defense of compositionality in light of exceptions like these should consider these exceptions in more detail, I will have to confine myself in this chapter to some very brief remarks in this footnote.

Consider first the fact that speakers who become familiar with the idiomatic meaning of 'pull strings' seems to be able to understand a range of related expressions idiomatically. Speakers familiar with the idiomatic meaning of 'John pulled some strings', for instance, is likely to also understand 'Some strings were pulled by John' idiomatically. This can be taken as evidence that 'pull strings' correspond to a complex grammatical term. But this might confuse the complexity of expressions with that of grammatical terms. For it seems that both surface strings can result from a grammatical term where the 'pull strings'-part is atomic, and the fact that 'pull' and 'strings' is separated in the expression need not have any deeper semantic significance. These kinds of idioms might be akin to the separable prefix verbs of German, such as 'anfangen' where the prefix (an-) of these words separates when the verb is conjugated. For instance, 'wir fangen an' means 'we begin'.

Consider second the fact that some idioms like 'leave no stone unturned' permits *internal modification* i.e. it permits what appears to be the modification of a part of the idiomatic expression (Nunberg, Sag et al. (1994)). 'leave no stone unturned'

## 5.1.2 Ambiguity

On a thick ('Russellian') interpretation of the claim '*the* meaning of a complex expression is determined by the meanings of its parts and its mode of composition', it entails that complex expressions have unique meanings. This seems to contradict the undeniable fact that many expressions of natural languages are *ambiguous*, i.e. have several meanings.[7] Consider, for instance, the following six examples.

(22)    Jane left the bank.
(23)    The ring is in the boot.
(24)    John loves kind men and women.
(25)    The hunter shot the moose with the rifle.
(26)    The philosophers lifted the piano.

---

means roughly the same as 'exhaust all options'. 'leave no *legal* stone unturned' does not seem to mean the same as 'legally leave no stone unturned' (i.e. 'legally exhaust all options') but the same as 'exhaust all legal options'. To see the distinction more clearly, consider a situation where there is a legally mandated cost for exploring more than five options say, but all the relevant options can be explored legally. One could thereby illegally explore all legal options (if there are more than five) by not paying the cost. So it seems that 'legal' does not modify the whole idiom but only a part of it. That this is possible suggest that there is a part to modify. However, in cases like these there is a pull to say that the constituents themselves have idiomatic meanings that they bestow to their hosts, and that there really is no problem with compositionality. To speakers familiar with the idiom 'leave no stone unturned', 'leave unturned' might have the meaning of 'remain unexplored' and once it is combined in the right way with an expression meaning 'stone' the idiomatic reading result. What the merits of this proposal are depends in part on whether one can avoid over-generalizing idiomatic readings. In this particular case it might be possible to assign 'leave unturned' a metaphorical meaning which is dependent on the occurrence of another meaning in order to influence the meaning of a complex expression since it seems that 'leave no legal rock unturned' also has a metaphorical reading.

  Although this is probably too concessive, these more complicated idioms might be made compatible with $C^{FI}$ by dropping the clause in $C^{FI}$ pertaining to explicit meaning assignments. The resulting principle would retain the explanatory utility of $C^{FI}$, and it might additionally avoid some of the counterexamples against reverse compositionality (see the previous chapter) that might also be construed as counterexamples to $C^{FI}$. But since the principle no longer entails $C^{FE}$ it is arguably not a compositionality explication.

  See Westerståhl (2002) for a very rewarding discussion of the general problem of how to extend different compositional semantics to incorporate idiomatic expressions.

[7] Pelletier (2000) has advanced this kind of argument.

(27)    Every linguist knows two languages.

All of these expressions seem to have several clearly distinct meanings. 'Jane left the bank' could either mean 'Jane left the establishment for safekeeping of money', or 'Jane left the slope'; 'The ring is in the boot' could either mean 'The ring is in the shoe covering both foot and ankle', or 'The ring is in the luggage compartment at the back of the car'; 'John loves kind men and women' could either mean 'John loves kind men and kind women', or 'John loves women and kind men'; 'The hunter shot the moose with the rifle' could either mean 'The hunter used a rifle to shoot the moose', or 'The moose with the rifle was shot by the hunter'; 'The philosophers lifted the piano' could either mean 'The philosophers lifted the piano together', or 'The philosophers each lifted the piano individually', and 'Every linguist knows two languages' could either mean 'There are two languages that every linguist knows' or 'Every linguist knows more than one language', The existence of expressions like these might conflict with thick explications of compositionality.[8]

   Notice first (which was also emphasized in the previous subsection) that on the explications of compositionality given so far in this essay, compositionality is a claim about *grammatical terms*, and not directly about expressions. So to the extent that these explications are thick, they entail the claim that grammatical terms have at most one meaning. But this means that a certain expression can have multiple meanings to the extent that it corresponds to several grammatical terms (and it is the case that expressions inherit the meanings of their grammatical terms). This might mean that both thin and thick versions of compositionality is compatible with some of the above constructions.

   Consider first sentences such as (22) and (23). In these cases, the multiplicity of meanings seems to be due to *lexical ambiguity*, i.e. a form of ambiguity which can be explained either in terms of a simple expression corresponding to multiple grammatical terms, or that expression corresponding to a single grammatical term with several

---

[8] I will leave open exactly what kind of judgments that constitutes evidence for and against an expression being ambiguous, hoping that the notion is fairly intuitive. One explication of 'ambigous' that would make the aforementioned six expressions ambiguous is *being synonymous with expressions with different meanings.* But they are also ambiguous on a range of other ambiguity tests. Cfr. Gillon (2004).

meanings. It is, for instance, intuitively very compelling to attribute the ambiguity of (22) to the ambiguity of 'bank', and the ambiguity of (23) to the ambiguity of 'boot'. This is suggested by the fact that the ambiguity persist across a range of expressions featuring 'bank' and 'boot' respectively. In light of this, the ambiguity can be explained either by holding that 'bank' and 'boot' each correspond to multiple grammatical terms (each of which might have a unique meaning) or that they each correspond to a grammatical term with several meanings, and that the multiplicity of meanings corresponding to the complex expressions is carried up by the composition rules. On the former alternative it follows that (22) correspond to grammatical terms that can be represented by (22a) and (22b), and that each of these terms has a unique meaning.

(22a) [$_{NP}$ She [$_{VP}$ left [$_{NP}$ the bank$_1$.]]]
(22b) [$_{NP}$ She [$_{VP}$ left [$_{NP}$ the bank$_2$.]]]

Since this form of explanation only makes use of grammatical terms with unique meanings, it is compatible with both thick and thin explications of compositionality. Hence, cases of ambiguity such as (22) and (23) seem to be unproblematic from the perspective of compositionality.

Consider next sentences such as (24) and (25). In these cases the multiplicity of meanings seems to be due *amphiboly* (*structural ambiguity*) i.e. a form of ambiguity that can be explained in terms of an expression corresponding to several structural derivations.[9] It is compelling to think that (25), for instance, instantiates this form of ambiguity since it seems plausible that prepositional phrases can combine with, and modify, both nouns and verbs. One can imagine something like the following syntactic theory, for instance.

$t_1$      *Syntax:*
     1. N    → rifle, moose, hunter
     2. D    → the
     3. P    → with
     4. V    → shot

---

[9] In these cases it does not seem that the ambiguity can be traced to an ambiguous part.

5.  NP  → D·N
6.  NP  → NP·PP
7.  PP  → P·NP
8.  VP  → V·NP
9.  VP  → V·NP·PP
10.  S   → NP·VP

That something like this theory is correct, and that prepositional phrases can be constitutents of noun phrases, and not only of verb phrases, is supported by the fact that, for instance, 'The hunter with the rifle shot the moose' is an acceptable sentence. Given $t_1$, (25) corresponds to two grammatical terms that can be characterized roughly by (25a) and (25b).

(25a)  [$_{NP}$ John [$_{VP}$ shot [$_{NP}$ the moose [$_{PP}$ with the rifle.]]]]
(25b)  [$_{NP}$ John [$_{VP}$ shot [$_{NP}$ the moose][$_{PP}$ with the rifle.]]]

Given these structural assumptions, the ambiguity of (25) can be explained in terms of different implicit semantic rules corresponding to the different modes of composition. This explanation is lent credence by the fact that that the ambiguity persists across a range of expressions of the same surface structure (i.e. instantiating the structure NP-VP-NP-PP, e.g. 'The bowler watched a park with a telescope'). Since this form of explanation, like the previous one, only makes use of grammatical terms with unique meanings, it is compatible with both thick and thin explications of compositionality. Hence, cases of ambiguity such as (24) and (25) seem to be generally unproblematic from the perspective of compositionality.

   But consider now sentences such as (26) and (27). In these cases the multiplicity of meanings does not seem to be due either to lexical or to structural ambiguity. It is not the case that, for instance, sentences containing the parts of (26) generally manifests the relevant ambiguity (in the same way as sentences containing 'bank' or 'boot' do for instance). For instance, the collective-distributive reading of 'the philosophers' as part of (26) seems to disappear when 'the philosophers' are considered in isolation; it is only once some kind of relation (e.g. lifted) is introduced, that the ambiguity arises. In addition, no structural hypothesis that can explain the ambiguity suggest itself (like the one suggested for (25)). Something like (26a)

seems to be able to roughly capture the underlying structure on both readings.

(26a)  [$_{NP}$ The Philosophers [$_{VP}$ lifted [$_{NP}$ the piano]]]

If these considerations are correct, a single complex term underlies (26). And yet (26) seems to be ambiguous between two different meanings.

The existence of is kind of ambiguity, which has been referred to as *essential ambiguity*, has been used as a general argument against compositionality.[10] But notice that complex grammatical terms' having several meanings is only part of the *thick* explications of compositionality.[11] The explications of compositionality that have mainly considered in the previous chapters have not been of this kind. $C^{FE}$, $C^{FI}$, and $C^{BU}$ are without additional assumption all *thin* explications of compositionality where nothing (or very little in the case of $C^{BU}$) is assumed about the number of meanings of complex terms. They are all perfectly compatible with expressions such as (26) and (27). Since the ambiguity seems to arise, in the case of (26), when the subject noun phrase is combined with the verb-phrase, a theory in accordance with $C^{FI}$, could posit two different implicit semantic rules corresponding to this mode of combination, resulting in expressions such as (26) having two meanings. This is plausible since it seems that the ambiguity persists across a range of expressions of the same surface structure; expressions denoting most animate objects can be

---

[10] According to Jeff Pelletier, these cases of ambiguity are incompatible with compositionality, since 'what compositionality *cannot* admit is that there be no lexical ambiguity, there be but *one* syntactic structure, and yet there be two (or more) meanings for that item.' (Pelletier 2000: 210-211). Westerståhl (2007) has referred to these cases of ambiguity as essential ambiguity.

Noun-noun compounds also exhibit essential ambiguity. 'murder weapon' is typically used to refer to a weapon that *has been used in a murder*, but it could also refer to a weapon that *is especially good for murdering* (or that *is often used for murdering*). These two interpretations cannot plausibly be traced to different lexemes or different structures. On his account of the semantics of noun-noun compounds, Weiskopf (2007) correspondingly posits two different implicit semantic rules in order to get these two readings. Note that these readings are independent of the contextually relevant connection between the meanings of the nouns (which was here assumed to be 'used for murder')

[11] This includes the somewhat thick proposals such as the proposals due to Hodges (2001) which we considered in chapter two.

substituted for 'the philosophers' and expressions denoting many actions that can be done collectively can be substituted for 'lifted the piano'. If the posited rules predict the extent to which the ambiguity shows up in natural language constructions, there does not seem to be anything from an explanatory perspective that would rule them out.[12] This is not possible for thick variants of compositionality since the result of positing two such rules would be grammatical terms with multiple meanings. So the proponent of thick compositionality is left to either deny the data – which seems unattractive – or try to locate lexical or structural ambiguities – which seem very difficult. So it seems that that since the thin explications of compositionality have no difficulties with ambiguity, but the thick versions do, the thin versions have a decisive advantage over the thick ones, and we have reasons to think that the thick explications are false.

### 5.1.3 Meaninglessness

As was already remarked in the previous subsection, a thick ('Russellian') interpretation of the claim '*the* meaning of a complex expression is determined by the meanings of its parts and its mode of composition' entails that complex expressions have unique meanings. This does not only seem to contradict the fact that many expressions have several meanings, it also seems to contradict the fact some expressions have none. Consider for instance the following expressions.

(28)   It surprised Jeeves that the pig had been stolen.
(29)   There are three pigs escaping.
(30)   Ideas are green.
(31)   This stone is thinking of Vienna.

In sentences (28) and (29) the expletives 'it' and 'there' seem to lack meaning even though (28) and (29) themselves are clearly

---

[12] It should be noted that it is very different, from the perspective of the possibility to explain speakers' ability to understand novel expressions, to discover that there are two grammatical terms with different meanings in spite of them having parts with the same meanings and having the same mode of combination, from discovering a single complex term which has two different meanings. In the first case it cannot be that only meanings of parts and modes of combination are used in order to arrive at the meaning of the complex expressions. But in the second case this is still possible.

meaningful. The sentences (30) and (31) seem to have meaningful parts but no meanings of their own. The existence of expressions like these seem to conflict with the thick interpretation of compositionality.[13]

What has been emphasized in the previous two sections – that compositionality primarily concerns grammatical terms and that it concerns expressions only indirectly – applies here as well, but is less important. For if we assume that expressions inherit their meanings from their grammatical terms it follows immediately that if an expression is meaningless, then all of its grammatical terms must be meaningless as well. So since thick explications of compositionality are incompatible with meaningless grammatical terms, they are also incompatible with expressions' being meaningless.

Consider the occurrence of 'it' in (28) and 'there' in (29). It is important to realize that these occurrences are not the same as those of 'it' and 'there' in sentences such as (32) and (33).

(32)   It bit John.
(33)   The book is there.

---

[13] I'm concerned with cases where it is *intuitively* compelling to say that the expressions lack meaning. There are many other ways in which expressions can be held to be meaningless. Consider Ayer for instance when he discuss 'the principle of verification'.

> The principle of verification is supposed to furnish a criterion by which it can be determined whether or not a sentence is literally meaningful. A simple way to formulate it would be to say that a sentence had literal meaning if and only if the proposition it expressed was either analytic or empirically falsifiable.
>
> (Ayer 1946: 5)

or consider George Orwell when he writes.

> ..looking back through my work, I see that it is invariably where I lacked a political purpose that I wrote lifeless books and was betrayed into purple passages, *sentences without meaning*, decorative adjectives and humbug generally
>
> (Orwell [1946] 2001: 463, emphasis added)

These uses of 'meaningless' are very different from ones that could plausibly cause problems for compositionality.

For the occurrences of 'it' and 'there' in (32) and (33) it holds that they can be questioned and that they can receive focal stress.[14]

(34)   What bit John?
(35)   *It* bit John.
(36)   The book is where?
(37)   The book is *there.*

But this does not seem possible with the occurrences of 'it' and 'there' in (28) and (29).

(38)   *What surprised Jeeves that the pig had been stolen?
(39)   **It* surprised Jeeves that the pig had been stolen.
(40)   *Where are three pigs escaping?
(41)   **There* are three pigs escaping.

The occurrences of 'it' and 'there' in sentences such as (28) and (29) are usually treated as *expletives*, expressions serving only to provide an overt subject to a sentence. It seems very plausible to think that expletives have no meaning since (28) and (29) seem to be synonymous with (42) and (43) respectively.

(42)   Jeeves was surprised that the pig had been stolen.
(43)   Three pigs are escaping.

This form of redundancy seems to indicate that the relevant expressions are meaningless.[15] Hence (28) and (29) give us reasons to

---

[14] Cfr. Haegeman (1995).
[15] Not all forms of redundancy indicates meaninglessness. 'His lovely, wonderful husband reads fast' seems to mean the same thing (barring some slight difference in emphasis) as 'His lovely husband reads fast'. These expressions also seem to mean the same thing as 'His wonderful husband reads fast'. In these examples of pleonastic expressions there is a redundancy but the redundant parts are not meaningless. Consider also 'His female sister is angry' which (again barring some slight difference in emphasis) seems to mean the same thing as 'His sister is angry'. Reflecting on these cases it seems that the meaninglessness-indicating redundancy is the form of redundancy where a part $e_1$ of an expression can be dropped without the meaning of $e$ changing, or there being another part of $e$, $e_2$ such that $e_1$ and $e_2$ are synonymous or $e_1$ following analytically from $e_2$.

think that thick explications of compositionality are false.[16] But note that for thin explications of compositionality these cases do not pose a problem. A theory in accordance with $C^{FI}$ for instance could include identity functions such as '$\mu_{(\text{XP}} X Y) = \mu(Y)$' to include expressions featuring expletives in the theory. In rules like these the semantic contribution of the expletives is simply ignored. Even theories in accordance with $C^{BU}$ can be in accordance with cases like (28) and (29) since the meanings of complex expressions can still be built up out of meanings of the non-expletive parts.

Consider now expression such (30) and (31). These are cases where complex expressions seem to be meaningless. Ideas are not the kind of things that can be colored, so it does not make sense to hold that they are green; stones are not the kinds of thing that have mental states, so it does not make sense to hold that a particular stone is thinking of something. Hence sentences like (30) and (31) seem to be meaningless. And even if one were to dispute these particular cases the problem seems to be very general; some predicates cannot apply to certain kinds of things, and attempting to apply these predicates to things of the relevant kinds result in nonsense.

Note that it does not seem plausible to maintain that these sentences are not meaningless but only necessarily false. For the negations (in a syntactic sense) of these sentences are not tautologies but lack truth-values (just like the original sentences seems to do).

Cases like (30) and (31) conflict with thick explications of compositionality to the same extent as cases like (28) and (29). In addition, cases like (30) and (31) conflict with (thin and thick versions of) $C^{BU}$. If a complex expression has meaningful parts, $C^{BU}$ mandates that the meaning of that expression has certain meanings as parts. But if that expression is meaningless, then there is nothing that can have these parts. Thin versions of $C^{FE}$ and $C^{FI}$ on the other hand are fully compatible with cases such as (30) and (31). If all expressions of a certain mode of combination are meaningless, then omitting, in a theory compatible with $C^{FI}$, an implicit semantic rule for this mode of composition will result in all the relevant expressions being

---

[16] Note that on the explication of compositionality offered by Hodges (2001) (which we discussed in chapter two) some grammatical terms could be meaningless. But it is incompatible with cases such as (28) and (29) where a part of a meaningful term is meaningless.

meaningless. And as long the domain of meanings for a certain semantic theory contains a null element (which represents meaninglessness) there is no problem with devising implicit semantic rules that result in meaninglessness for some restricted class of cases. We can thus sum up this subsection by saying that due to the fact that natural languages contain meaningless expressions we have a reason to think that thick explications of $C$ as well as $C^{BU}$ are false. But they give us no reason to think that thin $C^{FE}$ or thin $C^{FI}$ are false.

### 5.1.4 Interactionism

The principle of compositionality is sometimes associated with $I$; the principle of the context-independence of meanings.[17]

(I)        Expressions mean the same in all linguistic contexts.

The antithesis to $I$, i.e. the thesis that words mean different things in different linguistic contexts, is sometimes referred to as *interactionism*. Cases like the following seem to constitute compelling evidence in favor of interactionism, and thus serve as counterexamples to $I$.

(44a)  Four lions occupy the corners of Trafalgar Square
(44b)  Four stone lions occupy the corners of Trafalgar Square
(45a)  Most students here drop geography in their final year
(45b)  Most students here drop geography lectures in their final year
(46a)  Jane is a good philosopher
(46b)  Jane is a good dancer
(47a)  A red book is lying on the table
(47b)  A red newspaper is lying on the table

In (44a) 'lions' seems to mean 'lions', but in (44b) it seems to mean 'representations of lions'; in (45a) 'drop' seems to mean 'drop studying', but in (45b) it seems to mean 'drop attending'; in (46a) 'good' seems to mean, roughly, 'good for a philosopher', but in (46b) it seems to mean, roughly, 'good for a dancer', and in (47a) 'red' seems to mean 'red on the outside', but in (47b) it seems to mean 'red all the way through'. From these observations it follows that if

---

[17] The principles are equated by Lahev (1989) for instance.

compositionality entails *I*, then examples such as (44) - (47) are counterexamples to compositionality.[18]

The strength of the connection between *I* and *C* depends on how *C* is explicated. According to $C^{BU}$ the meaning of a complex expression contains the meanings of its parts as parts. This means that if $e_1$ and $e_2$ are both expressions having *e'* as an immediate part, $\mu(e_1)$ and $\mu(e_2)$ have a common part (i.e. $\mu(e')$). It seems very natural to identify this part with what *e'* means *in* $e_1$ and $e_2$. But, given this, $C^{BU}$ entails *I*. Thus, the counterexamples to *I* are counterexamples to $C^{BU}$ and they thus give us a reason to think that $C^{BU}$ is false. [19]

On the other hand, weaker explications of compositionality do not entail *I*. $C^{FE}$ does not require that there is any pattern across the meanings of all complex expressions $e_1 \ldots e_n$ which contain a certain expression *e'*, such that one can deduce from it what *e'* means *in* $e_1 \ldots e_n$. According to $C^{FE}$, no two complex expressions can have different meanings if they have parts with the same meanings and have the same mode of composition. But the examples (44) - (47) gives us no reason to think that there are such expressions because even though, for instance, (47a) and (47b) have different meanings,

---

[18] Some of these cases have been used as arguments against compositionality by Lahav (1989). Cohen (1986) makes use of some of these cases to throw doubt on something he calls insulationalism which is supposed to be inherent in, for instance, Davidsonian semantics. He is not explicitly committed to the position that these examples are counterexamples to compositionality.

[19] The defender of $C^{BU}$ might argue that, for instance, 'drop' does not mean something different in, for instance, 'Most students here drop geography in their final year' and 'Most students drop geography lectures in their final year'. Abstracting away from particular ways of dropping something she might say that what remains (something like 'quit') is the meaning of drop (in addition to the other things which are also meanings of drop like that corresponding to the noun 'drop') and this might be constant across all expressions featuring 'drop'. However, Cohen remarks that 'drop' seems to behave differently with respect to its subcategories than does other words. Consider (1) and (2).

(a)      Mary dropped geography and Jane did too.
(b)      Mary killed police and Jane did too.

In the first case, it is mandated that Mary and Jane did the same kind of dropping, i.e. dropped studying, whereas in the second case it is not mandated that they did the same kind of killing, e.g. (a) is still felicitous in a context where Mary strangles police to death and Jane stabs them to death. This asymmetry between (a) and (b) seems to indicate that 'dropped' really means something more specific in (b) because it is that which is available for anaphoric reference.

they do not have parts with the same meanings ('book' does not mean the same as 'newspaper'). So the counterexamples to *I* give us no reasons to think that $C^{FE}$ is false.

$C^{FI}$ might not be in conflict with these examples either. To see this, cases of interaction need to be divided into *systematic* and *unsystematic* cases of interaction since the reconcilation of $C^{FI}$ with interactionism might go along either of two routes.

Jonathan Cohen, who draws attention to cases like the aforementioned in his paper 'How is Conceptual innovation possible?' maintains, in connection with one case of interaction that he considers – the fregean idea that what words refer to in indirect speech differs from what they normally refer to – that 'the sense of an expression is being changed by its linguistic context in accordance with a *systematic rule* that is *inherent in any synchronic language-state.*'(Cohen 1986: 230, emphasis added). Other remarks in his paper seem to suggest that he thinks that the kinds of interactions he considers are generally systematic. This seems reasonable for cases like (44) - (45). It seems, for instance, that the change in meaning from 'lion' to 'representation of lion' is mirrored in a range of similar construction; 'lion' can be substituted for another kind of animal or plant, and 'stone' can be substituted for a range of different materials, 'brick', 'glass', 'metal' etc. and the 'representation of *x*' reading persist. Similarly, the meaning of 'drop' seems to co-vary in a predictable way with the meaning of the complement noun phrase; in 'Most students here drop geography lectures in their final year', 'drop' seems to mean 'drop attending', in 'Most students here drop geography lectures reading assignments in their final year', 'drop' seems to mean 'dropped executing', in 'Most students here drop geography lectures reading assignments library-fees in their final year', 'drop' seems to mean 'drop paying' etc. If these appearances are correct, then it is plausible to hold that what these interactions amount to are special kinds of composition. In these cases, what *x* means in *y* seems to be posterior to composition; 'lion' means what it does in 'stone lion' because 'stone lion' means what it does, 'drop' means what it does in 'drop geography classes' because 'drop geography class' means what it does. If this is correct, cases like (44) – (45) do pose a challenge to (the proponent of) $C^{FI}$, but not by exhibiting some feature recalcitrant to composition, but by displaying what composition must be like. It

is then up to the compositionalist to try to capture the systematicities inherent in these forms of interaction in her composition rules.[20]

   Others that have put forth cases such as (44) – (47) have instead emphasized how *unsystematic* interaction is. Ran Lahav, for instance, who is primarily concerned with adjective noun combinations emphasize this.

> [A]djectives have a non-compositional semantics, in the sense that their applicability conditions (and thus their semantic contribution to the expression in which they are embedded) varies from one linguistic context to another in a way *that cannot be analyzed in terms of a general (not vacuously disjunctive) rule or function*
>
> (Lahav 1989: 266, emphasis added)

The following passage illustrates the observed variation that supports his conclusion.

> Consider the adjective 'red'. What it is for a bird to count as red is not the same as what it is for other kinds of objects to count as red. For a bird to be red (in the normal case), it should have most of the surface of its body red, though not its beak, legs, eyes, and of course its inner organs. Furthermore, the red color should be the bird's natural color, since we normally regard a bird as being 'really' red even if it is painted white all over. A kitchen table, on the other hand, is red even if it is only painted red, and even if its 'natural' color underneath the paint is, say, white. Moreover, for a table to be red only its upper surface needs to be red, but not necessarily its legs and its bottom surface. Similarly, a red apple, as Quine pointed out, needs to be red only on the outside, but a red hat needs to be red only in its external upper surface, a red crystal is red both inside and outside, and a red watermelon is red

---

[20] Notice that it is important to distinguish between *a word making the same semantic contribution* to the expressions it is featured in, and that expression *meaning* the same thing in all these expressions. Say that $e_1$ and $e_2$ are both expressions having $e'$ as an immediate part. Whereas a principle such as $C^{+I}$ is committed to the claim that if the meaning of $e_1$ and $e_2$ both depend on the meaning of $e'$, then they both depend on the same thing, i.e. $e'$ makes the same semantic contribution to both expressions. However, it does not follow that $C^{+I}$ is committed to $e'$ meaning the same thing in $e_1$ and $e_2$. This can be seen clearly by noticing that whereas making the same semantic contribution to the meaning of an expression is something which is true or false prior to composition, meaning the same thing in an expression might, in some cases, be something that is posterior to composition.

only inside. For a book to be red is for its cover but not
necessarily for its inner pages to be mostly red, while for a
newspaper to be red is for all of its pages to be red. For a
house to be red is for its outside walls, but not necessarily its
roof (and windows and door) to be mostly red, while a red car
must be red in its external surface including its roof (but not
its windows, wheels, bumper, etc.). A red star only needs to
appear red from the earth, a red glaze needs to be red only
after it is fired, and a red mist or a red powder are red not
simply inside or outside. A red pen need not even have any
red part (the ink may turn red only when in contact with the
paper). In short, what counts for one type of thing to be red is
not what counts for another.

(Lahav 1989: 264)

Similar remarks can be given with respect to other adjectives. There
seems to be little in common, for instance, between a good
philosopher and a good knife.

Given that 'red' seems to mean very different things when
combined with different nouns, and that this variability is not
predictable from the nouns it combines with, then the previous way
to reconcile interactionism does not seem possible. Interaction can
not only be due to specific forms of composition.

However, another treatment suggested by Zoltán Szabó (for cases
such as those involving 'red' and 'good') seems attractive.[21] According
to Szabó the logical form corresponding to the relevant kinds of
adjectives are of the following kind

(46a)  Jane is a good philosopher.
(46a')  (good($R$))(Jane) $\wedge$ philosopher(Jane)

(47a)  A red book is on the table.
(47a')  $\exists x$, (red($C$, $P$))($x$) $\wedge$ book($x$) $\wedge$ on-the-table($x$)

$R$, $C$, and $P$ are variables ranging over roles, contrast classes and parts
of objects respectively. The expression '(good($R$))(Jane)' is supposed
to be understood as Jane being good *with respect to a role R*, where a
role is a technical term which range over ways in which an agent can
be good, i.e. being a philosopher, being a painter, being a fire-fighter

---

[21] See Szabó (2001).

etc. The idea is that '$R$' is contextually assigned a value which then influences the truth value of the relevant utterance of the expression. Similarly, the expression $(\text{red}(C, P))(x)$ is supposed to be understood as $x$ being red with respect to a contrast class $C$ and a part $P$ of $x$, where the contrast class is used to determine to what degree an object needs to be red in order to be counted as red (a person needs to have a less red face to count as having a red face than a traffic light needs to be red in order for it to be counted as red), and the relevant part is the surface that needs to be red for the object to be counted as red. Again, '$C$' and '$P$' are contextually assigned values which then influences the truth value of the relevant utterance of the expression. This accounts for the unsystematic (from the perspective of the meaning of the noun of the combinations) nature of the interactions.

That the interpretation of at least some of the relevant words (e.g. 'red') really are context dependent is suggested by the fact that, for instance, the ways in which an object must be colored in order for the relevant complex expression to apply, are really just defaults. One can easily imagine contexts where these are shifted. A juggler might have painted the exterior surfaces of watermelons (with red interiors) in different colors, say, red, blue and white, in order to make his watermelon juggling even more striking. When suggesting to his audience that he will at a certain point add a red watermelon to the blue and white watermelons he is currently juggling in order to make things more colorful, it seems clear that his utterance is *only* true, if he adds a watermelon which is *painted* red.

We could make use of Szabó's insights in either of two ways. We could either follow him quite closely and assume that there are variables in the underlying structure of the relevant expressions, or we can assume that the meaning of, for instance, 'red' is itself context dependent and depends on relevant contrast classes and salient parts for its meaning. On either proposal the interactions are reduced to the kind of context dependence displayed by overtly context dependent expressions such as 'I', 'him' and 'now' which are unproblematic from the perspective of a semantic theory in accordance with $C^{FI}$.

To sum up, even though $C^{BU}$ is incompatible with interactionism it seems that weaker explications of $C$ such as $C^{FI}$ are compatible with both systematic and unsystematic interactionism.

## 5.1.5 Context-Sensitivity

It has been maintained that some complex expressions have meanings that depend on contextual factors in a way that cannot be traced to a dependence of the meaning of their parts. Consider, for instance, the following example.

(48)    It's raining.

Whether or not an utterance of (48) is true or false depends on whether it's raining at a location which is salient in the context of utterance. But in contrast with 'It's raining *here*' no overt part of (48) seems to be responsible for this dependence on a salient location. This is usually expressed by saying that the proposition expressed by an utterance of (48) contains an *unarticulated constituent*.[22] If this is the case there is a divergence in the pattern of propositions expressed by utterances of (48) that cannot be traced to the semantic values of utterances of parts of (48) which means that the meaning of the complex expression (which is supposed to sum up the pattern of significance across utterances), is not completely predictable from the meanings of its parts and its mode of composition. The existence of unarticulated constituents is thus *prima facie* in conflict with compositionality.[23]

It is important to distinguish the aforementioned kind of context sensitivity from forms of context sensitivity that are compatible with compositionality. Consider the following illustration; according to a theory *t*, the meaning of 'red' is a function *r* from contexts to the set of things that are red there (given a contrast class and relevant parts along the lines of the proposal of the previous subsection), the meaning of 'I' is a function *i* from contexts to the speaker in that context and the meaning of 'I'm red' is a function *g* from contexts to truth values.[24] The rule $\mu(_S\ NP\ VP) = f(\mu(NP), \mu(VP))$ where $f(x, y) =$

---

[22] See Perry ([1986] 2000).

[23] Note that if the pattern of meanings across utterances of a complex expression is not predictable from the patterns of meanings across utterances of parts of those expressions and the assigned modes of composition, then the meaning of that expression is not predictable from the meaning of its part and mode of composition.

One can interpret Fodor (2001) as advancing this kind of argument against the compositionality of language.

[24] Cfr. Kaplan ([1977] 1989).

$g$ where $g$ is such that $g(c)$=true iff $x(c) \in y(c)$, is clearly a composition rule in the sense of $C^{FI}$ (it only makes reference to meanings of parts and it assigns meanings via a mode of composition variable). Hence, a theory can be in accordance with $C^{FI}$ even though the meanings of *the parts of complex expressions* that are invoked by the theory are context dependent and even though the meanings of *complex expressions* that are invoked by the theory are context dependent.[25]

If we generalize from this observation we can conclude that quite radical forms of context sensitivity are compatible with compositionality. Francois Recanati has argued that *modulation*, i.e. optional unconscious pragmatic processes operating on the meanings of sub-sentential utterances can influence what is said by an utterance of an expression.[26] Modulation includes the processes *free enrichment*, *loosening* and *transfer*.[27] It seems that all of these processes are compatible with compositionality.

Consider (49) and (50), which can be used to illustrate *free enrichment*.

(49)  Mary took out her key and opened the door.
(50)  He wears rabbit.

It is natural to understand the second conjunct of an utterance of (49) along the lines of 'opened the door *with the key*' although this goes beyond what is explicitly expressed by (49). This enrichment of 'opened' seems to occur due to the key being mentioned in the first conjunct. Similarly, it is natural to understand an utterance of (50) along the lines of 'He wears rabbit *fur*' although this goes beyond what is explicitly expressed by (50). This enrichment of 'rabbit' seems to occur due to knowledge about clothing customs.

What is said by an utterance can also be influenced by *loosening*. Consider (51).

---

[25] The view expressed here is also endorsed by Szabó (2001).

[26] That these processes are *optional* just means that they differ from the saturation of overtly context-sensitive expressions such as 'I', processes that are required to take place to get an interpretation at all. Recanati (2004) endorses the view that modulation can influence the truth conditional content of an utterance

[27] The cases of 'unsystematic' interaction considered in the previous subsections are plausibly treated as special cases of modulation. See Recanati (2004).

(51)   The ATM swallowed my card.

Here, an utterance of (51) could well be true even though ATM's cannot literally swallow anything due to them not having the appropriate bodily equipment for swallowing. But the conditions of applications for 'swallow' are loosened in many utterances of (51) so that the predicate does apply.

Consider finally (52) and (53) which are examples of *transfer*.

(52)   The ham sandwich left without paying.
(53)   I'm parked out back.

In a context where a person who has ordered a ham sandwich is contextually salient, it is quite natural to interpret an utterance of (52) along the lines of 'The person who ordered the ham sandwich left without paying'. The referent of 'the ham sandwich' is thus transferred to the referent of 'the person who ordered the ham sandwich'. Similarly, an utterance of (53) is quite naturally interpreted along the lines of 'My car is parked out back'. The reference to the speaker of 'I' is thus transferred to the reference of 'my car'.[28]

With all these cases of modulation it is the case that for each utterance of the expressions that intuitively has a meaning that is influenced by one of the pragmatic processes, one of the parts of those utterances also have a meaning that is so influenced. So the context sensitivity of the complex expression is traceable to the context sensitivity of one of its parts. Hence, it seems that free enrichment, loosening and transfer are not in conflict with compositionality. This rests on the fact that these primary pragmatic processes quite generally preserve semantic categories and structural roles. 'The ham sandwich' and 'The person who ordered the ham sandwich' are both noun phrases and are likely to combine semantically in the same way with the verb phrase 'left without paying'. This means that it is reasonable to think that the composition rules of the semantic theory

---

[28] There are many other examples of transfer as well. Many phenomena connected with expressions containing quotation marks can be treated as cases of transfer as well. See Jönsson (2007) for one such account.

can be appropriately adopted even after the primary pragmatic processes have been applied.[29,]

Not all cases of context sensitivity fit the previous mold. A sentence such (48) seems to be different. In order to get a better grip on sentences like (48) we can consider another example.

Some have argued that utterances of belief reports such as (54) also involve unarticulated constituents.

(54)  John believes that Hesperus can be seen in the morning.

Mark Crimmins and John Perry have suggested the following concerning expressions such as (54).[30]

> When we report beliefs, there is always some further condition that a belief with the specified content is claimed to meet. The belief report is true, only if a belief *meeting that further condition* has the right content […] this additional requirement is part of the proposition expressed by the belief report. Thus, it is a condition on the truth, not merely the felicity, of the report […] The general solution to the puzzles [about belief reports] is to allow a condition on particular beliefs, over and above a content condition, to be part of the claim made. The version of this strategy we shall pursue here is to take this further condition always to be a specification of the notions that are supposed to be involved in the ascribed belief.
>   We shall say that a notion that a belief report is about is an *unarticulated constituent* of the content of the report-it is a propositional constituent that is not explicitly mentioned.
>              (Crimmins and Perry 1989: 697, emphasis added)

The idea seems to be that belief reports are judged true or false not only on the basis of whether they involve attributions of the right content, but also on the basis of whether they attribute the right *notions*. On Crimmins and Perry's view, *notions*, as well as *ideas*, are constituents of beliefs, and correspond, respectively, to ways of thinking of individuals, and to ways of thinking about properties.

---

[29] This point is noted by Pagin and Pelletier (2007: 32-33). Pagin and Pelletier also show how widely modulation can account for context sensitivity, and thus how important it is that compositionality is compatible with modulation.
[30] See Crimmins and Perry (1989).

Beliefs, ideas and notions are cognitive particulars that can stand in causal relationships to perceptions and actions, belong to an agent, come into existence, endure, and go out of existence. Ideas and notions have contents, but unlike beliefs these contents are not propositions, but properties and individuals respectively. Different notions have different circumstances of origin and are associated with different ideas depending on which beliefs they are constituents of. Crimmins and Perry's main proposal is that when a belief report is uttered, the utterance is evaluated for its truth in part by involving an attribution of the correct notion.[31] Which notion that is relevant is not expressed by any of the syntactic parts of the belief report but is supposed to be contextually given. Consider the following illustration in terms of the two belief reports (54) and (55).

(54)    John believes that Hesperus can be seen in the morning.
(55)    John believes that Phosphorus can be seen in the morning.

Utterances of (54) and (55) can have different truth values even though uttered in identical contexts. This is explained on Crimmins and Perry's account by assuming that even though they ascribe beliefs with the same content to John, they might attribute John's belief with different constituents. The use of 'Hesperus' in one sentence and 'Phosphorus' in another context might signal that two different attributions are made, depending on, for instance, how John reports his own beliefs, known circumstances of the formation of his belief etc.[32]

---

[31] Crimmins and Perry do not assume that all belief reports are of this kind and the way they assume that notions can figure in belief reports is more complicated than I let on in the main text. Notions can also be *constrained* rather than *provided by* context. But I do not think anything in my overall argument hinges on my simplification of their position in the main text. The consequence for compositionality by the involvement of notions in the truth conditions of utterances of belief reports is the same regardless of whether notions are provided by the context or constrained by the context.

[32] Crimmins and Perry offer the following remarks to further clarify how notions are picked out.

> In the case of belief reports, in which notions are unarticulated, we do have rough and ready ways to clarify just which notions we mean to talk about […] we specify how [the believer] would or would not "put" his belief. Or we allude to the evidence which led [the believer]

The crucial thing in the present contexts is that what notions that are attributed to a believer as part of the utterance of a belief report is not determined by the meanings of the parts of that utterance or its mode of composition. This means that even $C^{FE}$ might be violated by the existence of unarticulated constituents since Crimmins and Perry's claims about belief reports suggest that there might be two expressions with different meanings even though they have parts with the same meanings and the same mode of composition.

It might be compelling to think that cases like (48) and (54) really are counterexamples to compositionality but this rests on two different assumptions that one might well deny. First, the person who advances these cases against the compositionalist must assume that the relevant context dependence really is not traceable to constituents of the relevant expressions. This is controversial. Jason Stanley has argued, for instance, that 'all effects of extra-linguistic context on the truth-condition of an assertion are traceable to elements in the actual syntactic structure of the sentence uttered.' (2000): 391). According to Stanley, these elements correspond to variables that are not articulated in the surface syntax, but which are present in the underlying structure. The existence of these variables is supported by the following argument.

($P_1$)   Operators in a sentence only interact with variables in the sentence that lie within their scope.

($P_2$)   If a constituent is unarticulated, it is not the value of any variable in the sentence.

(C)   If a supposed unarticulated constituent of the proposition expressed by a certain utterance can be bound by an operator in the sentence, then that constituent is not unarticulated

---

to form the belief, or to the actions it would be likely to bring about. Each of these devices can succeed in distinguishing among the two notions which in context can seem equally relevant, thus eliminating possible confusion about which notion we mean to talk about.

(Ibid: 701)

Stanley then goes on to show that the antecedent of his conclusion is satisfied for many of the supposed cases of unarticulated constituency. Consider (56) for instance.

(56)    Everywhere I go, it rains.

An utterance of (56) expresses the proposition that it rains at all the locations where the speaker goes (at the time the speaker goes there). This suggests that 'Everywhere' binds a location variable in 'it rains'. But if there is such a variable, then 'it rains' is not a case of unarticulated constituency.[33] To the extent that this is possible in general, compositionality can be salvaged since cases of supposed unarticulated constituency will be no harder to square with compositionality than overtly context-dependent expressions such as 'She is dangerous'.[34]

The second assumption made by the person advancing cases such as (48) and (54) against the compositionalist is that she assumes that the meaning of an utterance is identified with, or determines, the proposition expressed by that utterance. This could also plausibly be denied since it seems that the composition pertaining to some other form of significance has to take place prior to the determination of the unarticulated constituents anyway. Both utterances of (48) and utterances of (54) are such that the unarticulated constituent depends on the meaning of the rest of expression. Determining which location an utterance of (48) is about, i.e. fixing the unarticulated constituent of that utterance, might be affected by the kind of weather being reported. Imagine that you have a friend $T$ that has just called you and uttered (48). Imagine further that you know that $T$ has two homes $A$ and $B$, and that $A$ is located in an area where it currently might snow but where it never rains at this time of year, and $B$ is located in an area where it currently might rain but where it never

---

[33] Pagin (2004) offers another proposal that salvages compositionality without positing a location variable in 'it rains' but on a contextual dependence in 'rains'.

[34] Stanley's proposal is not uncontested. See Recanati (2002) for criticism. The issue is much too complicated to be dealt with satisfactory here. In order to settle if the unarticulated constituent analyses or the covert variable analyses are to be preferred global properties (such as simplicity or generality) of larger theories needs to be considered. In addition, Stanley does not discuss cases like (54) but since these are much more theoretically laden than (48) and there are several other theories about the semantics of belief reports (see section 5.2.6) these cases seem less pressing.

snows at this time of year. You know that *T* calls you from one of her homes, but it unknown to you from which. It seems to me that in order to determine the unarticulated location constituent of the proposition expressed, prior composition needs to take place (the meaning of 'rain' needs to be combined with the meaning of the present tense inflection on the verb). The case with (54) seems to be similar. In order to determine what notions that are attributed by an utterance of (54) we need to know which belief we are concerned with and we learn this by composing the meaning of the that clause of the utterance. So if the semantic notion that we are interested in with respect to utterances, is not what is said by that utterance, but some other meaning entity, then the unarticulated constituent examples are not counterexamples to the compositionality principles.[35]

   To sum up this section: not all kinds of context sensitivity is problematic from the perspective of compositionality. Cases of supposedly unarticulated constituents, which might be problematic for compositionality can be made compatible with compositionality either by finding covert variables revealing that the constituents are really articulated, or by recourse to utterance meanings other than truth-conditional content.

### 5.1.6 Propositional Attitude Reports

Some have maintained that there are general problems with combining *C* with an adequate semantic account of propositional attitude reports. Jeff Pelletier, for instance, has claimed that the following four assumptions lead to an inconsistency if combined with the principle of compositionality.[36]

(P$_1$)   There are synonymous sentences.
(P$_2$)   There is only one mode of composition for sentences
          of the surface form:
             name-of-believer·'believes'·'that'·sentence

---

[35] Bach (1994) makes use of the expression 'propositional radical' to denote one possible such meaning entity whera a propositional radical is an incomplete proposition.
[36] See Pelletier (1994b; 2004).

(P₃)    If ϕ and ψ are synonymous sentences, then they have
        the same truth value with respect to the same contexts of
        utterance.

(P₄)    If ϕ and ψ have different forms, then there is a context
        of utterance such that exactly one of (i) and (ii) is true:
          i) *B* believes that ϕ.
          ii) *B* believes that ψ.[37]

Given (P₁), there are synonymous sentences. Assume that (57) and
(58) are two of these.

(57)    Dogs are dangerous.
(58)    Canines are dangerous.

Given (P₂), (59) and (60) have the same mode of combination.

(59)    Kim believes that dogs are dangerous.
(60)    Kim believes that canines are dangerous.

Given compositionality, the fact that (57) and (58) have the same
meaning, and that 'Kim', 'believes' and 'that' means the same thing in
(59) and (60) leads to (59) and (60) having the same meaning. Given
(P₃), (59) and (60) have the same truth value with respect to the same
contexts of utterance. But this contradicts (P₄) since then there is no
context of utterance such that exactly one of (59) and (60) is true.
Hence, it seems that if we want to retain (P₁) – (P₄) we must give up
compositionality.[38]

   We should note first that the notion of compositionality at work
here is a very weak one, something along the lines of $C^{FE}$. Since $C^{FE}$ is
so weak it seems that we must give up compositionality entirely in
order to retain (P₁) – (P₄).

   But counter to Pelletier's claim, there is a consistent account of
propositional attitude reports compatible with (P₁) – (P₄) and with

---

[37] I will refer to a version of this thesis as 'The Mates Thesis' in the following chapter.
I avoid doing so here since it would obscure the exposition.

[38] The argument has been modified slightly. Pelletier's original first premise is in
terms of synonymous *sentences* having the same truth value. Since I think that
sentences have truth values only relative contexts of utterances (P₁) and (P₃) have been
adjusted accordingly. Pelletier (1994b: 313, fn. 4) condones such a reinterpretation.

$C^{FI}$: the paratactic theory of propositional attitude reports.[39] According to a simple variant of this view (59) should be broken down into two parts. Only the first of these is asserted when (59) is uttered.

(59')   Kim believes that. Dogs are dangerous.

We assume that 'that' in (59') acts as a genuine demonstrative and the second sentence is merely uttered to demonstrate (exhibit) a certain form and a certain meaning. Hence, on this view, 'that' refers to something different in an utterance of (59) than one of (60). This account is obviously compositional, since the belief reports are assumed to be of the same syntactic-semantic kind as constructions such as 'Kim loves him'. In addition, it is compatible with $(P_1)$, since there can be synonymous sentences, and we can again assume that (57) and (58) are synonymous. It is compatible with $(P_2)$ since there is only one mode of composition for sentences of the surface form 'name-of-believer·"believes"·"that"·sentence' (i.e. the form $[_S \, [_{NP} \, [_N$ name-of-believer]] $[_{VP}[_V$ believes] $[_{NP} \, [_N$ that]]]]). It is compatible with $(P_3)$ since synonymous sentences have the same truth values with respect to the same context of utterance. And it is compatible with $(P_4)$ since given that expressions (57) and (58) have different forms there are contexts of utterance such that exactly one of (59) and (60) is true. Hence, Pelletier's argument fails. Compositionality (explicated in terms of $C^{FI}$) is compatible with assumptions $(P_1) – (P_4)$.[40, 41]

### 5.1.7 Quantified Conditionals

James Higginbotham has suggested that the semantics of certain conditionals pose problems for compositionality.[42] In his paper

---

[39] This view has been developed from specific suggestions made by Davidson ([1968] 2001) on indirect speech reports. The underlying idea dates back to Carnap (1947).

[40] No analysis of belief reports is uncontroversial however and the paratactic account is no exception. See Schiffer (1987) for instance. See Lepore and Loewer (1989) for a defense.

[41] Another way out of the initial inconsistency is just to deny $(P_3)$ since we saw in Subsection 5.2.5 that we can avoid the problems with unarticulated constituents by doing so as well.

[42] See Higginbotham (1986; 2003). In his earlier paper Higginbotham really argues that natural languages do not satisfy what he calls the indifference principle. But in his later paper, he equates this principle with compositionality, 'narrowly construed'.

'Linguistic Theory and Davidson's Program in Semantics' he considers conditionals containing 'if' or 'unless' that are embedded under quantifiers, and view these as potential counterexamples to compositionality. He uses the following four examples.

(61)   Every student will succeed unless he goofs off.
(62)   No student will succeed unless he goofs off.
(63)   Every student will succeed if he works hard.
(64)   No student will succeed if he works hard.

Assume that (61) and (62) can be represented with something like the following structure

(61') $[_{NP}$ Every student $x_1]$ $[_{VP}$ $[x_1$ will succeed] unless [he$_1$ goofs off.]]
(62') $[_{NP}$ No student $x_1]$ $[_{VP}$ $[x_1$ will succeed] unless [he$_1$ goofs off.]]

In providing a compositional semantics for sentences like (61) and (62) we need to assign some semantic contribution to 'unless'. But the problem for the compositionalist, Higginbotham suggests, is that there is no suitable truth function that can be assigned to 'unless' that will give these sentences their intuitive truth conditions. In (61) it seems appropriate to use inclusive disjunction, but if this is used in (62) this sentence turns out to be false if there is a student who succeeds and goofs off. But this is incorrect, what (62) says is that there are no students who does not goof off and succeed. Conversely, in (62) it seems appropriate to assign 'unless' a truth function such that it yields true only if its first argument is true but the second false (something like '…and not…'). But if this is used in (61) this sentence turns out to be true only if every student succeeds. But this is clearly incorrect. Similar problems arise with respect to (63) and (64) and 'if'. So these constructions constitute a *prima facie* problem with giving a compositional semantics.[43]

Notice first that the problem Higginbotham draws attention to is not a problem for $C^{FE}$. Whatever meaning that is assigned to 'unless'

---

[43] Hintikka (1980: 43) offers similar but simpler examples of the same type of phenomena. So 'any' in 'Chris will beat any opponent' and 'Chris will not beat any opponent' does not seem to mean the same thing. In the first sentence it seems to mean 'every' and in the second 'a single'.

and 'if', neither (61) and (62), nor (63) and (64) jointly constitute counterexamples to $C^{FE}$ since for neither pair it is true that the two complex expressions have different meanings even though they have parts with the same meanings and the same mode of composition, since the two expressions do not have parts with the same meanings (the quantifiers 'all' and 'no' have different meanings).

The problem might be a problem for $C^{FI}$. Higginbotham concludes from his examples that 'the interpretation of [a] subordinate phrase [sometimes] depends on information not locally available...' (Higginbotham 2003: 181). If this is the case, then (61) - (64) do pose a problem for $C^{FI}$. For according to $C^{FI}$, only the meanings of the *immediate* parts of a complex expression can be cited in the composition rule corresponding to the mode of composition of that expression. Higginbotham's proposal amounts to saying, for instance, that the meaning of the VP 'will succeed unless he goofs off' in (61) depend on the presence of 'every' in the sister NP.[44] So if Higginbotham's conclusions about cases like (61) - (64) are correct then we have a reason to think that $C^{FI}$ is false.

But there are various points where one can resist Higginbotham's conclusions. One could for instance hold that the meaning of 'will succeed unless he goofs off' *is* determined by the meanings of its parts and its mode of combination, by assuming that a vague meaning is assigned to 'unless' (and 'if'). This vague meaning is then transmitted to the meaning of 'will succeed unless he goofs off' but once this meaning is combined with that of the subject noun phrase, the different quantifiers have different features that will make the vague meaning determinate ('no' for instance might have a 'negativity' feature that will transform the vague connective into 'and not').[45] This would make (61) - (64) in accord with $C^{FI}$. Despite what one might think of the naturalness or generality of this proposal, the moral is simple: it is possible to 'postpone' semantic effects from embedding constructions. It is not necessary that their influence is exerted immediately.

---

[44] If we reflect on the discussion of Subsection 5.1.4 we can see this phenomenon as a case of non-local interaction.
[45] This is one of the proposals offered by Pelletier (1994a).

One could also maintain that that 'unless' and 'if' are ambiguous.[46] The idea is that 'unless' and 'if' correspond to two different grammatical terms each with its own meaning (corresponding to the two different truth functions). This is sufficient by itself to preserve $C^{FI}$. But one might also want to explain why on this view (61), for instance, does not have two readings corresponding to the two terms. Jeff Pelletier does this in terms of feature agreement. The two lexical entries corresponding to 'unless', for instance, might vary in the feature [neg]. One has [+neg] and the other has [-neg]. Correspondingly different quantifiers have one of these features as well. If a feature mismatch occurs the relevant interpretation is blocked.[47]

Finally, one could sever the links to Davidsonian truth-conditional semantics that Higginbotham presuppose, and give an account of the meanings of (61) – (64) that is in part in terms of *modal properties*.[48] On such a view sentences such as (63) – (64) quantify the number of students actually having a certain modal property; that of *succeeding if he works hard*. So instead of treating the verb phrase as if it concerned some logically complex property that students are supposed to *actually* have, which applies to a person depending on whether or not, for instance, the student *actually* succeeds and works hard, it concerns the students modal properties, whether the student *if he works hard will succeed*. Whether the student actually works hard or succeeds is immaterial on this analysis (except it being ruled out that he cannot actually work hard and not succeed). On this reading, both (61) and (64) concern the same property and the original difficulty is dissolved. As Sarah-Jane Leslie has remarked, this proposal takes seriously that not only the actual agreement between individuals and the properties described by the conditional are relevant, but also that the counterfactual agreement is important. For instance, if there is a student who will not succeed not matter how hard he works, but knows this, and does not work hard as a result, then the he would not be a counterexample to the reading of (63) according to which the

---

[46] This is another proposal by Pelletier (1994a). It is also endorsed by Pagin (2004)).

[47] By adding this second part to his proposal I feel that Pelletier(1994a) avoids a critique in Higginbotham's original discussion that the ambiguity proposal 'merely label what we would like to explain' (Higginbotham 1986: 36).

[48] Solutions like these have been proposed by von Fintel and Iatridou (2002) and Leslie (2007).

'if'-clause corresponds to the material conditional, since the antecedent of the conditional is false. But intuitively under these circumstances (63) is false since if the student would work hard he would still fail.[49] More importantly, by adopting this proposal we can make (61) – (64) compatible with $C^{FI}$. This result exemplifies this general rule: if we restrict the semantic options available it will be harder to make the distinctions that we need in order to get a compositional semantics. This should not lead us to think in these circumstances that meaning is not compositional, but rather that our semantic notions might not be powerful enough.[50]

## 5.2 Concluding Remarks

The more convincing of the previous arguments against compositionality pertained to the thick notions of compositionality or to the $C^{BU}$-explication of compositionality. The existence of essentially ambiguous expressions and the existence of meaningless expressions give us very good reasons to think that thick explications of compositionality are false. And the existence of meaningless expressions with meaningful parts and the existence of semantic interaction between the words of a phrase give us very good reasons to think that $C^{BU}$-explications of compositionality are false.

Among the arguments that could be construed as arguments against weaker compositionality explications (such as thin $C^{FE}$ and thin $C^{FI}$), matters were much less convincing. It seems that thin $C^{FI}$ (and thus thin $C^{FE}$) can be made compatible with i) the existence of idiomatic expressions – since these are often possible to treat as syntactic simple; ii) the existence of essential ambiguity and iii) the existence of meaningless expressions – since thin $C^{FI}$ does not make any claims about the number of meanings that an expression can have; iv) interactionism – since interactionism either amounts to special cases of composition or a form of context sensitivity that squares with $C^{FI}$; v) supposed cases of unarticulated constituents – since there is room to deny that unarticulated constituents exist, and even if they do exist

---

[49] This way to spell out of this idea is much too simplistic. But space prohibits a more detailed discussion. See Leslie (2007) for a more satisfactory proposal in the same vein.

[50] This is also the general moral reached by Janssen (1997: 421-425) in his dicussion of a range of putative problem cases for compositionality.

theories in accordance with $C^{FI}$ are only in conflict with the existence of unarticulated constituents when the meanings assigned to utterances determine the propositions expressed; vi) giving an adequate semantics of belief reports – since the paratactic analysis of belief reports squares with the suggested desiderata for such reports; and vii) the semantics of quantified conditionals – since these could, for instance, be interpreted as involving quantification over individuals having modal properties.

I do not want to deny that there is tension between the aforementioned phenomena and thin $C^{FI}$ (or even thin $C^{FE}$) or that some of the concessions that one has to make in order to retain these principles in the light of these phenomena are significant. But it seems to me that whether semantic theories in the end should be compositional has to be settled with the construction and comparison between different particular semantic theories. It cannot be settled by attempting to provide examples that *cannot* be handled by a compositional (explicated in terms of $C^{FI}$) account since there does not seem to be any such cases. This is compatible with saying that once theories have been constructed that take into account several of these and other phenomena, the compositional ones might not be better than the alternatives. But the reason for this will involve properties of the particular theories such as their scope or simplicity. This points in the same direction as my argument in chapter three (in Subsection 3.1.5) that the ranking of explanations of $LC^U$ sometimes needs to be done at the level of particular theories and not at the level of the determination principles: theory comparison is the key to progress. My stance echoes a position taken by Barbara Partee concerning the place of compositionality within Montague grammar.

> The principle [of compositionality] is so deeply constitutive of Montague grammar and most of its close relatives that it must be considered to be a methodological principle: there is no way to test it without testing an entire theory in which it is embedded. So the claim that natural languages have a compositional semantics amounts to the claim that natural languages can be fruitfully described with a theory that includes compositionality as one of its principles (…it is not that […] empirical evidence is not relevant, it is just that it is whole theories that must be evaluated).
>
> (Partee 1997:22)

It seems to follow from this position that the strategy of providing isolated supposed counterexamples to compositionality will provide less compelling reasons to reject compositionality than the strategy of constructing non-compositional theories which are superior to their compositional alternatives.

# 6

# The Informative Potential of Compositionality

The goal of inquiry in this essay is to ascertain the extent to which the principle of compositionality (*C*) can be justifiably imposed as a constraint on semantic theories and thereby provide information about what meanings are.

(C)    The meaning of a complex expression is determined by the meaning of its parts and its mode of composition.

Chapter Three and Chapter Four revealed that the reasons usually given in support of the idea that a semantic theory should be compositional are not (at least not by themselves) reasons to suppose that a semantic theory must accord with *C specifically*. Instead they are reasons merely to think that a semantic theory should accord with *some member of a set of determination principles*. This means that *C* cannot be justifiably imposed on semantic theories – at least, not on the basis of these specific reasons – and *ipso facto* that we cannot learn anything about meanings by justifiably imposing *C* as a constraint on these theories.

The question of $C$'s informative *potential* – of what we *could* learn about meanings from $C$ if we had been justified in imposing it as a constraint on semantic theories – is not without interest, however. For if there is no difference between the informative potential of $C$ and the informative potential of what we might call $D$, the property of being in accordance with one of the relevant determination principles, then there is a sense in which we might still be able to learn about meanings from the former. We would not, then, be able to learn anything from $C$'s justifiable imposition as a constraint on semantic theories. The lessons would emerge from the fact that it coincides, in informative potential, with a constraint that we are justified in imposing. Additionally, $C$'s informative potential might be interesting if reasons surface for thinking that it, specifically, is true. Finally, it is of some intrinsic interest to explore exactly how constraints like the ones discussed can provide us information about meanings.

The present chapter therefore asks i) what the informative potential of $C$ is, and ii) how the informative potentials of $C$ and $D$ are related. Since i) and ii) depend on the way in which $C$ is explicated, these issues will be addressed repeatedly, explication by explication. In the subsections below the explications will be addressed in order of logical strength: first $C^{FE}$ (Subsection 6.1.1), then $C^{FI}$ (Subsection 6.1.2), and finally $C^{BU}$ (Subsection 6.1.3). It will be concluded (in Section 6.2), first, that all three explications have significant informative potential; and second, and unfortunately for the compositionalist, that the informative potential of $C^{FE}$, $C^{FI}$ and $C^{BU}$ goes far beyond that of $D$. As a methodological corollary it will also be concluded that the explications of $C$ have their informative potential in virtue of our knowledge of the way candidate meanings distribute over expressions, and, in the case of $C^{BU}$, in virtue of such knowledge together with our knowledge of the structural relations that hold between candidate meanings.

My conclusions clash with the spirit, if not the letter, of the passage quoted in Chapter One.

> Over the last few years, we have just about convinced ourselves that compositionality is the sovereign test for theories of lexical meaning. So hard is this test to pass, we think, that it filters out practically all of the theories of lexical

meaning that are current in either philosophy or cognitive science. Among the casualties are, for example, the theory that lexical meanings are statistical structures (like stereotypes); the theory that the meaning of a word is its use; the theory that knowing the meaning of (at least some) words requires having a recognitional capacity for (at least some) of the things that it applies to; and the theory that knowing the meaning of a words requires knowing criteria for applying it.

(Fodor and Lepore [2001] 2002: 43)

Had we obtained good reasons for thinking that the principle of compositionality is true, we would now have a very useful source of information about meanings, but we did not, so we do not.

## 6.1 The Determination Principles

A constraint on semantic theories will provide information about what meanings are in virtue of certain theories being incompatible with that constraint. Depending on the details of the constraint, the theories that violate it will often differ, and therefore different constraints are likely to provide different information. Three explications of compositionality – $C^{FE}$, $C^{FI}$ and $C^{BU}$ – occupied center stage in the previous chapters, and the present chapter will be primarily concerned with the informative potential of these. It will turn out that there are indeed significant differences in the informative potential of the three explications, but that, due to the entailment relations between them, stronger theses are at least as informative as weaker ones. The informative potential of $C^{BU}$ will be discussed here for the sake of completeness in spite of its explanatory redundancy (which we noted in Chapter Four) and despite the fact that it seems incompatible with some salient linguistic phenomena, including certain kinds of meaninglessness and interactionism (noted in Chapter Five). $C^{BU}$ is in any case of real interest, because its informative potential differs in kind from that of the other explications.

Note that the concept of informative potential is, as it were, coarser than that of logical strength. Two constraints of differing logical strength could be such that we would learn exactly the same things about the nature of meaning from them. It is thus an open question, prior to investigation, whether the three explications of $C$ actually

possess different informative potential, or whether their informative potential differs from $D$'s.

Since we have no reason to think that any of the explications of $C$, specifically, are true, we are not justified in requiring a semantic theory to accord with any of them. This means that what we could have learnt about meaning by imposing these explications of $C$ on semantic theories, had we had good reason to think they were true, might differ from what we can *actually* learn about meaning by imposing the constraint that we are justified in imposing. Since we have seen that theories according with several other determination principles such as $F^{FI}$, $X^{FI}$, and $K^{FI}$ can provide the same kinds of explanation that theories in accord with $C^{FI}$ can, the informative potential that is really interesting is that associated with the constraint $D$ – the constraint that a semantic theory should be in accordance with one of these principles.

> A semantic theory $t$ accords with $D$ iff $t$ accords with $C^{FI}$, $F^{FI}$, $X^{FI}$, or $K^{FI}$

Although the list of determination principles that we discussed in the previous chapters was not meant to be exhaustive, we can assume for present purposes that $D$ can be defined in this way. None of the conclusions to be drawn about its informative potential requires these principles alone to be included.

The following subsections will be structured in the following way. Each will begin by asking what, exactly, gives a certain explication of $C$ its informative potential. Examples of this potential will then be provided. A discussion of the difference, in informative potential, of this explication and that of $D$ will follow.

## 6.1.1 $C^{FE}$ vs. $D$

Although $C^{FE}$ is a very weak form of compositionality, it is still a significant, potentially informative, constraint on semantic theories.

A semantic theory accords with $C^{FE}$ iff

there is a function $r$ such that for every complex term $s = \sigma(e_0, \ldots, e_{n-1})$,

$$\mu'(s) = r(\sigma, \mu'(e_0), \ldots, \mu'(e_{n-1})).$$

Since $C^{FE}$ is precisely defined, it is a straightforward matter to determine whether a semantic theory accords with it or not. It does not follow directly, however, from a particular theory's not being in accord with $C^{FE}$, that meanings cannot be of the type that that theory presupposes that they are. Given a certain theory in violation of $C^{FE}$, according to which, say, meanings are extensions, we cannot conclude that meanings are not extensions. There may be other theories in which it is assumed that meanings are extensions, and these may accord with $C^{FE}$. Indeed, since $C^{FE}$ is a purely distributive principle, *whatever* meanings are assumed to be, there will be semantic theories, according with $C^{FE}$, that make use of meanings of that sort.

However, this does not mean that $C^{FE}$ could not provide information about meanings. For we can learn something from $C^{FE}$ – and the other constraints of interest to us – once we have information about the way in which candidate meanings distribute over expressions; and information of this kind is often available. It is sometimes given by intuition and is sometimes empirical. Thus we *intuit* what expressions refer to, or denote (e.g. that 'George W. Bush' and 'the president of the United States of America' currently co-refer); what the truth conditions of sentences are (that 'the sun is shining' is true if and only if the sun is shining); what the typical instances of general terms are (that a hammer is a more typical tool than a knife); in what contexts it would be appropriate to use a certain expression (that 'hello' can be used in order in start up a conversation); what can be inferred from utterances of sentences (that 'John likes clowns' can be inferred from a true utterance of 'John likes animals and clowns'); and what conditions must obtain in order to verify the utterance of a sentence (that an utterance of 'This is square' is verified if the corners of the object indicated is counted and they add up to four).

In addition to the information available from intuition, distributional information about candidate meanings can be acquired

by means of *empirical inquiry* (and complementary operational-ization). For instance, on the intensional construal of a *prototype* – the construal according to which a prototype is a structured set of weighted features (mental correlates of properties) – we have no intuitive way of telling which prototypes correspond to which general terms; but by operationalizing what it means for the prototype corresponding to a certain expression to have a certain feature with a certain weight (e.g. in terms of the performances of subjects in feature-listing experiments) we can obtain information about the distribution of prototypes over expressions.[1]

No matter how we obtain the distributional information, as long as it is satisfactorily reliable, we can use it, together with $C^{FE}$, to devise arguments with which we can draw conclusions about what meanings are. More exactly, once we have reliable distributive information about candidate meanings, we can use $C^{FE}$ to draw conclusions about what meanings are *not*, and thus indirectly – given a space of candidate meanings – draw conclusions about what meanings are. Potentially, then, we can learn about what meanings are by *comparing a known distribution with certain permissible distributions*. The argument will be along the following lines:

($P_1$)  Candidate meanings of kind *m* distribute over expressions in manner *x*.

($P_2$)  $C^{FE}$ requires meanings to be distributed over expressions in manner *y*.

($P_3$)  *x* and *y* are incompatible.

(C)  Meanings are not of kind *m*.

This *modus operandi* is like determining that a wedding is not a traditional wedding by seeing how the guests are seated. Suppose that in traditional weddings no two people of the same sex sit next to each other during dinner. Then, if we know how a group of wedding guests are seated (i.e. how they are distributed) and each person's sex, we can compare this distribution with those tolerated in the traditional seating arrangement; and if it turns out that two people of the same sex are sitting next to each other, we can conclude that it is not a traditional wedding.

---

[1] Prototypes are discussed at greater length in Subsection 6.1.3.

*Two Applications of $C^{FE}$*

Consider the view that meanings are extensions – i.e. that the meaning of a singular term is the individual to which the term refers; that the meaning of a general term is the set of individuals that term denotes; and that the meaning of a sentence is its truth-value.[2] A standard way to criticize theories of this kind is by applying a constraint like $C^{FE}$.[3] The relevant argument usually departs from the observation that we know, say, that (1) and (2) are co-extensive, but also (given appropriate information, in the case below about Kim) that (3) and (4) differ in truth-value.

(1)     dogs
(2)     canines
(3)     Kim believes that dogs are dangerous.
(4)     Kim believes that canines are dangerous.

This knowledge of the distribution of extensions over expressions seems likely to be incompatible with $C^{FE}$. According to $C^{FE}$ no two complex grammatical terms can differ in semantic range (and thus in meaning) if they have the same mode of composition and parts with the same semantic ranges. But given that (3) and (4) correspond to the same modes of composition – in something like the following ways – it follows that we have a conflict.[4]

---

[2] We will charitably ignore that it follows from this theory that out of every three meaningful sentences, at least two are synonyms.

[3] This argument is a staple in textbooks on the philosophy of language. See Lycan (2000) for instance. It can be reconstructed from passages in Frege's 'Uber sinn and Bedeutung'. The argument is usually presented as a substitution argument, i.e. that co-extensive expressions cannot be substituted for each other in all sentences without changing truth-values. $C^{FE}$ is equivalent to the substitution thesis that parts of complex terms be possible to substitute for other terms that have the same semantic range, without the semantic range of the complex term thereby changing. So the arguments are really the same. (Grammatical terms are often assumed to have unique meanings however, so the argument is usually presented in terms of meanings and not semantic ranges.)

[4] Although the presentation in the main text follows the way the argument is usually presented it should be noted that this way of putting it might be misleading. It focuses on the parts 'dogs' and 'canines' even though these are not immediate parts of (3') and (4'). The violation really occurs with [$_{NP}$ that [$_{NP}$ dogs [$_{VP}$ are dangerous]]]]],

(3')    [$_{\text{NP}}$ Kim [$_{\text{VP}}$ believes [$_{\text{NP}}$ that [$_{\text{NP}}$ dogs [$_{\text{VP}}$ are dangerous.]]]]]
(4')    [$_{\text{NP}}$ Kim [$_{\text{VP}}$ believes [$_{\text{NP}}$ that [$_{\text{NP}}$ canines [$_{\text{VP}}$ are dangerous.]]]]]

Thus it seems to be part of $C^{FE}$'s informative potential that extensions cannot be meanings.[5]

The same line of argument has been applied to the view that meanings are functions from possible worlds to extensions – i.e. to the thesis that the meaning of a singular term is a function from possible worlds to the individual that the term refers to in each world; that the meaning of a general term is a function from possible worlds to the set of individuals that the term denotes in each world; and that the meaning of a sentence is a function from possible worlds to its truth-value in these worlds. We know that (1) and (2) have the same extensions in all possible worlds, but also (given appropriate information about Kim) that there are possible worlds in which (3) and (4) differ in truth-value (as in the actual world, if one of the sentences is true and the other is false here). But given something like the modes of composition in (3') and (4'), there is a conflict with the distributions tolerated by $C^{FE}$. Thus it is part of $C^{FE}$'s informative potential that functions from possible worlds to extensions cannot be meanings.

It seems, therefore, that if a semantic theory is required to accord with $C^{FE}$, meanings are neither extensions nor functions from possible worlds to extensions. But before moving on we need to address some *prima facie* difficulties with the idea that $C^{FE}$ (or any of the other explications of $C$) has any informative potential.

## Some Worries about the Informative Potential of C

The notion that an explication of $C$ has informative potential presupposes that the grammatical terms underlying expressions are

---

[$_{\text{NP}}$ that [$_{\text{NP}}$ dogs [$_{\text{VP}}$ are dangerous]]]]], [$_{\text{VP}}$ believes [$_{\text{NP}}$ that [$_{\text{NP}}$ dogs [$_{\text{VP}}$ are dangerous]]]], and [$_{\text{VP}}$ believes [$_{\text{NP}}$ that [$_{\text{NP}}$ caniness [$_{\text{VP}}$ are dangerous]]]]].
[5] Extensional (as well as intensional) semantic theories are sometimes thought of as assignments not of semantic entities to syntactic entities, but as assignments of *types* of entities to syntactic *categories*. Thomason (1974: 48-50) advocates this view and it seems to be a pervasive view in the tradition in semantics following Montague. See Heim and Kratzer (1998: 300) for an illustration of how the problem discussed in the main text arises in the context of Montague style semantics.

constrained in some way.[6] Were there no constraints, one could always defend a particular conception of meaning against *C* by transforming problematic complex grammatical terms to simple ones. Quite what constraints, precisely, to adopt would depend on the pros and cons of different syntactic theories, and it is not possible to decide these issues here. But to illustrate the constraints, and to defend the conclusion that it is part of the informative potential of $C^{FE}$ that meanings cannot be extensions, we can note that two possible strategies for reconciling $C^{FE}$ with an extensional semantics are ruled out if we adopt the following version of a standard constituency test.[7]

Q        '*x*' corresponds to a part (at some level of immediacy) of
          the grammatical term corresponding to '…*x*…', if '*x*'
          can stand alone as an answer to a question.

Consider (5).

(5)     Paul ate at a really fancy restaurant.

Whereas some parts of (5) seem to constitute appropriate responses to various questions, others do not. Compare (6), which seems to be an appropriate response to the question 'What did Paul do yesterday afternoon?' with (7), which does not seem to be an appropriate answer to any question.

(6)     Ate at a really fancy restaurant.
(7)     Ate at.

Hence, according to *Q*, (6) but not (7), corresponds to a grammatical term which is a part (at some level of immediacy) of the grammatical term corresponding to (5).

   The paratactic account of belief reports discussed in the previous chapter (which offered a way out of Pelletier's argument from synonymy) can be seen as one way to reconcile $C^{FE}$ with the idea the meanings are extensions. This theory violates *Q*. It assumes that a

---

[6] Cfr. the discussion in Chapter Two.
[7] See Carnie (2002) for instance.

sentence such as (8) corresponds roughly to the following two structures given by (8').

(8)    Kim believes that clowns love monkeys.
(8')    [$_S$ [$_{NP}$ [$_N$ Kim]] [$_{VP}$[$_V$ believes] [$_{NP}$ [$_N$that.]]]]
        [$_S$ [$_{NP}$ [$_N$ Clowns]] [$_{VP}$[$_V$ love] [$_{NP}$ [$_N$monkeys.]]]]

In this case, there is no pair of complex grammatical terms with different meanings, the same mode of composition, and parts with the same meaning. However, if we adopt *Q*, this structural analysis is ruled out, since (9) suggests that 'that clowns love monkeys' corresponds to a part of the grammatical term corresponding to (8).

(9)    What does Kim believe? That clowns love monkeys.

*Q* also rules out a position sometimes attributed to Quine.

> Quine suggests, in fact, that it is a mistake to speak of "embedding" a referring expression within a that-clause at all. He suggests that an expression like "believed that Hesperus often rose in the evening" functions rather like a primitive one-place predicate with no 'logically germane' constituent parts. In that case, [the relevant] sentences […] are misleadingly spelled. A more revealing spelling would treat this predicate as a single, though very long word somewhat like: "believed-that-Hesperus-often-rose-in-the-evening". The presence of the string "Hesperus" in this very long word is merely an accident of orthography. It would be no more correct to regard "Hesperus" as a grammatical constituent of this very long word, than it would be to regard the word 'cat' as grammatical constituent of the word "cattle".
> (Taylor 1998)

According to this position (8) corresponds approximately to the following structure.

(8")    [$_S$ [$_{NP}$ Kim] [$_{VP}$ believes that clowns love monkeys.]]

But the latter is also incompatible with *Q*. This is again due to (9), but also due to(10) and (11).[8]

(10)    What does Kim believe that clowns love? Monkeys.
(11)    What does Kim believe that clowns do? Love monkeys.

So it seems that, already, by adopting *Q* we have a fairly firm grip on the structure of the underlying grammatical terms.

It should be noted that we need not require the constituency tests we adopt to determine *unique* modes of composition in order for an explication of *C* to have any informative potential. Were the tests to determine only a set of permissible modes of composition, a certain explication of *C* could still have informative potential in virtue of ruling out meanings being a particular kind of thing *x qua* that explication being incompatible with *all* the theories presupposing that meanings are *x* that are in accordance with the available distributional information and that invoke permissible structural assumptions.

Let us turn now to a different possible complication with the claim that an explication of *C* has a certain informative potential, one due to Paul Horwich.[9] Horwich's point can be put in the following way. In some cases in which there is an incompatibility between the distribution of some candidate meanings and compositionality, the conflict can be resolved by maintaining that, although the candidate meanings really are the meanings of the atomic expressions, they are not the meanings of complex expressions. Hence, if we drop the assumption that a semantic theory must be semantically *uniform*, i.e. that the meanings of all expressions must be of the same general kind, compositionality is, according to Horwich, compatible with *lexical* meanings being *anything*. This would mean that no information about what lexical meanings are follows from compositionality.

Two things should be noted in response to Horwich's remarks. First, it would appear that not all assignments of meaning to lexemes are compatible with $C^{FE}$, since in some cases, no matter what meanings are assigned to complex terms, we might have intuitions

---

[8] As Davidson ([1965] 2001: 13-14) also points out, such a proposal has the additional drawback that it seems incompatible with explaining why speakers are linguistically creative with respect to propositional attitude reports.
[9] See Horwich (2005).

about the divergence of meaning, which, if respected, will lead to $C^{FE}$ violations no matter what entities are substituted for the meanings of the complex expressions. Hence, $C^{FE}$ might rule out hypotheses about the nature of the meanings of lexemes if they lead to a very coarse-grained lexical semantics.

Second, Horwich's remarks are not really in conflict with the claims about informative potential that I have made so far. It is still the case that meanings cannot be extensions *in general*. It is still the case that meanings cannot be functions from possible worlds to extensions *in general*. This is compatible with not being able to say that *meanings of lexemes* cannot be extensions or functions from possible worlds to extensions.[10] So Horwich's observation does not exclude the possibility that the explications of compositionality have rich informative potential.

A final complication with the notion that $C^{FE}$ (or any of the other explications of *C*) has informative potential is that the distributive information is often contested. For instance, it is not uncontroversial to claim that 'Jane believes that clowns love canines' would be false if Jane would not assent to an utterance of the belief report. Some would instead argue that there is an explanation of why we think that the belief report (or utterances of it) is false other than the fact that it (or utterance of it) expresses a false proposition. I think that this point must be conceded: the distributive information is not always clear and might be obscured by irrelevant intuitions. Nonetheless, not all information is contested, and in the situations where it is contested, we can, on the assumption that the conflicts that arise can be resolved, use whatever distributive theses result from these debates in applications of $C^{FE}$. So even if the distributive information is not always straightforward, this does not strip the compositionality explications of informative potential.


### Two Applications of $C^{FE}$ (continued)

The main problem with the informative potential of $C^{FE}$ is not that it is contingent on the syntactic structure of grammatical terms being

---

[10] Horwich is under no illusion at regarding this point. His emphasis on lexical meanings is due, I think, to his criticism of Fodor and Lepore's ([1991] 2002) who emphasize (cfr. the quote in Section 6.1) the informative potential of compositionality with respect to *lexical* meanings.

determinate, nor that it is contingent on us knowing how candidate meanings distribute over expressions, or that it does not contain that much information about lexical meanings. The main problem is that since we have no reason to think that $C^{FE}$ *specifically* is true, we cannot *learn* anything from imposing it on a semantic theory. For if we now turn to the question how the informative potential of $D$ relates to that of $C^{FE}$, it is clear that $D$ has much more limited informative potential than $C$ does. Hence, what follows from imposing on semantic theories what we are justified in imposing differs from what follows from the imposition of $C^{FE}$.

   Consider the two previous applications of $C^{FE}$. They depended on the fact that two complex expressions with different meanings, but with parts with the same meaning and the same modes of composition, constituted a counter-example to $C^{FE}$. But expression pairs like these are not incompatible with $D$. A theory is compatible with $D$ if it is compatible with $F^{FI}$, and expression pairs like those just described are not counter-examples to $F^{FI}$, since this thesis allows reference to the forms of expressions as part of the implicit semantic rules. And since expressions like (3) and (4) have different forms, they are not in any immediate conflict with $F^{FI}$. This can be illustrated with the following non-compositional theory in accordance with $F^{FI}$.[11]

---

[11] This theory just illustrates the general point. See Larson and Ludlow (1993) for a more worked out, more plausible semantic theory along similar lines.

$t_1$   *Syntax:*                  *Semantics:*

   N → 'Max',      $\mu$('Max')      = Max

        'Peter',     $\mu$('Peter')     = Max

        'Judy'       $\mu$('Judy')      = Judy

   $V_1$ → 'killed',    $\mu$('killed')     = {$<x, y>$: $x$ killed $y$}

        'hugged'     $\mu$('hugged')    = {$<x, y>$: $x$ hugged $y$}

   $V_2$ → 'believed',  $\mu$('believed')   = {$<x, y>$: $x$ believed $y$}

        'knew'       $\mu$('knew')      = {$<x, y>$: $x$ knew $y$}

   NP → N        $\mu(_{NP}$ N)      = $\mu$(N)

   VP → $V_1 \cdot$NP     $\mu(_{VP}$ $V_1$ NP)   = $f(\mu(V_1), \mu$(NP))

   VP → $V_2 \cdot$S      $\mu(_{VP}$ $V_2$ S)    = $f(\mu(V_1)$, S)

   S → N$\cdot$VP      $\mu(_{S}$ N VP)     = $g(\mu$(N), $\mu$(VP))

 

   where...

   $f(x, y) = \{z : <z, x> \in y\}$, and

   $g(x, y) = 1$ iff $x \in y$, else 0.

Note that this theory does not accord with $C^{FE}$, since although 'Max hugged Judy' and 'Peter hugged Judy' mean the same thing, 'believed Max hugged Judy' and 'believed Peter hugged Judy' can mean different things. Nonetheless, the theory is in accord with $F^{FI}$ and hence *D*. Therefore we cannot conclude from the existence of $C^{FE}$ violations that we have a violation of *D*, and consequently we are not entitled to conclude that meanings cannot be extensions from these observations. Exactly the same line of argument applies to possible world semantics. From these reflections it can be seen, then, that it is no part of the informative potential of *D* that meanings cannot be extensions or functions from possible worlds to extensions.

## *Addendum to the $C^{FE}$ discussion*

The informative potential of *D* vis-à-vis propositional attitude constructions seems to be quite limited, since theories like $t_1$ might be available to account for them. In general, it seems that the informative potential of *D* relative to opaque contexts is fairly limited

when *M* – that is, what we might call the *Mates Thesis* – is true of the relevant contexts.[12]

(M)    No two formally distinct expressions can be substituted for each other without changing the meaning of the containing expression.

*M* seems plausible when applied to propositional attitude reports. For, in addition to the differences between expressions such as (3) and (4), quite superficial differences seem to have an impact on the meaning of such reports. Larson and Ludlow give the following example.[13] Assume that 'Hahvahd' is synonymous with 'Harvard', and that the former expression is used in order to indicate that 'Harvard' have been pronounced in a certain way. Then if some speaker, Jason, is unfamiliar with this way of pronouncing 'Harvard', (12) might well be true even though (13) is not.

(12)    Jason believes that Harvard is a fine school.
(13)    Jason believes that Hahvahd is a fine school.

Since *M* seems plausible for propositional attitude reports, it renders *D* less informative with respect to these. The same goes for many contexts featuring *quotation* where *M* also seems to be true.[14]

But *D* might not be entirely without informative potential, since *M* does not seem to be true in modal contexts. (9) and (10) seem to mean the same thing even though 'dogs' and 'canines' are formally distinct.[15]

---

[12] Mofett (2002) attributes the claim that belief contexts are semantically opaque to Benson Mates (1952). It seems reasonable that Mates did believe this but he does not explicitly commit to this thesis in his article.

[13] See Larson and Ludlow (1993).

[14] For instance, although 'Spiderman' and 'Peter Parker' necessarily co-refer (in the Spiderman fiction) the sentence 'Mary Jane Watson likes "Spiderman".' and 'Mary Jane Watson likes "Peter Parker".' might differ in truth value depending on Mary Jane Watson's attitudes to different words. (The context we are imagining is one where Mary Jane is considering two different names. Hence the quotation marks.)

[15] I'm assuming that 'dogs' and 'canines' are synonymous and not merely co-extensional. If one has conflicting intuitions the words can be exchanged for a more compelling synonymy pair.

(9)    Necessarily, dogs are nice.
(10)   Necessarily, canines are nice.

However, modal contexts can be recruited in order to throw doubt on extensional theories. For although 'cordates' and 'renates', for example, co-denote, and (14) and (15) are thus materially equivalent, (16) is true while (17) is false.[16]

(14)   Cordates are cordates
(15)   Cordates are renates
(16)   Necessarily, cordates are cordates.
(17)   Necessarily, cordates are renates.

So it seems that extensions violate $C^{FE}$, but that no attempted explanation couched in terms of $F^{FI}$ can account for this. Depending on the way in which $D$ is filled out (over and above $C^{FI}$, $F^{FI}$ $X^{FI}$ and $K^{FI}$), it still might be the case that some external factor influences the meaning of these constructions. But, owing to the failure of $M$ in these contexts, it seems unlikely that an extensional account of them employing $F^{FI}$ can be given. (Nor does any of the remaining principles seem appropriate.) So, at least at this stage it could be argued in view of the differences between constructions such as (16) and (17) that $D$ and $C^{FE}$ converge on the claim that meanings cannot be extensions. The aforementioned constructions are not a problem for possible world semantics.

To sum up this subsection, it seems to be part of the informative potential of $C^{FE}$ that meanings cannot be, say, extensions and functions from possible worlds to extensions. This is less significant than might appear, however. Since we are only licensed to require semantic theories to accord with $D$, we cannot infer, from the fact that their distribution over expressions is incompatible with $C^{FE}$, that extensions and functions from possible worlds to extensions are not meanings. Moreover, the informative potential of $D$ is significantly less rich than that of $C^{FE}$. Much distributional data that is incompatible with $C^{FE}$ is compatible with $D$. The exception to this general pattern is that both $D$ and $C^{FE}$ seem to be incompatible with

---

[16] I'm assuming that 'cordates' and 'renates' are not synonymous but merely co-extensive.

what we know about the way in which extensions behave in modal contexts. They both point to the conclusion that meanings are not extensions. But whereas, for instance, possible world semantics seem problematic from the perspective of $C^{FE}$, this is not the case with $D$.

### 6.1.2 $C^{FI}$ vs. D

The informative potential of $C^{FI}$ is greater than that of $C^{FE}$.

> A semantic theory $t$ accords with $C^{FI}$ iff

> i)   There are no explicit semantic rules for complex expressions in $t$.
> ii)  Each implicit semantic rule in $t$ is of the form '$\mu(x)=y$' where

>> 1)  '$x$' is a mode of composition variable, and
>> 2)  '$y$' is a *c-term*, where a *c-term* is defined in the following way:

>>> a)  '$\mu(x_1)$' is a *c*-term if '$x_1$' is one of the variables from which '$x$' is built up.
>>> b)  '$g(x_1,\ldots, x_n)$' is a *c*-term if '$x_1$',…,'$x_n$' are *c*-terms and '$g$' is a functor corresponding to a humanly computable function defined on the set of meanings.

Since $C^{FI}$ is logically stronger than $C^{FE}$, the informative potential of the later is contained in that of the former. But $C^{FI}$ also embodies additional informative potential, since in addition to the constraint embodied by $C^{FE}$ it requires a theory to contain implicit semantic rules of a certain kind. Given distributional information about candidate meanings, we can learn that meanings are not of this kind if it turns out that there are no *correct* implicit semantic rules that *accord with what we know about the way in which they distribute*. We remember from Chapter Three and Chapter Four that an implicit semantic rule's being correct was a matter of it correctly capturing causal dependencies, and it is evidence that the implicit semantic rules get the dependencies wrong that can utilized in applications of $C^{FI}$.

There need be no $C^{FE}$ violations among the known distributions for it to be possible to apply $C^{FI}$ in this way. All that is required is that the distribution across certain complex expressions cannot be distilled into a correct statement in terms of the distributions across the meanings of parts of the complex expressions. In fact, both the examples of $C^{FI}$-violations to be discussed in this section involve theories that accord with $C^{FE}$.

The argument featuring applications of $C^{FI}$ is the following.

(P$_1$)    Candidate meanings of kind *m* distribute over expressions in manner *x*.
(P$_2$)    $C^{FI}$ requires that there be composition rules.
(P$_3$)    No correct composition rules that accord with *x* exist.
(C)     Meanings are not of kind *m*.

*Two Applications of $C^{FI}$*

The first semantic theory we will consider identifies meanings with what we might call *fuzzy sets\**, but before attending to this theory we should briefly consider *fuzzy sets* (note the absence of the asterisk). Originally, fuzzy sets were conceived as functions – which we can refer to as *f*-functions – from entities to numbers between 0 and 1, which were assigned to categories with vague borders.[17] The function for a certain category was taken to indicate the degree to which each of the entities in the domain of the function belonged to that category. We can refer to these numbers as *membership scores*, or *degrees of membership*.

A fuzzy set-theoretical approach to semantics is quite similar to an extensional approach. However, on the fuzzy set approach it is not the case that each general term is assigned the set of things that fall under that term. Instead things are assigned, relative to each term, a membership score between 0 and 1 that can be interpreted as their degree of membership in the associated category. The higher the score is, the higher their degree of membership. On this view, an ordinary set can be understood as a fuzzy set where each object is assigned a score of either 0 or 1.

---

[17] Zadeh (1965) introduced the notion of a fuzzy set.

Now the compositionality arguments I want to address originally targeted the combination of fuzzy set theory and prototype theory. However, in order for these arguments to apply, the *f*-functions have to be interpreted as functions from entities to *typicality scores* rather than membership scores. This difference seems to be important, since there is a distinction between typicality and degree of membership: although, for instance, an ostrich and a robin differ in their typicality as birds, they are both indisputably birds.[18] Since the arguments that we will consider seems to depend on typicality judgments rather than membership judgments, I think the arguments are better construed as arguments directed at the notion of a fuzzy set* i.e. a function from entities to *typicality scores*, rather than ordinary fuzzy sets.[19]

Suppose a theory assumes that the meanings of general terms are fuzzy sets*. By adopting $C^{FI}$ we impose on that theory the requirement that it contain implicit semantic rules which feature correct humanly computable functions and which yield the typicality scores of complex expressions as values when fed the typicality scores of their parts.

The problem with fuzzy set* theory, in the present context, is that problems may arise when we try to devise functions that accord with what we know of the typicality scores of the members of the fuzzy sets*. Consider the simple adjective-noun combination 'red apple'. In his original formulation of fuzzy set theory and fuzzy logic, Zadeh did not present any rules for such combination, but if the composition of the meaning of such combinations can be assimilated to conjunctive composition, Zadeh's original treatment of conjunctive composition entails the following.[20]

$$f(x)_{\text{RED APPLE}} = \text{Min}(f_{\text{RED}}(x), f_{\text{APPLE}}(x))$$

where…

$$\text{Min}(x, y) = x \text{ iff } x \leq y, \text{ else } y.$$

---

[18] See Osherson and Smith (1997) and Hampton (2006b) for a discussion.
[19] The arguments were originally supposed to be directed at the combination of fuzzy set theory and prototypes but as far as I can tell they only depend upon the fuzzy set (or rather fuzzy set*) part and I will simplify the exposition by dropping the prototype part. We will return to prototypes below.
[20] See Zadeh (1975).

Here is an example-theory incorporating this rule.

$t_2$   *Syntax:*                        *Semantics:*

N   → 'apple',    $\mu$('apple')   $= f_{APPLE}$:   $\lambda o \in O.\ [0, 1]$
         'orange'    $\mu$('orange')  $= f_{ORANGE}$:  $\lambda o \in O.\ [0, 1]$
A   → 'green',    $\mu$('green')   $= f_{GREEN}$:   $\lambda o \in O.\ [0, 1]$
         'red',         $\mu$('red')       $= f_{RED}$:     $\lambda o \in O.\ [0, 1]$
         'striped'    $\mu$('striped')  $= f_{STRIPED}$: $\lambda o \in O.\ [0, 1]$
NP  → A·N          $\mu(_{NP}$ A N)   $= g(\mu(A), \mu(N))$

where…
$O$ is a set of objects
$[0, 1]$ is the reals between 0 and 1
$g(x, y) = \lambda o \in O.\ \text{Min}(x(o), y(o))$

Although this theory accords with $C^{FI}$, the composition rule it employs yields counterintuitive results. An object can have a higher typicality score relative to an adjective-noun combination than it has relative to either the adjective or the noun: a striped apple, for instance, might be more typical of a striped apple (it might be very typical) than it is of either striped things in general or of apples in general (either of which it might be very atypical of).[21] Unfortunately, for fuzzy set* theory, however, the problem is not just that this particular composition rule is inappropriate, for there is real difficulty, also, with finding an alternative that is not. The following rules also yield erroneous results.[22]

---

[21] This example was suggested by Osherson and Smith (1981: 267). Although Osherson and Smith later insist on the distinction between typicality and set-membership they are not kept clearly apart in this article where a single characteristic function is referred to as measuring both prototypicality (p. 263) and set membership (p. 266).

[22] Osherson and Smith (1982) has empirically validated that subjects typicality judgments deviate from the predictions of these rules. They also point out a problem with the following proposal made by Zadeh (1982).

$$f_{AN}(x) = \text{Min}\ (f_A\ (x), f_N\ (x))\ /\ \text{Max}\ (\text{Min}\ (f_A\ (i_1), f_N\ (i_1)),\ldots,\text{Min}(f_A(i_n), f_N\ (i_n)))$$

where…

$i_1 \ldots i_n$ is the set of all relevant objects.

$f_{AN}(x) = \text{Max}(f_A(x), f_N(x))$, where $\text{Max}(x, y) = x$ iff $x \geq y$, else $y$.

$f_{AN}(x) = \text{Prod}(f_A(x), f_N(x))$, where $\text{Prod}(x, y) = x^*y$.

$f_{AN}(x) = \text{Avg}(f_A(x), f_N(x))$,   where $\text{Avg}(x, y) = (x + y)/2$

At this point one might conclude, on the basis of the belief that a correct rule will never be found, that fuzzy sets* cannot be meanings. But the fact that we have merely *inductive* support for the third premise in the argument from $C^{FI}$ to the conclusion that meanings cannot be fuzzy sets* makes me hesitant to say that we have *learnt* that meanings cannot be fuzzy sets if we assume that $C^{FI}$ is true. The situation is better described by saying that we have *some reason to believe* that meanings are not fuzzy sets*. What we really need is a general argument to the conclusion that there are not any correct rules (of the appropriate sort) not just the observation that we have not been able to come up with any so far.

In point of fact such arguments can be found in the literature, and I think, therefore, that it is safe to say that it is part of $C^{FI}$'s informative potential that fuzzy sets* are not meanings. Let us consider now some of these general arguments. Let us begin with a rather general argument that does *not* establish this but that occurs so frequently in the literature that it seems important to highlight why it does not establish that fuzzy sets* cannot be meanings.

Osherson and Smith offer a general, quite ingenious, and very compelling argument against the possibility of there being *any* intuitively compelling composition rule that yields the typicality score for an object relative to a complex expression on the basis *only of the*

---

According to this function something can be more typical of a conjunctive combination than of either conjuncts. The problem with this proposal pointed out by Osherson and Smith is the following. Consider two plane figures *r* and *s*, both having the same color, an imperfect shade of red. *s* is squarer than *r*, and *r* is more square than it is red. Now, even if *s* can be indefinitely squarer than *r*, this is immaterial from the viewpoint of the new composition function. The dividend will be the same for *r* and *s* since they are equally red and more square than red. The divisor will be the same since they are both members of the same two categories. In this situation differences in squareness are ignored by the composition rule and thus yield counterintuitive results.

*typicality scores for that object relative the parts of that expression.*[23]
Imagine a round ball $r$ that is gradually turned into a square block by
slowly flattening it. At some stage in this process $r$, the object, is as
much ball as it is square. Hence,

(P$_1$)    $f_{\text{BLOCK}}(r) = f_{\text{BALL}}(r)$

Now, let $g$ be an arbitrary candidate composition function which
takes pairs of typicality scores as arguments and yields typicality scores
as values. No matter how this function is defined it will make (P$_2$)
true.

(P$_2$)    $g(f_{\text{ROUND}}(r), f_{\text{BLOCK}}(r)) = g(f_{\text{ROUND}}(r), f_{\text{BALL}}(r))$

(P$_2$) follows from (P$_1$) together with the fact that functions yield the
same values for the same arguments. If $g$ is understood as a *general*
composition function applying to several constructions (or at least to
'Round block' and 'Round ball'), (P$_3$) and (P$_4$)  are true as well.

(P$_3$)    $f_{\text{ROUND BLOCK}}(r) = g(f_{\text{ROUND}}(r), f_{\text{BLOCK}}(r))$
(P$_4$)    $f_{\text{ROUND BALL}}(r) = g(f_{\text{ROUND}}(r), f_{\text{BALL}}(r))$

(P$_2$) - (P$_4$) entail (P$_5$).

(P$_5$)    $f_{\text{ROUND BLOCK}}(r) = f_{\text{ROUND BALL}}(r)$

However, something that is a cross between a round ball and square
block would be much more typical of a round block than a round
ball, because typical round balls are completely round. Hence, no
matter how $g$ is defined, it will not be an adequate composition
function of the kind mentioned above; and hence no adequate theory
that is compatible with $C^{FI}$ incorporates such a function. If $C^{FI}$
requires there to be such a function, then, it can be safely inferred
from the previous argument that we can learn from $C^{FI}$ that fuzzy
sets* are not meanings.
   However, $C^{FI}$ does not require that there is such a function. The
previous argument applies only to functions $g(x,\ y)$ such that in

---

[23] See Osherson and Smith (1982).

computing, for instance, the fuzzy sets corresponding to adjective-noun combinations, $f_{AN}(x)$ depends only on $f_A(x)$ and $f_N(x)$ *but not* on the distribution of membership scores for other members of the relevant A and N categories. This restriction is not imposed by $C^{FI}$ since there is more information available to exploit by the composition functions.

Consider now the following general argument suggested by James Hampton.[24] Hampton discovered that typicality judgments of members of categories corresponding to relative clause conjunctions (e.g. 'sports that are also games') are subject to *the dominance effect*, i.e. the fact that conjuncts differ with respect to how predictive their typicality scores are of the typicality scores of their conjunction. So, for instance, the typicality distribution corresponding to 'sports' yields a better prediction of the typicality distribution corresponding to 'sports that are also games' and 'games that are also sports' than the typicality distribution corresponding to 'games' does. So in order to predict typicality for conjunctions one has to be able to predict dominance. But, the argument goes, on the extensional view of fuzzy set theory, there does not seem to be anything available that can predict dominance. Set size, for instance, does not seem to yield the right predictions and not much else is available. This suggests that there are no *correct* composition rules that a fuzzy set* semantics can contain. For such rules can only allow for a dependence on meanings of parts and modes of composition, and are correct or incorrect depending on whether they get the dependency right. But what Hampton's argument suggest is that there is some third thing which the fuzzy set corresponding to a complex expression depend on, hence, that no *correct* composition rules exist. This is compatible with there not having been any violation of functionality, hence no $C^{FE}$ violation, since the dependence on the third factor might not have revealed itself in this strong sense. To conclude, it is part of $C^{FI}$'s informative potential that fuzzy sets* cannot be meanings.

Before moving on to another argument with the same conclusion, we should stop and note that it is not part of the informative potential of $D$ that fuzzy sets* cannot be meanings. This means that we are not in a position to conclude that meanings are not fuzzy sets*

---

[24] See Hampton (1988:24).

since it is $D$ and not $C^{FI}$ that we can justifiable require that a semantic theory accords with.

Hampton explained the dominance effect in terms of *prototypes*, i.e. intensional abstractions from the relevant categories formally modeled as sets of weighted features, and the feature inheritance process that he supposed was part of prototype combination. On his model, if one conjunct corresponded to more salient attributes, then the feature set corresponding to the conjunction would be dominated (in terms of number of features) by that conjunct and hence (due to the mechanisms that was built in to generate typicality judgments) that conjunct would be more important in determining the typicality distribution corresponding to the conjunction. However, since a semantic theory is not required to accord with $C^{FI}$ but only with $D$ there is a satisfactory semantic theory according to which meanings are fuzzy sets that draws on Hampton's insights. For if a semantic theory is in accordance with $K^{FI}$, it is in accordance with $D$ and there is a theory in accordance with $K^{FI}$, that can mirror Hampton's predictions (by including Hampton's model as part of the theory). $t_3$ is a sketch of one such theory.

$t_3$  *Syntax:*                          *Semantics:*

   N  $\rightarrow$ 'sports',          $\mu$('sports')  $= f_{\text{SPORTS}}$: $\lambda o \in O.\ [0, 1]$

            'games',          $\mu$('games')  $= f_{\text{GAMES}}$: $\lambda o \in O.\ [0, 1]$

            'weapons',       $\mu$('weapons') $= f_{\text{WEAPONS}}$: $\lambda o \in O.\ [0, 1]$

            'tools'           $\mu$('tools')   $= f_{\text{TOOLS}}$: $\lambda o \in O.\ [0, 1]$

   NP $\rightarrow$ N· 'that are also'·N  $\mu(_{\text{NP}} N_1 N_2)$  $= g(\mu(A), \mu(N), b)$

where…
$O$ is the set of objects
$b$ is a set of beliefs
$[0, 1]$ is the reals between 0 and 1
$g(x, y, \text{z}) = \lambda o \in O.\ [0, 1]$.

The function $g(x, y, z)$ works in the following way. It constructs a composite prototype from the beliefs in $z$ that a speaker has about the fuzzy sets* $x$ and $y$, and then it uses this prototype to derive the

relevant fuzzy set* for the complex.[25] So given that Hampton's explanation of dominance is correct, a fuzzy set* theory can accord with dominance effects without conceding that meanings are not fuzzy sets*. Where Hampton's model assigns prototypes to expressions as their semantic values and then makes use of these directly to obtain a composite prototype that can be used to induce a fuzzy set, the previous theory makes use of the prototypes indirectly without supposing that they are semantic values of any expressions. This means that we cannot validly move from $D$ to the claim that meanings are not fuzzy sets. Hence it is not part, by dint of the previous argument, of the informative potential of $D$ that meanings are not prototypes. If Hampton's argument is read as an argument against meanings being fuzzy sets its validity depends on something like $C^{FI}$.[26]

   Consider next another general argument based on $C^{FI}$ to the effect that meanings cannot be fuzzy sets*. The typicality of an object relative a certain complex expression will often not be completely predictable by the typicality of its parts. A guppy for instance, is a typical pet fish, although an atypical pet and an atypical fish. The problem is not simply that the typicality of the guppy relative 'pet fish' exceeds its typicality relative 'pet' and 'fish', it is that a guppy is a typical fish partly in virtue of the distribution of pet fish in the world. As a matter of fact, guppies are common pet fish, but it could have been that trout be common pet fish. The pet fish in the world, and people's pet preferences determine which fish that are kept as pets. And this in part determines how typical something is for a pet fish. No composition rule will be able to get this right since the fuzzy sets* of the parts of the relevant complex expressions and their modes of

---

[25] This can be done if the weighted features are, or can be constructed from partial beliefs. Saying that the prototype corresponding to, say, 'dog' contains the feature 'barks' with weight $x$ is very close to saying that the subject believes that dogs bark with strength $x$. Once the features are in place, the composite prototype can be derived along the lines of Hampton's model. (See the next subsection for the details). If one wants to deny that prototypes can be reconstructed from partial beliefs, we can simply devise a new determination principle, call it $P^{FI}$ according to which prototypes can influence meaning derivations, in the same way that $K^{FI}$ allows for general knowledge to influence meaning derivation. $t_3$ would then be in accordance with $P^{FI}$. It seems clear that $P^{FI}$ would have the same explanatory utility as $K^{FI}$.

[26] Hampton's argument can also be interpreted as aiming only for the conclusion that the relevant composition has to *involve* prototypes.

combination are not sufficient determinants for doing this. We need knowledge of the world as well. Thus, if this is true no theory in accordance with $C^{FI}$ will be able to include *correct* composition rules, since these suggest that the dependence is more restricted. Hence, it is part of the informative potential of $C^{FI}$ that meanings are not prototypes.

But, again, fuzzy set* theory is only refuted on the assumption that we are required to adhere to $C^{FI}$, and not if what we are required to adhere to is *D*. For if theories compatible with $K^{FI}$ might be explanatory adequate semantic theories, general knowledge can play a role in meaning derivation. In this case the fuzzy sets* corresponding to complex expressions might be determined by the fuzzy sets* corresponding to the simple expressions, their modes of composition, *and general knowledge*. So the typicality of a guppy relative 'pet fish' is determined by its typicality relative 'pet' and 'fish' and by the things we know about guppies, e.g. that they are fish and that people often keep them as pets. This means that there might be correct $K^{FI}$-rules. Hence, the general argument to the conclusion that meanings are not fuzzy sets* do not go through. Again, it does not seem that it is part of the informative potential of *D* that meanings are not fuzzy sets*.

To sum up; even though it seems to be part of the informative potential of $C^{FI}$ that meanings are not fuzzy sets*, this is not part of the informative potential of *D*. Since we are only licensed to require that a semantic theory accords with *D* and not that it should accord with $C^{FI}$ we have not at this stage learnt that meanings are not fuzzy sets*.

## Another Application of $C^{FI}$

Consider next semantic theories that are species of *Inferential Role Semantics*. Paul Boghossian has given the following brief but useful characterization of these theories.

> Let's suppose that we think in a language of thought and that there are causal facts of the following form: the appearance in *O*'s belief box of a sentence $S_1$ has a tendency to cause the appearance therein of a sentence $S_2$ but not of $S_3$. Ignoring many complications, we may describe this sort of fact as consisting in *O*'s disposition to *infer* from $S_1$ to $S_2$, but not to $S_3$.

Let's call the totality of the inferences to which a sentence is capable of contributing, its *total inferential role*[…]Against this rough and ready background, an *inferential role semantics* is just the view that there is some construct out of an expression's total inferential role that constitutes its meaning what it does.

(Boghossian 1993: 73-74, italics in original)

On this view, the meaning of a sentence such as 'John is happy and is smiling' might be constituted by the inferences 'John is happy and is smiling' → 'John is happy', 'John is happy and is smiling' → 'John is smiling', 'John is happy and is smiling' → 'Someone is happy and is smiling' etc. Sub-sentential expressions can be understood along similar lines. The meaning of 'dog' might thus be constituted by the inferences '*a* is a dog' → '*a* is an animal', '*b* is a dog' → '*b* barks', '*c* is a dog' → '*c* has four legs' etc. The following theory is a simple inferential role semantics compatible with $C^{FI}$.

$t_4$     *Syntax:*
  N   → 'cow', 'cat'
  A   → 'brown', 'black'
  NP → A·N

*Semantics:*
$\mu$('cow')     = { 'x is a cow' → 'x is an animal',
                      'x is a cow' → 'x produces milk',
                      …                                        }
$\mu$('cat')     = { 'x is a cat' → 'x is an animal',
                      'x is a cat' → 'x meows',
                      …                                        }
$\mu$('brown')  = { 'x is brown' → 'x is colored',
                      'x is brown'→ 'x is non-black',
                      …                                        }
$\mu$('black')   = { 'x is black' → 'x is colored',
                      'x is black'→ 'x is non-brown',
                      …                                        }
$\mu(_{NP}$ A N)   = $\mu(A)\cup\mu(N)$.

It has been pointed out that a theory such as $t_4$ does not really accord with what we know about the inferential roles corresponding to

various expressions, and once this is remedied, we will no longer have a theory that will conform with compositionality.

> Suppose, for example, that you happen to think that brown cows are dangerous; then it's part of the inferential role of "brown cow" in your dialect that it does (or can) figure in inferences like "brown cow" → "dangerous". But, first blush anyhow, this fact about the inferential role of "brown cow" doesn't seem to derive from corresponding facts about the inferential roles of its constituents…it doesn't look like either "brown" or "cow" entails "dangerous", so, to this extent, it doesn't look like the inference from "brown cow" to "dangerous" is compositional.
>
> (Fodor and Lepore [1991] 2002: 16-17)

Since we know, for instance, that 'brown cow' → 'dangerous' is part of the inferential role of 'brown cow' if we believe that brown cows are dangerous, and if this is not predictable from the inferential roles of 'brown' and 'cow', then there will not be a *correct* composition rules for inferential roles: the inferential roles for complex expressions would depend on something additional to the determinates that the composition rules can cite. This means that semantic theories presupposing that meanings are inferential roles that accord with our distributive knowledge of inferential roles will not be in accord with $C^{FI}$.[27] Hence, it is part of $C^{FI}$'s informative potential that meanings are not inferential roles.[28]

Note that it is not part of $D$'s informative potential that meanings are not inferential roles. A theory in accordance with $K^{FI}$, is in accordance with $D$, and there are theories in accordance with $K^{FI}$ that are in accordance with what we know about the distribution of inferential roles. For since the troubling inferences are due to what we believe about the world, a theory that leaves a place in the meaning

---

[27] Here is another instance where the distributive information about candidate meanings have been disputed. Block (1993) and McCullagh (2003) have argued that the inferential roles corresponding to simple expressions contain conditional inferences (such as 'is dangerous if brown' for $\mu$('cow')) which would preserve compositionality.

[28] Fodor and Lepore are explicit about the argument relying on there not being a principled analytic-synthetic distinction so I'm going to assume that no such distinction can be made to make their argument as strong as possible. See Block (1993), McCullagh (2003), Horwich (2005) and Greenberg and Harman (2006) for other solutions.

derivation of meanings for complex expressions for such beliefs can account for the data. Consider the theory that we discussed in Chapter Three for instance.

$t_8$     *Syntax:*
N   → 'cow', 'cat'
A   → 'brown', 'black'
NP  → A·N

*Semantics:*
$\mu$('cow')     = { 'x is a cow' → 'x is an animal',
                          'x is a cow' → 'x produces milk',
                          …                                    }
$\mu$('cat')      = { 'x is a cat' → 'x is an animal',
                          'x is a cat' → 'x meows',
                          …                                    }
$\mu$('brown') = { 'x is brown' → 'x is colored',
                          'x is brown'→ 'x is non-black',
                          …                                    }
$\mu$('black')    = { 'x is black' → 'x is colored',
                          'x is black'→ 'x is non-brown',
                          …                                    }
$\mu(_{NP} A N)$     = $\mu(A) \cup \mu(N) \cup f(\mu(A) \cup \mu(N), b)$

where…
$b$ is a set of extensional feedback
$f(x, y)$= the set of inferences corresponding to the subset of $y$ that is the   extensional   feedback   which   is about the things that the inferences in $x$ pertain to.

Again, the idea here is that if we assume that the speaker has come across, for instance, animals of a certain color that produce milk and learnt about them that they are dangerous, this belief will be part of $b$. And when $f$ is fed {'x is a cow' → 'x is an animal', 'x is a cow' → 'x produces milk'…}, $b$ it will yield something like 'x is a brown cow' →

'x is dangerous' as value.[29] Hence, by drawing on the beliefs a speaker has about the world, a theory like $t_5$ can account for the inferential roles corresponding to various complex expressions. Which means that the previous problem is not a problem for a theory in accordance with $K^{FI}$, and *eo ipso*, not a problem for a theory in accordance with $D$. It is thus not part of the informative potential of $D$ that meanings are not inferential roles. We are thus not in a position where we can conclude that meanings are not inferential roles, in spite of them being incompatible with $C^{FI}$.[30]

To sum up this subsection; it seems that it is part of the informative potential of $C^{FI}$ that meanings cannot be, for instance, fuzzy sets* or inferential roles. This is not very consequential however, since it is not part of the informative potential of $D$ – the constraint that we are actually justified in imposing on semantic theories – that meanings cannot be fuzzy sets* or inferential roles. We are thus in a situation with $C^{FI}$ that is similar to that we were in at the end of the previous subsection with $C^{FE}$. Although it is part of the informative potential of the relevant explication of $C$ that meanings cannot be certain things, we have not by noticing this learnt that meanings cannot be these things, since it does not follow from what we are justified in imposing on semantic theories that meanings cannot be these things.

### 6.1.3 $C^{BU}$ vs. $D$

Just like the informative potential of $C^{FI}$ included but went beyond that of $C^{FE}$, the informative potential of $C^{BU}$ includes but goes beyond that of $C^{FI}$.

---

[29] Note that $t_5$ does not accord with $C^{FI}$ since it contains the implicit semantic rule '$\mu(_{NP}\ A\ N) = \mu(A) \cup \mu(N) \cup f(\mu(A) \cup \mu(N),\ b)$' the right-hand side of which is not a *c*-term since it contains the term '*b*'.

[30] I'm ignoring, of course, all other reasons that might exist that allow us to conclude that meanings are not inferential roles. I'm just concerned with a very particular line of though based on the $C^{FI}$ and $D$.

A theory $t$ accords with $C^{BU}$ iff

i)   There are no explicit semantic rules for complex
     expressions in $t$.

ii)  Each implicit semantic rule in $t$ is of the form
     '$\mu(x)=y$' where

     1)  '$x$' is a mode of composition variable, and
     2)  '$y$' is a *c-term*, where a *c-term* is defined in the
         following way:

         a)  '$\mu(x_1)$' is a *c*-term if '$x_1$' is one of the
             variables from which '$x$' is built up.
         b)  '$f(x_1,\ldots, x_n)$' is a *c*-term if '$x_1$',…,'$x_n$' are *c*-
             terms and '$f$' is a functor corresponding to
             a humanly computable function defined
             on the set of meanings.

iii) For each complex grammatical term $e=\sigma(e_0,\ldots,e_{n-1})$
     and each meaning $m\in\mu'(e)$ it is the case that $m$ is a
     structure built up from the parts $m_0,\ldots,m_{n-1}$, where
     $m_0\in\mu'(e_0),\ldots,m_{n-1}\in\mu'(e_{n-1})$, and for no two meanings
     $m$, $m'\in\mu'(e)$ it is the case that $m$ and $m'$ is built up
     from the same parts.

iv)  For each syntactic operation $\sigma$, and for each
     sequence of expressions $e_0,\ldots,\ e_{n-1}$ for which $\sigma$ is
     defined, given that $\sigma(e_0,\ldots,e_{n-1})=e$ then for each $n$-
     tuple $m_0\in\mu'(e_0),\ldots,m_{n-1}\in\mu'(e_{n-1})$ it is the case that
     there is a meaning $m\in\mu'(e)$ which is a structure
     built up from the parts $m_0,\ldots,m_{n-1}$.

The part of the informative potential of $C^{BU}$ that goes beyond that of
$C^{FI}$ rests on the assumption that we not only know that candidate
meanings distribute over expressions in a certain way, we also know
something about how they are structurally related, i.e. we have
information about what things that are parts of which other things.
Perhaps this is easiest illustrated with those semantic theories that
identify meanings with certain kinds of sets, where the subsets (and

possibly the elements) of these sets can be seen as the parts of these sets.[31] We know for instance, that the parts of the set of pigs (the denotation of 'pigs') correspond to individual pigs (on the element reading of 'part') or sets of pigs (on the subset reading of 'parts'). Other candidate meanings such as propositions and properties can also be complex, and thus have parts. Since $C^{BU}$ requires that meanings are structurally related in a certain way, we can learn what meanings are by recourse to the relevant structural information, and the relevant distributive information. The argument that provides $C^{BU}$ with additional informative potential is thus the following.

(P₁)   Candidate meanings of kind *m* distribute over expressions in manner *x*.

(P₂)   Candidate meanings of kind *m* are structurally related in manner *y*.

(P₃)   $C^{BU}$ requires that meanings are structurally related in manner *z*.

(P₄)   *z* is incompatible with the combination of *x* and *y*.

(C)   Things of kind *m* are not meanings.

## *An Application of $C^{BU}$*

Consider again a semantic theory where meanings are assumed to be extensions, i.e. a semantic theory according to which the meanings of singular terms are individuals, meanings of general terms are sets of individuals, and the meaning of sentences are truth-values. According to this kind of theory adjective-noun combinations might be handled in terms of set intersection in a familiar way.

| $t_6$ | Syntax: | | Semantics: | |
|---|---|---|---|---|
| | N → | 'apple', | $\mu$('apple') | = {*x*: *x* is an apple} |
| | | 'pear' | $\mu$('pear') | = {*x*: *x* is a pear} |
| | A → | 'green', | $\mu$('green') | = {*x*: *x* is green} |
| | | 'brown' | $\mu$('brown') | = {*x*: *x* is a brown} |
| | NP → | A·N | $\mu$(NP A N) | = $\mu$(A) ∩ $\mu$(N) |

---

[31] The subset relation is ideal for the part-relation since it accords with all the standard mereological axioms on parts. The subset relation is reflexive, antisymetrical, transitive and obeys strong supplementation. See Simons (1987).

We know that the sets of individuals that these expressions denote are those given by the theory. In addition we know, if we take the subset relation as a part relation, that certain of the meanings are parts of certain others. The meaning of 'pear' for instance has the meaning of 'brown pear' as a part. So we know things about the distribution and structural relations of extensions. But according to $C^{BU}$ the meaning of a complex expression must have as parts the meanings of the parts of that expression. Hence, we know that the meaning of 'brown pear' must have the meaning of 'pear' as part. This is incompatible with $t_6$. The extension corresponding to 'brown pear' is not a part (/subset) of the extension corresponding to 'pear'. In fact, as we have just noticed, the reverse situation is the case. In general, the extension of a non-pleonastic adjective-noun expression does not contain as parts the extensions corresponding to the parts of that expression.[32, 33] This incompatibility is not limited to adjective-noun combinations but is also true of definite descriptions. The reference of 'the king of Sweden', for instance, does not contain as part the reference of 'Sweden', or the things denoted by 'king'. Indeed, the parts of the king of Sweden seems completely unrelated to the extensions of the parts of 'the king of Sweden'. It can thus be concluded that it is part of $C^{BU}$'s informative potential that meanings cannot be extensions.[34]

---

[32] A *pleonastic* expression is an expression that contains a part that is semantically superfluous, i.e. that can be removed without change in meaning of the original expression.

Not only intersective adjective-noun combinations (i.e. ones where the extension corresponding to the combination is the intersection of the extensions of the adjective and the noun), violate $C^{BU}$. Subsective adjectives (e.g. skilful) and privative adjectives (e.g. fake) does so as well. But it is particularly interesting that the the intersective adjective-noun combinations are incompatible with this form of compositionality since these kind of combinations are the ones where accounting for the compositionality of complex expressions purely in terms of extensions seems to be the most promising; the meaning (extension) of a complex expression can, by definition, be arrived at by means of taking the intersection of the meanings (extensions) of its parts. Subsective adjectives and privative adjectives are more recalcitrant to extensional treatment. See Kamp and Partee (1995).

[33] Note that the problem is avoided if the meanings of general terms are properties rather than sets. Then, the meaning of 'grey dog' could be the complex property BROWN PEAR that contains both BROWN and PEAR as parts. This is in conformity with $C^{BU}$.

[34] Although this is also a part of the informative potential of $C^{FE}$, it is part of $C^{BU}$ regardless of the considerations pertaining to opaque constructions that were discussed in Subsection 6.2.1.

It should be noted that in spite of their incompatibility with $C^{BU}$, theories that suppose that meanings are extensions we have not thereby learnt that these entities are not meanings. For all the relevant theories are perfectly compatible with *D*. If a theory is compatible with $C^{FI}$ it is compatible with *D*. $t_6$ is compatible with $C^{FI}$ since it contains implicit semantic rules the right-hand side of which only featured *c*-terms. Hence, the previous considerations do not give us any reasons to think that it is part of *D*'s informative potential that meanings are not extensions. Hence we find no reasons among these considerations to think that meanings cannot be extensions.[35]

---

[35] Depending on how extensionally one understands functions, the $C^{BU}$ constraint might actually be incompatible with several other kinds of semantic theories. Many semantic theories assume that meanings can be understood in terms of functions of various kinds. If this is understood extensionally (i.e. that the relevant functions are understood in-extension), and the relevant theories are interpreted as really assuming that meanings *are* the relevant functions, then these theories are incompatible with $C^{BU}$. Although the view that meanings really are these functions is surely a minority view (hence the placement of this remark in a footnote), it is still remarkable, I think, that *none* of these theories accord with $C^{BU}$, since the incompatibility actually arises from what has been considered to be one of the most important contributions made by Richard Montague to formal semantics, that of seeing 'function-argument structure as the basic semantic glue by which meaning are combined' (Partee 1997: 27).
   Consider the following theory to see the incompatability.

| $t_7$ | Syntax: | | Semantics: | |
|---|---|---|---|---|
| | N $\rightarrow$ | 'apple', | $\mu$('apple') | = {x: x is an apple} |
| | , | 'pear' | $\mu$('pear') | = {x: x is a pear} |
| | A $\rightarrow$ | 'green' | $\mu$('green') | = $\lambda o \in O. \, p \in O$ |
| | | 'brown' | $\mu$('brown') | = $\lambda o \in O. \, p \in O$ |
| | NP $\rightarrow$ | A·N | $\mu(_{NP} A \, N)$ | = $\mu(A)(\mu(N))$ |

   where…
   *O* is a set of extensions

If the functions (those corresponding to 'green' and 'brown' for instance) invoked by theories like $t_7$ are understood in-extension, then the inclusion of these in the theory results in a $C^{BU}$ violation. Since functions are ordered pairs, it is not the case that, for instance, the extension of 'green apple' which is the set of green apples, will contain the meaning of 'green' which is a set of ordered pairs of extensions. It is clear that any semantics that tries to combine meanings by function-argument application where one of the meanings is a function, the other is an appropriate argument for that function and the function is understood extensionally will not be in accordance with $C^{BU}$ (assuming a well founded set theory). It is also true for various other theories which assume that meanings are functions. Say that the meaning of sentence is its a function from possible worlds to truth-values. Then it is not the case that the

function corresponding to a sub-sentential expression is part of these functions. Ordered pairs consisting of possible worlds and truth-values do not contain sets of ordered pairs consisting of possible worlds and (for instance) sets of objects as parts, they contain ordered pairs as parts. This problem generalizes to other similar semantic entities such as Kaplanian characters, i.e. functions from contextual factors to contents. It can thus be concluded that it is part of $C^{BU}$'s informative potential that meanings cannot be, for instance, functions from possible worlds to extensions or Kaplanian characters (if these are understood in-extension).

It is unclear to me to what extent one needs to regard the assumptions made about semantic values in the Montague tradition as pertaining to what meanings really are. Most researchers in this tradition are not concerned with ontological questions. Nonetheless, consider the following paragraph from an introduction to semantics in this tradition.

> Montague's general framework leaves wide latitude for choices at many points, including in particular many choices about the nature of the model structures which make up the semantic algebras. The choices involved in the semantic algebras are of two main sorts: structural and ontological. The structural choices involve the type theory and the choices of semantic operations; the ontological choices concern principally the nature of the basic elements, such as domains of entities or of possible entities, a domain of moments or intervals of time, a domain of truth values, a domain of possible worlds, etc.
>
> (Partee 1997: 24-25)

If the choice of basic entities is ontological, why shouldn't the choice of the derived types (in particular which derived types correspond to which syntactic categories) also be ontological? I will not make any assumptions at all about how these theories should be best treated. If they are interpreted as candidate meanings, and the functions are interpreted extensionally, then they are incompatible with $C^{BU}$.

An additional complication is that in formal semantics certain constructions are treated as equivalent (a set and its characteristic function for instance) even though they are mereologically distinct. The application of $C^{BU}$ requires that a particular theory is identified with a particular mereological structure.

It should be noted that in spite of their incompatibility with $C^{BU}$, theories that suppose that meanings are functions from possible worlds to extensions or characters, we have not thereby learnt that these entities are not meanings. For all the relevant theories seems perfectly compatible with $D$. If a theory is compatible with $C^{FI}$ it is compatible with $D$. $t_7$ is compatible with $C^{FI}$ since it contain implicit rules the right-hand side of which only feature c-terms. Hence, the previous considerations do not give us any reasons to think that it is part of $D$'s informative potential that meanings are not functions from possible worlds to extensions or characters, hence we find no reasons among these considerations to think that meanings cannot be functions from possible worlds to extensions or characters.

*Another Application of C^{BU}*

Consider now *Prototype Theory*. This theory invokes *prototypes* – a specific kind of mental representation – to play the role of meanings.[36] A prototype is at its core a set of weighted features, where features are, roughly, mental correlates of properties and relations. Accordingly, features are as diverse as properties and relations are and include, for instance, *blue*, *square*, *lives in a domestic home* and *dangerous*.[37] Prototype theory was initially constructed in order to account for speakers' judgments about typicality and set membership relative to everyday natural language categories such as those corresponding to 'birds', 'tools', or 'fruit'.[38] These judgments include, for instance, the judgment that robins are more typical birds than penguins, that hammers are more typical tools than knifes, and that apples are more typical fruits than coconuts. The weight assigned to a feature with respect to a certain prototype reflects how important that feature is when determining, relative a certain category, how typical something is of that category or to what degree something belongs to that category.[39, 40]

---

[36] See Rosch (1975) and Rosch and Mervis (1975) for some early seminal work on prototype theory. It bears emphasis that prototype theory is an empirical theory of *concepts*. As such it might be felt to be out of place in the present exposition. However, i) prototype theories can be seen as fleshed out use theories of meaning and as such they have a natural place in the discussion; although prototype theory is sometimes explicitly stated in terms of 'concepts', the psychological notion of a concept is, not least due to the operationalization procedures employed, very close to word meanings. For instance, a concept being simple or complex is often a matter of it being expressed by a simple or complex expression. See, for instance, Murphy (1988: 530-531 ) for an explicit definition and Smith and Osherson et al. (1988: 355) for implicit agreement on this point; and iii) the distinction between word meaning and concepts is seldom emphasized by psychologists . Hampton, writes for instance that 'The central insight of prototype theory is that word *meanings*, and the conceptual classes that the words name, are distinguished one from another not in terms of an explicit definition, but in terms of similarity to a generic or best example.' (Hampton 2006a: 79, emphasis added). All in all, there is a sufficiently close connection between the meaning of an expression and a concept in this literature to treat a theory of the latter as a theory of the former.

[37] I will use italicized expressions to refer to features.

[38] It is controversial whether prototypes can be employed to account for set membership judgments or whether they can only be used to account for typicality judgments. See Osherson and Smith (1997) and Hampton (2006b) for some of the recent discussion of this issue.

[39] Exactly how 'weight' is operationalized vary between models. For Hampton, a prototype corresponding to a certain expression contains a certain feature (roughly) if

There are several theories that complement the previous picture by detailing how prototypes combine, i.e. theories describing how prototypes corresponding to complex expressions can be derived.[41] I will discuss the relation between prototype theory and $C^{BU}$, by using as an example *The Composite Prototype Model*, a model developed by James Hampton. The composite Prototype Model was developed in order to account for typicality and set membership judgments relative to relative clause conjunctions, i.e. combinations of the form 'N's that are also N's'.[42] In particular, the model was an attempt to account for two phenomena – overextension and domination – that Hampton discovered commonly characterized these kinds of combinations.[43] The model can be described in the following way.[44]

---

at least three out of ten subjects report that the relevant feature is an attribute of things denoted by that expression. But the weight of a feature does not correspond directly to its frequency of production. Instead, subjects were asked to judge to what extent an attribute was useful in defining the relevant expression and the resulting classification of attributes, from 'necessarily true of all possible examples of the concept' to 'necessarily false of all possible examples of the concept' were ranked from +4 to -2. The feature weight is then the mean rated importance. So a feature is only assigned, for instance, the maximum weight of +4 if all subjects deemed the attribute to be 'necessarily true of all possible examples of the concept'. This measure often correlates significantly with production frequency although the average correlation might be quite low.

[40] This core representation has become increasingly fleshed out in order to meet the demands of new empirical discoveries. For instance, the features of a prototype are often assumed to be organized into various *dimensions* such as those corresponding to 'color', 'shape', 'habitat' etc. And in addition to features being assigned weights, these dimensions are sometimes assigned *diagnosticity values* which are estimates of how good that dimension is for discriminating between different categories (relative some set of categories), where these estimates also play a role in accounting for typicality and set membership judgments. In addition to organizing features into dimensions many variants of prototype theory assume that prototypes are such that their features are related by various forms of *dependency links*. It is assumed, for instance, that the prototype corresponding to 'bird' not only includes the feature *flies* and *has wings*, but also a link between these features which encodes the information that a bird is able to fly *because* it has wings. To avoid as much unnecessary complication as possible I will only be concerned with prototypes in the core sense of a set of weighted features. See Hampton (1993) for a good discussion.

[41] See, for instance, Hampton (1987; 1988a; 1988b; 1991; 1997), Murphy (1988), Smith and Osherson et al. (1988), Wisniewski (1997) and Wisniewski and Love (1998).

[42] See, for instance Hampton (1987; 1988b).

[43] Storms and de Boeck et al (1996) and Storms and Ruts et al. (1998) have demonstrated that the two phenomena discovered by Hampton are fairly general. Not only relative clause conjunctions, but also adjective-noun combinations and

$t_8$   *Syntax:*                          *Semantics:*

   N  →  'sports',              $\mu($ 'sports') = {  *is done for fun* (2),
             'games',                                     *uses a ball* (1),
             'birds',                                            …}
             'pets',                 $\mu($'games')  = {  *has rules* (4),
             'furniture',                                  *has a goal* (3),
             'machines',                                       …}
             'buildings',             $\mu($ 'birds')  = {  *has feathers* (5),
             'weapons',                                     *flies* (3,6),
'           'furniture'                                        ...}
             'vehicles',              $\mu($ 'pets')    = …
             'dwellings',                 …
             'tools'                         …

  NP → N·'that are also'·N      $\mu(_{NP} N_1 N_2) = h(\mu(N_1), \mu(N_2), K)$

where…

$h(P_1, P_2, K) =$    the prototype $P$ resulting from the following procedure;

        i)    each feature which is a member of either $P_1$ or $P_2$ is added  to $P$.

---

noun-noun combinations exhibit both phenomena. And after a test for dominance in over fifty compounds it was shown that about two thirds of them exhibited dominance.

[44] The theory I describe here deviates from Hampton's on at least the following counts: 1) the syntax is slightly more general than the one proposed by Hampton. His model only makes predictions about the combinations that result from combining the members of seven pairs, e.g. 'sport' and 'game', 'bird and pet', 'furniture' and 'household appliance',…, and 'tool' and 'weapon'. His model does not make predictions about combinations across these pairs. It doesn't strictly speaking, for instance, make any predictions about the combination that results from combining 'sport' and 'household appliance'. So the theory is more general than Hampton's original model, but using it has a certain expository convenience. 2) Step iii) is not explicitly part of Hampton's original model but I understand Hampton to mean that the feature sets that correspond to relative clause conjunctions include features that are not parts of the feature sets corresponding to the conjuncts, but are derived, for instance, from extensional feedback, i.e. the speaker having been directly exposed to members of the relevant category. So, strictly speaking, in order for the model to generate the right feature sets, something like step iii) must be part of it. See Hampton (1991) for an elaboration of the model.

ii)   each feature $f \in P$ such that $k(w(f, P_1), w(f, P_2)) < 1,5$ is removed from $P$ unless $f \in n(P_1) \cup n(P_2)$. Each feature $f \in P$ such that $f \in i(P_1) \cup i(P_2)$ is also removed.

iii)  $P$ is modified by adding or subtracting features so that the resulting feature set is coherent and in accordance with $K$.

For each feature $f \in P$ it is the case that $w(f, P) = k(w(f, P_1), w(f, P_2))$,

$k(x, y)=$     a rising monotonic function of $x$ and $y$ (a weighted average have proven to fit the middle range of values),

$w(x, P) =$     the weight of a feature $x$ in prototype $P$,

$n(P) =$     the set of features that at least one subject reported are necessary for something to fall under $P$, and

$i(P) =$     the set of features that at least one subject reported are impossible for something falling under $P$.

This theory is not compositional since general knowledge plays a role in determining the prototypes corresponding to complex expressions. Less formally expressed, $h(x, y, k)$ (where $x$ and $y$ are prototypes, and $k$ is a set of beliefs) is such that it generates a prototype which corresponds to the union of the feature sets $x$ and $y$ (the prototypes corresponding to the parts of the relevant expression) where the feature weights for the features have been shifted in accordance with a rising monotonic function from the feature weights of each feature in $x$ and $y$. Some features with very low resulting feature weights have been dropped, features that were deemed impossible for one of the prototypes have also been dropped, and the resulting feature set has been made coherent and is in accordance with the general knowledge in $K$.

Consider the following example. Assume that the feature sets corresponding to 'birds' and 'pets' are correctly described by the following. [45]

---

[45] Scores are taken from the data reported by Hampton (1987).

$\mu$('Birds') =          { *migrates* (3.75),
                            *is an animal* (3.00),
                            *has feathers* (3.62),
                            …                                    }

$\mu$('Pets') =           { *is an animal* (3.25),
                            *provides companionship* (2.75),
                            *lives in a domestic home* (2.75),
                            …                                    }

$t_8$ then predicts that the feature set corresponding to 'Birds that are also Pets' is something like the following.

$\mu$('Birds that are also Pets') =  { *is an animal* (2.50),
                                        *has feathers* (3.25),
                                        *provides companionship* (1.75),
                                        *lives in a domestic home* (1.87),
                                        *is kept in a cage* (2.00),
                                        …                                    }

Some of the features that are part of $\mu$('Birds') and $\mu$('Pets') – *is an animal* and *has feathers* for instance – are also parts of $\mu$('Birds that are also pets'). Others are dropped since they are deemed impossible for something falling under one of the conjuncts or to avoid low overall coherence – *migrates* for instance, is a feature of $\mu$('Birds') which is inconsistent with *lives in a domestic home* (all the time) which is a feature of $\mu$('Pets'), so in order to preserve coherence one of them has to be dropped. Some features, so called 'emergent features' are not part of either $\mu$('Birds') or $\mu$('Pets') but are nonetheless part of $\mu$('Birds that are also pets') – *is kept in a cage*, for instance, is such feature. It is likely that most of the emergent features (and some of the features being dropped) are due to *extensional feedback*, i.e. the speaker's exposure to members of the conjunctive category. It can also be the case that some features are dropped due to low average weight

(although this is not illustrated in this example)[46]. In some cases, the feature weights of a feature relative the prototype corresponding to a complex expression is predictable from the feature weights relative the parts of that expression, but features that are due to extensional feedback or are influenced by extensional feedback are obviously not predictable in the same way.

   A theory such as $t_8$ is not compatible with $C^{BU}$. Through the operationalization of what it means for a prototype to have a certain weighted feature, we know which prototypes that correspond to the relevant expressions. In addition we know, since a prototype is a kind of set, some things about how they are built up. The members of the prototypes corresponding to complex expressions are weighted features, not feature sets. So the *elements* of the prototype corresponding to a complex expression, are not the prototypes corresponding to the parts of that expression. Neither are the *subsets* of the prototype corresponding to a complex expression the prototypes corresponding to the parts of that expression; if a weighted feature is a matter of an ordered pair of a feature and a number, then any modification of feature weights would result in a $C^{BU}$ violation. Even if only the features (and not their weights) are assumed to be mereologically relevant, it is the case that the theory predicts that features are dropped and added as prototypes are combined, and it will not be the case that all the features of a prototype corresponding to a complex expression is a part of a prototype corresponding to one of the parts of that expression. Neither will it be the case that all the features of the prototypes corresponding to parts of a complex expression, are parts of the prototype corresponding to that expression. So we can conclude that it is part of the informative potential of $C^{BU}$ that meanings are not prototypes.[47] This conclusion is

---

[46] According to the model, a feature is not supposed to be part of a representation ('it fails to be inherited') if its weight (importance) is under a certain threshold value. Nonetheless, such features do play a role in subsequent combinations (by influencing feature drops due to perceived impossibility for instance).

[47] It is also part of the informative potential of $C^{FI}$ that meanings cannot be prototypes. But $C^{BU}$ also rules out prototypes on account of their behaviour in simple adjective-noun combinations. Smith and Osherson et al's (1988) selective modification model is also incompatible with $C^{BU}$, but this theory is in accordance with $C^{FI}$. Murphy (1988) and Wisniewski (1997) provide other models that are incompatible with $C^{BU}$. One could also note that the problems generalize to the

not restricted to the kind of prototypes, or the kind of composition presupposed by Hampton's model. It is a universal assumption of theories of prototype combination that the prototypes corresponding to the parts of a certain complex expression are 'merged' in order to arrive at the prototype for the complex expression. This assumption gives rise to $C^{BU}$ violations and all theories of prototype combination are thus in violation of $C^{BU}$.[48]

However, since we are not licensed to require that a semantic theory must accord with $C^{BU}$, we are not licensed in drawing the conclusion that meanings are not prototypes. The theory above is (and theories of prototype combination in general are) compatible with $K^{FI}$. This means that they are compatible with $D$. And since it is thus not part of the informative potential of $D$ that meanings are not prototypes, we do not know that meanings are not prototypes.

To sum up this subsection; applications of $C^{BU}$ suggest that it is part of the informative potential of $C^{BU}$ that meanings cannot be extensions or prototypes. However, we cannot conclude from this that meanings are not these things since it is not part of the informative potential of $D$ that meanings cannot be these things. We are thus in a situation with $C^{BU}$ that is similar to that we were in at the end of the previous subsections with $C^{FE}$ and $C^{FI}$: although it is part of the informative potential of the relevant explications of $C$ that meanings cannot be certain things, we have not by noticing this, learnt that meanings cannot be these things, since it does not follow from what we are justified in imposing on semantic theories that meanings cannot be these things.

---

feature based precursors of prototype theory namely those suggested by Katz and Fodor (1963) and Katz and Postal (1964).

[48] Although prototype theories have received a substantial amount of criticism based on their supposed non-compositionality, none of previous lines of criticism have been both convincing and such that they apply to the kind of prototypes described above. The classical sources of compositionality based criticism of prototypes are Osherson and Smith (1981) and Fodor and Lepore ([1996] 2002). But see also Connolly and Fodor et al. (2007) and Fodor and Lepore ([1991] 2002; [2001] 2002). For rebuttals and discussion see Jönsson and Hampton (2008) and Hampton and Jönsson (Forthcoming).

### *6.2 Concluding Remarks*

The results of each of the subsections converge. Although the explications of compositionality have substantial informative potential, this is rendered largely inconsequential by the conclusions of Chapter Three and Chapter Four. We are not justified in imposing the compositionality explications as constraints on semantic theories, and thus we do not learn anything about meanings from imposing them. Although compositionality, on some explication or other, seems to rule out that meanings are extensions, functions from possible worlds to extensions, fuzzy sets*, inferential roles or prototypes, these candidate meanings are all (with the possible exception of extensions) compatible with the weaker assumption *D*, which embodies what we are really justified in imposing.

We are thus in a position where we both are compelled to agree and to disagree with the claim that the compositionality constraint is 'the sovereign test' on semantic theories; on the one hand the informative potential of *C* is very rich, but on the other we can learn next to nothing about meaning from it.

*I must go on. I can't go on. I will go on.*
*– Samuel Beckett*

# 7

# Concluding Remarks

## *7.1 Where we are with Respect to Compositionality*

The goal of inquiry in this thesis has been to ascertain the extent to which the principle of compositionality can be justifiably imposed as a constraint on semantic theories and thereby provide information about what meanings are. It has been concluded that compositionality cannot – on any of the three explications $C^{FE}$, $C^{FI}$, or $C^{BU}$ – be justifiably imposed as a constraint on semantic theories, and that the information that it provides about meanings is very limited. Although this was where we ended up, several findings along the way suggested that there actually was a path from compositionality to meaning.

First, $C^{FE}$ (and thus the logically stronger $C^{FI}$ and $C^{BU}$) is a non-trivial constraint on semantic theories. $C^{FE}$ is not trivial *tout court* (i.e. it effects a proper partition of the semantic theories into compositional and non-compositional theories) and it seems likely that $C^{FE}$ is neither $\tau$-trivial nor $\varphi$-trivial (i.e. it seems likely that it is not the case that corresponding to each non-compositional theory there exists an equivalent compositional alternative, or that there are constraints stronger than $C^{FE}$ that we have good reason to believe all correct semantic theories must satisfy). These conclusions hold in

contexts, like the present one, where a semantic theory's accordance with certain syntactic and semantic data is important, in spite of the formal proofs, due to Zadrozny, Janssen, and van Benthem, that are sometimes used in order to support triviality claims about compositionality, and in spite of the arguments due to Horwich to the same effect. So up to this point, the potential threats to the idea that compositionality can be used as an interesting, substantial, constraint on semantic theories can be avoided.

Second, in addition to being able to rule out certain semantic theories, $C^{FE}$ and the other explications of compositionality can rule out certain hypotheses about what meaning is (and thereby potentially provide information about meaning). Given that the information we have about the way in which candidate meanings distribute over expressions (and, in the case of $C^{BU}$, about the structural relations between the candidate meanings) is reliable, there are valid arguments featuring the different compositionality explications as premises showing that some entities cannot be meanings. The combined informative potential of the three explications is such that, given a license to impose the explications as constraints on semantic theories, we are in a position to rule out, at least, the hypotheses that meanings are either extensions, functions from possible worlds to extensions, fuzzy sets*, inferential roles or prototypes. So at this point it seems that compositionality might serve as very revealing constraint.

Third, it seems that the incompatibility between compositionality and certain natural language phenomena has been overestimated, and that compositionality, at least on some explications, is compatible with a wide range of phenomena. $C^{BU}$ faces problems, as do thick versions of all the compositionality explications, but thin versions of $C^{FE}$ and $C^{FI}$ seem to be generally compatible with i) the existence of idioms, ii) essential (i.e. non-lexical, non-structural) ambiguity, iii) meaninglessness, iv) interactionism, and v) other forms of context sensitivity. These thin versions also appear to accord with satisfactory semantics for vi) propositional attitude reports and vii) quantified conditionals. So even after consideration of these phenomena, things look very promising for the compositionalist wishing to deploy compositionality in her inquiries into meaning.

From the perspective introduced by these reflections, then, it could well be the case that the principle of compositionality, explicated in

terms of $C^{FE}$ or $C^{FI}$, is true; and that it can function as a constraint from which we can learn much about meanings. However, as we have seen, things are much more bleak for the compositionalist than this lets on.

First, when, in search of a license to impose compositionality as a constraint on semantic theories, we look closer at the reasons that are usually given in support of compositionality, we find only reasons to affirm that *some* determination principle, but not necessarily compositionality is true. Each of several contrasting determination principles – $F^{FI}$, $X^{FI}$, $K^{FI}$ *or* $C^{FI}$ – is such that a semantic theory according with it, if it is correct and interpreted in the way detailed in the previous chapters, could potentially provide explanations of why speakers understand novel expressions, why they understand an infinite number of expressions, and why their understanding comes in chunks of syntactically related expressions. So a need to account for these explananda does not license one to reject non-compositional theories. Nor do compositional explanations seem preferable to non-compositional ones *in general*. We also noted that decisions about what constitutes the better explanation often have to be settled at the level of specific theories, where matters such as scope, simplicity, correctness, and the amount of understanding predicted, can be used as a basis of evaluation, not at the level of the determination claims. Whether one semantic theory is better than another depends on a range of factors, and we are not licensed to reject a semantic theory *qua* non-compositional theory.

Second, when we ask what considerations of linguistic creativity, productivity and systematicity *do* license, the answer – i.e. the requirement that semantic theories accord with $D$ – turns out to have informative potential falling well short of that of compositionality: few of the conclusions the compositionalist reaches by way of compositionality can be validly inferred in a similar way from corresponding arguments featuring $D$.

Our general conclusion, then, can be summed up in this way: $C^{FI}$ is the best explication of compositionality among those considered. It has important explanatory utility *and* it avoids problems facing stronger explications. Yet, in spite of its rich informative potential, we cannot learn much about meanings from $C^{FI}$, since we are not licensed to require that semantic theories must accord with it.

## 7.2 How to Go on

The next three subsections sketch reasonable responses to the position reached above. The approaches to be discussed are: i) using the Principle of Compositionality as a methodological constraint; ii) searching for additional data that show the imposition of the Principle to be justified after all; or iii) pursuing investigations into meaning independently of the Principle of Compositionality.

### 7.2.1 Settling for a Methodological Principle

We could continue to regulate semantic theorizing by compositionality in full awareness of that we lack justification for thinking that compositionality is true. This would not necessarily be unreasonable, although it is a research strategy from which we would not (at least, directly) learn anything about meanings. Since compositionality is a non-trivial constraint on semantic theories, it remains possible to use it as a *methodological principle*, i.e. a principle that is primarily employed to delimit an unwieldy hypothesis space. In cases where the space of hypotheses is simply too big to allow progress without some kind of restriction, one might argue that some methodological principle should be adopted so that researching becomes more feasible.

By treating compositionality as a methodological principle we can continue to use it to rule out hypotheses about what meanings are, then, but the motivation for using it in this way is now, in effect, something like an agreement among researchers to research in a certain way, and not that the truth of the principle has been established in any proper way.

To appreciate the way in which methodological principles are commonly used in scientific theorizing, consider the assumption that the threshold level of statistical significance is 0.05.[1] This assumption is often employed to refute or throw doubt on hypotheses: the null hypothesis that two samples (e.g. the IQ-scores of two groups of children) belong to the same population is frequently rejected if the calculated $p$-value is less than 0.05. Although this *modus operandi* is often employed, there does not seem to be any good justification for

---

[1] For the purposes of the illustration the precise threshold is unimportant: one could employ alternative levels such as 0.01 or 0.001.

holding that 0.05 is in general a better threshold of statistical significance than, say, 0.04. What seems to matter first and foremost is that the threshold should guard against error in a satisfactory way. But these considerations do not determine unique threshold values. It is thus a matter of convention which level, exactly, to adopt.

The assumption *that the threshold-level of statistical significance is 0.05* appears, then, to be methodological. Just as compositionality can be used in order to rule out certain hypothesis in the absence of proper justification, the assumption that the level of statistical significance is 0.05 is used to rule out hypotheses in the absence of proper justification. Both assumptions owe more to an agreement among researchers to proceed in a certain way than they do to the belief that they effect a partition between the true and the false.

For another example of a methodological assumption, consider the assumption often made in syntactic theory that branching is binary when it is coupled with the following argument in favor of it.

> A grammar which allows only binary branching nodes is more constrained than a grammar which allows any type of branching node: in the former type of grammar lots of imaginable representations are ruled out in principle […] If the ultimate goal of our grammar is to account for language acquisition, then it will be natural to aim for the more restricted type of grammar in which fewer decisions have to be made by the child.
>
> (Haegeman 1995, p. 139-143)

We can illustrate Haegeman's argument with the following example. If a child knows that branching must be binary, she knows, when exposed to (1), that the structure of (1) is either that of (1') or that of (1").[2]

(1)     John loves Mary.
(1')    [[John loves] Mary.]
(1")    [John [loves Mary.]]

---

[2] We assume for simplicity that only connected words can be part of the same constituent (e.g. in 'John loves Mary', 'John' and 'Mary' cannot both be part of a constituent which does not have 'loves' as part) and that all and only the surface words correspond to simple grammatical terms.

But if branching can be of any arity, then (1''') is also permissible, as are any number of structures involving unary branching.

(1''')   [[John ][loves ][Mary.]]

So, when branching can be of any arity it is harder to conjecture which mode of composition is the right one (since one has more alternatives to choose from) than it is when branching is binary. Hence, it would help the child trying to figure out how expressions are composed to know that branching is always binary.

   The foregoing is correct as far as it goes, but it is important to note that the argument as it stands supports, not binary branching per se, but binary branching over unconstrained branching.[3]

   Assume instead that only flat branching is permissible. Then the only possible mode of combination for (1) is (1'''). If the child knows that branching is always flat, things will be even easier for her (when it comes to assigning modes of composition). Flat branching might be unsuitable for generative purposes, but the point is simply that what is important from the perspective of minimizing structural hypotheses is merely that branching be *uniform*. Consider, for instance, the principle that branching is ternary where this is possible and binary otherwise. This is even more constraining than the binary branching hypothesis.[4] Indeed any principle that combines binary branching

---

[3] I do not mean to suggest that this is the only reason to think branching is binary. Far from it. I wish merely to point out the similarity, in terms of support, between this argument and binary branching, on the one hand, and the reasons discussed in Chapter Three and Chapter Four and compositionality, on the other.

[4] This is substantiated by the following. The number of structural hypotheses available for a sentence of length *n* where *n* >1 (we assume that $f(1)=1$) given the binary branching constraint is captured by the following formula.

$$f(n) = \sum_{i=1}^{n-1} f(i)f(n-i)$$

The number of structural hypotheses available for a sentence of length *n* where n > 2 (assuming that $f(1) = 1$, and $f(2) = 1$) given the restriction that branching is ternary if possible and otherwise binary is captured by the following formula.

with some other form of higher arity branching will be more constraining than the binary branching hypothesis. From the perspective of minimizing hypotheses, therefore, any such principle is better.

To deploy the hypothesis that branching is binary, in the absence of further justification, to delimit the space of possible syntactic theories is to treat the principle that branching is binary as a methodological principle.[5] The analogy with compositionality as a methodological constraint should be clear by now. In both cases we assume that theorizing should be regulated by a certain principle even though the principle is really underdetermined by the explanations that we want to provide. Imposition of the principles has to be seen (so long as no further justification is provided) as a kind of paradigm-defining restriction on how to proceed rather than as a warranted exclusion of inept alternatives.

Although this is seldom spelled out explicitly, it seems likely that some researchers *do* in fact treat compositionality as a methodological assumption rather than a well supported conclusion. We can turn, for

$$f(n) = \sum_{i\,=\,1}^{n-2} \sum_{j\,=\,i+1}^{n-1} f(i)f(j\text{-}i)f(n\text{-}j)$$

The number of structural hypotheses available for a sentence of length $n$ given the uniform flat branching constraint is, of course, captured by the following formula.

$$f(n) = 1$$

[5] As an example of how the binary branching assumption can be used to rule out hypotheses consider the following argument due to Kayne (1984: ix-x). If adverbs can be inserted between verbs and their direct objects, then there are expressions of the surface form [verb]-[adverb]-[noun phrase] (e.g. 'hugged carefully john'). These expressions must correspond to one of the following three grammatical terms.

(2)      [[V adv] NP]
(2')     [V [adv NP]]
(2")     [V adv NP]

However, given the assumptions that i) NPs must be assigned thematic roles, ii) V's can only assign thematic roles to their sisters, and iii) branching must be binary, none of these grammatical terms is legitimate. (2) and (2') are ruled out by i) and ii), and (2") is ruled out by iii). So, given that branching is assumed to be binary (and assuming i) and ii)), the hypothesis that adverbs can be inserted between verbs and their direct objects can be ruled out.

instance, to Davidson, in his seminal 'Truth and Meaning'. It seems that he was not completely convinced that the compositional avenue of research he suggested was the only one.

> [A] theory of meaning for a language *L* shows "how the meanings of sentences depend upon the meanings of words" if it contains a (recursive) definition of truth-in-L. *And so far at least,* we have no other idea how to turn the trick.
>                    (Davidson [1967] 2001: 101, emphasis added)

As the emphasized caveat suggests, Davidson did not rule out alternative accounts. He saw himself as describing (and recommending) one way – the only way he knew of at the time – in which the desiderata of a semantic theory could be provided.

### 7.2.2 Making Headway by Recruiting More Assumptions

One might be dissatisfied with this methodological approach – with its willingness to root the deployment of compositionality in agreement between researchers rather than compelling evidence about matters of fact. Feeling this way, one might instead react to the conclusions reached in this thesis by attempting to ground additional assumptions that render the imposition of compositionality on semantic theorizing legitimate.

We have seen that possible explanations of linguistic creativity, productivity and systematicity all include reference to the meaning of parts of complex expressions, and to their mode of composition, but that the non-compositional explanation also made reference to additional determinants. So if we can obtain plausible hypotheses about speakers that limit the number of possible determinants, we might actually be in a position to justify the imposition of compositionality, as a constraint, on semantic theories and thus learn something about meanings from it.

This approach is illustrated by the thesis that the mind is modular – a thesis famously presented by Jerry Fodor in 'The modularity of Mind'. This alternative was briefly considered in Chapter Three. It was dismissed there, since the available data suggested that the kinds of meaning assignment that concerned us (those pertaining to synonymy and ambiguity judgments) could not issue from modules. However, the modularity thesis might still serve to illustrate a

reasonable *kind* of response to the findings of the previous chapters since it is fairly clear how it would remove potential determinants from consideration.

The modularity thesis has recently been recruited by Emma Borg to argue in favor of a formal, compositional, Davidsonian truth-conditional semantics, for a purpose very similar to the one we are interested in here. So her discussion of the position might be a good point of departure. On Borg's account a *module* is an 'encapsulated body of information, together with processes operating only over that information, which is responsible for realizing a cognitive function' (Borg 2004: 80). In the sense of 'module' used by Borg, it is maintained that the five senses correspond to one or more modules. The claim, then, that linguistic comprehension, or grasp of literal linguistic meaning (understood as something much more narrow than our full communicative ability), is modular is the claim that the assignment of meanings to expressions involves applying a limited set of rules to a proper subset (only) of the information that is accessible to the mind as a whole.

Borg follows Fodor in arguing that, given that a capacity exhibits a sufficient number of certain features, it should be given a modular explanation, and goes on to argue that linguistic comprehension exhibits all the relevant features. She argues, to mention just a few of these features, that linguistic comprehension is *domain specific* (i.e. such that only a small subset, the expressions, of the stimuli that the speaker is exposed to  triggers it), *mandatory* (such that a speaker cannot, upon being exposed to an expression, normally help but assign a meaning to it), *very fast* and *exhibits specific patterns of breakdown and acquisition* (such that a speaker can acquire, and lose, the ability to linguistically comprehend independently of the addition or loss of many other capacities). Although this is only a subset of the phenomena discussed by Borg, and only an indication of what these amount to, the essence of the argument for the modularity of linguistic comprehension is, I hope, clear.

It appears that once adopted Modularity can be used to delimit the number of possible determinants of the meanings of complex expressions. Modularity thus appears to undermine some of the explanations discussed in Chapter Three and Chapter Four. Consider, for instance, semantic theories involving the general knowledge of the speakers they are about – that is, the $K^{FI}$ theories

that are not in accord with $C^{FI}$. Since a speaker's general knowledge concern matters of fact in the world, rather than specifically linguistic or semantic subject matter, it seems that reference to such knowledge is ruled out if modularity is accepted. Consider also the $X^{FI}$ theories that are not in accord with $C^{FI}$. According to these, the speaker's dispositions to attend to certain things in conversational contexts are involved in his or her semantic competence. But the case that we considered in connection with the $X^{FI}$ theories (the unarticulated constituent analysis of 'it rains') might feature *abductive* inference. Determining what location an utterance of 'it rains' pertains to seems to be a clear example of inference to the best explanation. Hence, any piece of evidence might bear on the inference. But if linguistic comprehension is modular, not all kinds of information are accessible in the underlying processing. So, many $X^{FI}$ theories will also be ruled out by the modularity assumption simply as theories presupposing abductive inference.[6]

When modularity is assumed, therefore, the space of possible explanations of linguistic creativity, productivity and systematicity seems to contract. Importantly, since our resistance to some of the arguments in Chapter Six depended essentially on the acceptability of $K^{FI}$ theories, some of the arguments against certain meaning theories (those invoking fuzzy sets* or inferential roles) now go through. However, Modularity itself (together with considerations of linguistic creativity and so on) does not license the imposition of $C^{FI}$ on semantic theories, since, for instance, $F^{FI}$ theories seem to be compatible with modularity. However, we do appear, at this juncture, to know more about meanings, given that modularity can be assumed.

Although, as has already been mentioned, it seems likely that the modularity thesis fails with respect to the kinds of meaning that we have been concerned with in this thesis, I meant here only to clarify a general strategy, the strategy of identifying plausible assumptions that delimit the possible determinants of the meanings of complex expressions, in order to license the imposition of a compositionality constraint on semantic theories.

---

[6] This line of reasoning is recruited by Borg in her criticism of what she calls a 'dual pragmatic account'.

### 7.2.3 Leaving Compositionality Behind

One might, of course, take the entirely different path of pursuing inquiries into meaning in complete disregard of whether or not they respect compositionality. One might be dissatisfied with the first approach on the grounds that it does not really give us any information about meanings, yet be unhappy, also, with the second approach, because one is skeptical about efforts to identify empirical assumptions capable of licensing the compositionality requirement. Approaching meaning in a less presumptuous way, by dropping the requirement that our account accords with this or that determination principle, and by instead making use of resources that seem to be explanatorily motivated, seems to me to be the most compelling way to proceed.

As an illustration of this third approach, consider the development of the Composite Prototype Model in cognitive psychology.[7] Certain features of relative clause conjunctions – in particular, that they exhibit overextension and dominance – demand explanation. As we saw in Chapter Six, Hampton accounts for the semantics of these constructions in terms of combinations of certain kinds of prototype. The resources (such as 'extensional feedback') that the model needs in order to get the predictions right are included in the model. The compelling explanation that ensues happen to depend on a non-compositional model, but since it is compatible with patent facts about speakers (such as their being linguistically creative) this should not be a source of serious concern.

It is unfortunate that the project of accounting for our fascinating ability to speak of things we have never encountered with phrases we have never before heard has become near synonymous with respecting compositionality, for the space of options is much larger than this equivocation suggests.

---

[7] This model is due to Hampton (1987). It was detailed in Chapter Six.

# Appendix

# Notation, Definitions and Principles

---

### *Basic Definitions of Set theory and Formal Semantics*

| | |
|---|---|
| {a, b, c,…} | The *set* containing *a*, *b*, *c* etc. The things contained in a set are its *elements* (or *members*). E.g. {Lisa, Carl, John} is the set containing the elements Lisa, Carl and John. The order in which the elements are listed is immaterial to the identity of the set. Thus: {John, Lisa, Carl} = {Carl, Lisa, John}. |
| {*x* : …*x*…} | The set containing each element *x* such that …*x*… (i.e. the set containing those elements that meet the criterion specified after the colon). Thus: {*x* : *x* is a bear} is the set of all bears. |
| ∅ | *The empty set*, i.e. the set without any elements. |
| *a* ∈ *A* | The statement that *a is an element of the set A*. |
| *A* ⊆ *B* | The statement that (the set) *A is a subset of* (the set) *B*. *A* is a subset of *B* |

iff every element of *A* is an element of *B*. Note that this means for all sets *X* (including $\varnothing$), $\varnothing \subseteq X$.

$A \cup B$      *The union of the sets A and B*, i.e. the set containing exactly those elements that are either elements of *A* or *B* (or both).

$A \cap B$      *The intersection of the sets A and B*, i.e. the set containing exactly those elements that are elements of both *A* and *B*.

*A-B*      *The relative complement of B in A,* i.e. the set containing exactly those members of *A* that are not members of *B*.

$\wp A$      *The power set of the set A*, i.e. the set of all subsets of *A*.

$<a_1, \ldots, a_n>$      *An ordered n-tuple*, i.e. a structure of *n* elements where the order in which the elements are given matters to the identity of the tuple, thus <John, Lisa> $\neq$ <Lisa, John>. A 2-tuple is also referred to as an *ordered pair*.

$A_1 \times \ldots \times A_n$      *The cartesian product of the sets $A_1 \ldots A_n$* i.e. the set $\{< a_1, \ldots, a_n> : a_1 \in A_1 \& \ldots \& a_n \in A_n\}$.

$A^n$      *The cartesian product corresponding to* $A \times \ldots \times A$ where *A* occurs *n* times.

$g : A \rightarrow B$      *A function from the members of the set A to the members of the set B*, i.e. a set

of ordered pairs such that for each ordered pair $<x, y> \in g$, it is not the case that $<x, z> \in g$ if $y \neq z$.

$g(x)$ — *The value of a function $f$ given $x$ as argument.* If $<x, y> \in g$, then $g(x)=y$.

$\text{dom}(g)$ — *The domain of a function $g$*, i.e. the set $\{x : \exists y, <x, y> \in g\}$.

$\text{ran}(g)$ — *The range of a function $g$*, i.e. the set $\{y : \exists x, <x, y> \in g\}$.

A *partial* function — *A function $g$ is partial relative to some set $C$ iff $\text{Dom}(g) \subseteq C$.*

A *total* function — *A function $g$ is total relative to some set $C$ iff $\text{Dom}(g) = C$. Note that this means that all functions that are total relative a set $C$ are also partial relative to $C$.*

A *surjective* function — A function $f : A \to B$ is surjective if $\text{ran}(f)=B$

$\lambda x_{1} \in X_{1} \ldots , x_{n} \in X_{n}.\ \phi$ — This is a concise way to describe functions. A description of this form is ambiguous between denoting

i)     the function from elements of the sets $X_{1} \ldots X_{n}$ to the truth-value true, if $\phi$ is true given the arguments, and to the truth-value false if $\phi$ is false given the arguments.
E.g. '$\lambda p \in P.$ p smokes' describes – given that $P$ is the set of persons – the function from a person to the truth-value true if

that person smokes, and to the truth value false if that person does not smoke.

ii) the function from elements of the sets $X_1 \ldots X_n$ to the values of $\phi$ given the arguments.
E.g. '$\lambda n_1 \in N,\ n_2 \in N.\ n_1 + n_2$' describes – given that $N$ is a set of numbers – the function from two numbers to their sum.

Whether the first or the second interpretation is intended can always be decided on the basis of whether $\phi$ is truth-evaluable or not.

| | |
|---|---|
| An *operation* | *An n-ary operation on a set A is a function $f: A^n \to A$.* |
| An *algebra* | An ordered pair <*A,O*> of a set *A* and an ordered n-tuple *O* of (total) operations on *A*. |
| *A partial algebra* | An ordered pair <*A,O*> of a set *A* and ordered n-tuple *O* of partial operations on *A*. |
| $E_1 \cdot E_2$ | The concatenation of expressions $E_1$ and $E_2$. E.g. 'red'·'car'= 'red car'.[1] |

---

[1] The notation used in this section is due to Enderton (1977), Grätzer (1979) and Heim and Kratzer (1998).

## Definitions of the Core Concepts of the Thesis

$Ę = \langle E_Ę, A_Ę, \Sigma_Ę \rangle$   *A partial syntactic algebra (a grammar)*

- $E_Ę$ is a set of expressions.
- $A_Ę$ is a set of expressions (*the simple expressions*), $A_Ę \subseteq E_Ę$.
- $\Sigma_Ę$ is a set of partial operations $\sigma_0, \ldots, \sigma_{n-1}$ (*the syntactic modes of composition*) from expressions to expressions.

  $A_Ę$ and $\Sigma_Ę$ are both finite, $E_Ę$ can be either finite or infinite. $E_Ę$ is the closure of $A_Ę$ under $\Sigma_Ę$. The members of $E_Ę$-$A_Ę$ are *complex* expressions.

$G(Ę)$   *The set of grammatical terms for a syntactic algebra* $Ę$, i.e. the set s.t.

i)  if $e \in A_Ę$, then $e \in G(Ę)$.
ii) if $\sigma \in \Sigma_Ę$, $e_0, \ldots, e_{n-1} \in G(Ę)$, and $\sigma$ is defined for $e_0, \ldots, e_{n-1}$, then '$\sigma(e_0, \ldots, e_{n-1})$' $\in G(Ę)$ (*the complex grammatical terms*).

When the syntactic algebra is clear from context, '$G(Ę)$' will be abbreviated '$G$'.

$val: G(Ę) \rightarrow E_Ę$   A total surjective function from grammatical terms to expressions s.t.

i)  if $e \in G(Ę)$ and $e \in A_Ę$, $val(e) = e$.
ii) if $e \in G(Ę)$ and $e =$'$\sigma(e_0, \ldots, e_{n-1})$', $val(e) = \sigma(e_0, \ldots, e_{n-1})$.

$M$   *A set of meanings*

| | |
|---|---|
| $\boldsymbol{M} = \wp M$ | *A set of semantic ranges* |
| $\mu': G \to \boldsymbol{M}$ | *An interpretation function*, i.e. a total function from grammatical terms to semantic ranges |
| A *language* | *A language* is an ordered pair $\langle \underline{E}, \mu \rangle$ consisting of a syntactic algebra $\underline{E}$ and an interpretation function $\mu'$. |
| An *explicit syntactic rule* | An expression of the form '$E \to e_1,\dots,e_n$' where '$E$' is a variable over expressions, and '$e_1$',…, '$e_n$' are names of expressions. |
| An *implicit syntactic rule* | An expression of the form '$E_0 \to E_1\cdot,\dots,\cdot E_n$' where '$E_0$',…,'$E_n$' are variables over expressions. |
| An *explicit semantic rule* | An expression of the form '$\mu(e_1) = \phi$' where '$e_1$' is the name of an expression and '$\phi$' is a name or a description of a meaning. |
| An *implicit semantic rule* | An expression of the form '$\mu(_{E0} E_1,\dots, E_n) = \phi$' where '$E_0$',…,'$E_n$' are variables over expressions, and '$\phi$' is a name or a description possibly featuring '$E_0$',…,'$E_n$'. |
| | Note that while the semantic rules are stated in terms of '$\mu$', the interpretation function is denoted by '$\mu'$'. This marks the distinction between meaning and semantic range assignments. The relationship between the semantic rules and $\mu'$, is spelled out by the conditions on the semantic rules that are part of the |

definition of a semantic theory (see below). Similarly, the relationship between the syntactic rules and $\underline{E}$ is spelled out by the conditions on the syntactic rules that are also part of the definition of a semantic theory.

A *rule*

*A rule* is either an explicit syntactic rule, an implicit syntactic rule, an explicit semantic rule or an implicit semantic rule.

$\Psi$

*A finite set of rules*

- $\underline{\Sigma}$ is the subset of $\Psi$ of the *syntactic* rules
- $\underline{\Sigma}^e$ is the subset of $\Psi$ of the *explicit syntactic* rules.
- $\underline{\Sigma}^i$ is the subset of $\Psi$ of the *implicit syntactic* rules.
- $\underline{\mu}$ is the subset of $\Psi$ of the *semantic* rules
- $\underline{\mu}^e$ is the subset of $\Psi$ of the *explicit semantic* rules.
- $\underline{\mu}^i$ is the subset of $\Psi$ of the *implicit semantic* rules.

A *Semantic Theory*

*An n-tuple t=<$\underline{E}$, $\mu$', $\Psi$,…>* the first three members of which are a syntactic algebra $\underline{E}$, an interpretation function $\mu$', and a rule set $\Psi$ which accords with the following syntactic and semantic constraints.

A *Semantic Theory* (*continued*)  *Constraints on the Syntactic Rules:*

i) For each implicit syntactic rule $r = {}^\prime E_0 \rightarrow E_1 , \ldots , E_n {}^\prime \in \Psi$ there is a sequence of 0 or more *l-substitutions* by which $r$ can be transformed into a statement of the form ${}^\prime E_0 \rightarrow Y_1 \cdot \ldots \cdot Y_m {}^\prime$ such that there for each $Y_i$ exist an explicit syntactic rule $r' = {}^\prime Y_i \rightarrow e_1 , \ldots , e_k {}^\prime \in \Psi$.

An *l-substitution* is the replacement of an occurrence of a variable $x$ with ${}^\prime x_1 \cdot \ldots \cdot x_p {}^\prime$ if there is an implicit syntactical rule $r' = {}^\prime x \rightarrow x_1 \cdot \ldots \cdot x_p {}^\prime \in \Psi$.

ii) If there is an explicit syntactic rule ${}^\prime E \rightarrow e_1 , \ldots , e_n {}^\prime \in \Psi$, $e_1 , \ldots , e_n \in A_E$.

iii) If an expression $e \in A_E$, there is an explicit syntactic rule ${}^\prime E \rightarrow \ldots e \ldots {}^\prime \in \Psi$.

iv) There is a 1-1 function $g : \underline{\Sigma}^i \rightarrow \Sigma_E$ s.t.

for each $r = {}^\prime X \rightarrow X_1 \cdot \ldots \cdot X_n {}^\prime \in \underline{\Sigma}^i$, $g(r)$ is of arity $n$, and $g(r)$ is defined exactly for those $x_1 , \ldots , x_n$ such that for each $x_i$, where $1 \leq i \leq n$, there is either

1) an implicit syntactic rule $r' = {}^\prime X_i \rightarrow Y_1 \cdot \ldots \cdot Y_m {}^\prime \in \Psi$ s.t. $x_i \in \operatorname{ran}(g(r'))$, or

A *Semantic Theory* (*continued*)

2) an explicit syntactic rule
$r'=$ '$X_i \rightarrow y_1, \ldots, y_m$' $\in \Psi$
and $\exists j$ s.t. $y_j = x_i$

*Constraint on the Semantic Rules:*

v) For each implicit semantic rule
$r=$ '$\mu(_{E0}\ E_1, \ldots,\ E_n) = \phi$' $\in \Psi$
there is an implicit syntactic
rule $r'=$ '$E_0 \rightarrow E_1 \cdot \ldots \cdot E_n$' $\in \Psi$

vi) For each explicit semantic rule
'$\mu(e_1) = \phi$' $\in \Psi$ it is the case
that $\phi \in \mu'(e_1)$.

vii) For each grammatical term
$e \in G$, the statements
'$\mu(e)=x_1$',...,'$\mu(e)=x_n$' that are
possible to arrive at by a
sequence of 0 or more licensed
substitutions on the semantic
rules of $\Psi$, are such that
$\mu'(e) = \{x_1\} \cup \ldots \cup \{x_n\}$.

The following are *licensed
substitutions*

1) If $\Psi$ contains an implicit
syntactic rule $r=$ '$E_0 \rightarrow$
$E_1 \cdot \ldots \cdot E_n$', and an
implicit semantic rule
$r'=$ '$\mu(_{E0}\ E_1, \ldots,\ E_n) = \phi$', a
left-hand side occur-
ence of '$E_0$' can be
replaced by '$E_1 \cdot \ldots \cdot E_n$' if a
corresponding right
hand occurrence of '$E_0$'
is replaced by '$\phi$'.

A *Semantic Theory* (*continued*)

2) If $\Psi$ contains an explicit syntactic rule $r=$ '$E\rightarrow \ldots e_i\ldots$', and an explicit semantic rule $r'=$ '$\mu(e_i)=$ $\phi$', a left-hand side occurrence of '$E$' can be replaced by '$e_i$' if the corresponding right hand occurrence of '$E$' is replaced by '$\phi$'.

3) An expression containing concatenation marks can be replaced by the same expression without those marks.

# *Principles*

**(C)** The meaning of a complex expression is determined by the meanings of its parts and its mode of composition.

**($C^{FE}$)** A semantic theory accords with $C^{FE}$ iff

there is a function $r$ such that for every complex term $s=\sigma(e_0,\ldots,e_{n-1})$,

$$\mu'(s) = r(\sigma, \mu'(e_0),\ldots, \mu'(e_{n-1})).$$

**($C^{FI}$)** A semantic theory $t$ accords with $C^{FI}$ iff

    i)    There are no explicit semantic rules for complex expressions in $t$.

    ii)   Each implicit semantic rule in $t$ is of the form '$\mu(x)=y$' where

        1)   '$x$' is a mode of composition variable, and

        2)   '$y$' is a *c-term*, where a *c-term* is defined in the following way:

            a)   '$\mu(x_1)$' is a *c-term* if '$x_1$' is one of the variables from which '$x$' is built up.

            b)   '$g(x_1,\ldots, x_n)$' is a *c-term* if '$x_1$',…,'$x_n$' are *c*-terms and '$g$' is a functor corresponding to a humanly computable function defined on the set of meanings.

**($C^{BU}$)** A theory $t$ accords with $C^{BU}$ iff

    i)    There are no explicit semantic rules for complex expressions in $t$.

    ii)   Each implicit semantic rule in $t$ is of the form '$\mu(x)=y$' where

1) '$x$' is a mode of composition variable, and
2) '$y$' is a *c-term*, where a *c-term* is defined in the following way:

     a) '$\mu(x_1)$' is a *c*-term if '$x_1$' is one of the variables from which '$x$' is built up.
     b) '$f(x_1,\ldots, x_n)$' is a *c*-term if '$x_1$',…,'$x_n$' are *c*-terms and '$f$' is a functor corresponding to a humanly computable function defined on the set of meanings.

iii) For each complex grammatical term $e=\sigma(e_0,\ldots,e_{n-1})$ and each meaning $m\in\mu'(e)$ it is the case that $m$ is a structure built up from the parts $m_0,\ldots,m_{n-1}$, where $m_0\in\mu'(e_0),\ldots,m_{n-1}\in\mu'(e_{n-1})$, and for no two meanings $m$, $m'\in\mu'(e)$ it is the case that $m$ and $m'$ is built up from the same parts.

iv) For each syntactic operation $\sigma$, and for each sequence of expressions $e_0,\ldots, e_{n-1}$ for which $\sigma$ is defined, given that $\sigma(e_0,\ldots,e_{n-1})=e$ then for each $n$-tuple $m_0\in\mu'(e_0),\ldots,m_{n-1}\in\mu'(e_{n-1})$ it is the case that there is a meaning $m\in\mu'(e)$ which is a structure built up from the parts $m_0,\ldots,m_{n-1}$.

**(F$^{FI}$)** A semantic theory $t$ accords with $F^{FI}$ iff

i) There are no explicit semantic rules for complex expressions in $t$.
ii) Each implicit semantic rule in $t$ is of the form '$\mu(x)=y$' where

     1) '$x$' is a variable over expressions, and
     2) '$y$' is a *f-term*, where a *f-term* is defined in the following way:

         a) '$\mu(x_1)$' is a *f*-term if '$x_1$' is one of the variables from which '$x$' is built up.

b) '$x_I$' is a *f*-term if '$x_I$' is one of the variables from which '$x$' is built up.

c) '$g(x_1,…, x_n)$' is a *f*-term if '$x_1$',…,'$x_n$' are *f*-terms and '$g$' is a functor corresponding to a humanly computable function defined on meanings and/or expressions.

**($X^{FI}$)** A semantic theory *t* accords with $X^{FI}$ iff

i) There are no explicit semantic rules for complex expressions in *t*.

ii) Each implicit semantic rule in *t* is of the form '$\mu(x)=y$' where

1) '$x$' is a mode of composition variable, and

2) '$y$' is a *x-term*, where a *x-term* is defined in the following way:

a) '$\mu(x_I)$' is a *x*-term if '$x_I$' is one of the variables from which '$x$' is built up.

b) '$c$' is a *x*-term if '$c$' refers to a contextual function.

c) '$g(x_1,…, x_n)$' is a *x*-term if '$x_1$',…,'$x_n$' are *x*-terms and '$g$' is a functor corresponding to a humanly computable function defined on the set of meanings and/or a set of contextual functions.

**($K^{FI}$)** A semantic theory *t* accords with $K^{FI}$ iff

i) There are no explicit semantic rules for complex expressions in *t*.

ii) Each implicit semantic rule in *t* is of the form '$\mu(x)=y$' where

1) '$x$' is a mode of composition variable, and

2) '*y*' is a *k-term*, where a *k-term* is defined in the following way:

   a) '$\mu(x_1)$' is a *k*-term if '$x_1$' is one of the variables from which '*x*' is built up.
   b) '*b*' is a *k*-term if '*b*' refers to a state of general knowledge.
   c) '$g(x_1,\ldots, x_n)$' is a *k*-term if '$x_1$',…,'$x_n$' are *k*-terms and '*g*' is a functor corresponding to a humanly computable function defined on the set of meanings and/or a set of general knowledge.

**(D)** A semantic theory *t* accords with *D* iff *t* accords with $C^{FI}$, $F^{FI}$, $X^{FI}$, or $K^{FI}$.

The Entailment Relations between the Principles:

$$C^{FE} \longleftarrow C^{FI} \longleftarrow C^{BU}$$

$$F^{FI} \qquad X^{FI} \qquad K^{FI}$$

$$D$$
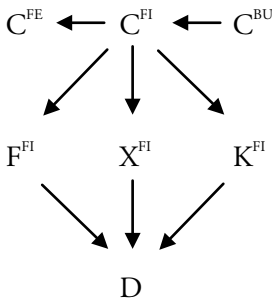
# References

Aizawa, K. (1997a). "Exhibiting versus Explaining Systematicity: A Reply to Hadley and Hayward." <u>Minds and Machines</u> **7**: 39-55.

Aizawa, K. (1997b). "Explaining Systematicity." <u>Mind & Language</u> **12**(2): 115-136.

Aizawa, K. (1997c). The role of the systematicity argument in classicism and connectionism. <u>Two Sciences of Mind</u>. S. O'Nualláin, P. McKevitt and E. MacAogáin. Amsterdam, The Netherlands, John Benjamins Publishing Company: 197-218.

Aizawa, K. (2003a). Cognitive Architecture: The Structure of Cognitive Representations. <u>Blackwells Companion to the Mind</u>. T. Warfield and S. Stich: 172-189.

Aizawa, K. (2003b). <u>The Systematicity Arguments</u>. Boston, United States, Kluwer.

Armstrong, D. M. (1978). <u>Universals and Scientific Realism (II): A Theory of Universals</u>. Cambridge, Cambridge University Press.

Ayer, A. J. (1946). <u>Language, Truth and Logic 2:nd edition</u>. London, Victor Gollancz Ltd.

Bach, K. (1994). Semantic Slack: What is Said and More. <u>Foundations of Speech Act Theory</u>. S. L. Tsohatzidis, Routledge: 267-291.

Barwise, J. and J. Perry (1983). <u>Situations and Attitudes</u>. Cambridge, Massachusetts, The M.I.T. Press.

Block, N. (1993). "Holism, Hyper-Analyticity and Hyper-Compositionality." <u>Philosophical Issues</u> **3**: 37-72.

Boghossian, P. A. (1993). "Does an Inferential Role Semantics Rest upon a Mistake?" <u>Philosophical Issues</u> **3**: 73-88.

Boolos, G. S., J. P. Burgess and R. C. Jeffrey (2002). <u>Computability and Logic</u>. New York, Cambridge University Press.

Borg, E. (2004). <u>Minimal Semantics</u>, Oxford, Oxford University Press.

Cappelen, H. and E. Lepore (2005). "Varieties of Quotation Revisited." <u>Belgian Journal of Linguistics</u>.**17**: 51-75.

Carnap, R. (1947). <u>Meaning and Necessity</u>. Chicago, University of Chicago Press.

Carnie, A. (2002). <u>Syntax: A Generative Introduction</u>. Malden, Massachusetts, Blackwell Publishing Ltd.

Chalmers, D. (1990a). "Syntactic Transformations on Distributed Representations." <u>Connection Science</u> **2**(1&2): 53-62.

Chalmers, D. (1990b). <u>Why Fodor and Pylyshyn Were Wrong: The Simplest Refutation</u>. The Twelfth Annual Conference of the Cognitive Science Society.

Chierchia, G. and S. McConnell-Ginet (1990). <u>Meaning and Grammar: An Introduction to Semantics</u>, Cambridge, The Mit Press.

Chomsky, N. (1957). <u>Syntactic Structures</u>. The Hague, Mouton & Co.

Chomsky, N. (1965). <u>Aspects of the Theory of Syntax</u>. Cambridge, Massachusetts, The Mit Press.

Chomsky, N. (1966). <u>Topics in the Theory of Generative Grammar</u>. Paris, Mouton & Co.

Chomsky, N. (1986). <u>Knowledge of Language: Its Nature, Origin, and Use</u>. London, Praeger.

Church, A. (1951). <u>The Calculi of Lambda-Conversion</u>. Princeton, Princeton University Press.

Cohen, L. J. (1986). "How is Conceptual Innovation Possible?" <u>Erkenntnis</u> **25**: 221-238.

Cohnitz, D. (2005). Is compositionality an a priori principle? <u>The Compositionality of Meaning and Content. Volume 1: Foundational Issues</u>. M. Werning, E. Machery and G. Schurz. Frankfurt, Ontos Verlag**:** 23-58.

Collins, J. (2003). "Horwich's Schemata Meet Syntactic Structures." <u>Mind</u> **112**(447): 399-431.

Connolly, A. C., J. A. Fodor, L. R. Gleitman and H. Gleitman (2007). "Why Stereotypes Don't even Make Good Defaults." Cognition **103**(1): 1-22.

Crimmins, M. and J. Perry (1989). "The Prince and the Phone Booth: Reporting Puzzling Beliefs." The Journal of Philosophy **86**(12): 685-711.

Cummins, R. (1996). "Systematicity." Journal of Philosophy **93**: 591-614.

Davidson, D. ([1965] 2001). Theories of Meaning and Learnable Languages. Inquiries into Truth and Interpretation. Oxford, Clarendon Press.

Davidson, D. ([1967] 2001). Truth and Meaning. Inquiries into Truth and Interpretation. D. Davidson. Oxford, Clarendon Press: 17-36.

Davidson, D. ([1968] 2001). On Saying That. Inquiries into Truth and Interpretation. D. Davidson. Oxford, Clarendon Press: 93-108.

Davidson, D. ([1970] 2001). Semantics for Natural Language. Inquiries into Truth and Interpretation. D. Davidson. Oxford, Clarenden Press.

Davidson, D. ([1979] 2001). Quotation. Inquiries into Truth and Interpretation. D. Davidson. Oxford, Clarendon Press: 79-92.

Davidson, D. ([1992] 2001). The second person. Subjective, Intersubjective, Objective. D. Davidson. Oxford, Oxford University Press: 107-122.

Davies, M. (1987). "Tacit Knowledge and Semantic Theory: Can a Five per cent Difference Matter?" Mind **96**(384): 441-461.

Davis, S. and Gillon, B. S. (2004). Introduction. Semantics. S. Davis and B. S. Gillon. Oxford, Oxford University Press: 3-130.

Dever, J. (1999). "Compositionality as Methodology." Linguistics and Philosophy **22**: 311-326.

Dever, J. (2006). Compositionality. The Oxford Handbook of Philosophy of Language. E. Lepore and B. Smith. Oxford, Oxford University Press: 633-666.

Enderton, H. B. (1977). Elements of Set Theory. London, Academic Press.

Enderton, H. B. (2001). <u>A Mathematical Introduction to Logic</u>. Burlington, Harcourt/Academic Press.

Evans, G. ([1981] 1985). Semantic Theory and Tacit Knowledge. <u>Collected Papers</u>. G. Evans. Oxford, Clarenden Press: 322-342.

Fernando, T. (2001). "Ambiguous Discourse in a Compositional Context. An Operational Perspective." <u>Journal of Logic, Language, and Information</u> **10**: 63-86.

Fodor, J. A. (1983). <u>The Modularity of Mind</u>. Cambridge, Massachusetts, Mit Press.

Fodor, J. A. (1987). <u>Psychosemantics</u>. Cambridge, MA, MIT Press.

Fodor, J. A. (1998a). <u>Concepts</u>, Mit Press.

Fodor, J. A. (1998b). There are No Recognitional Concepts-Not Even RED. <u>In Critical Condition</u>. J. A. Fodor. London, England, The MIT Press: 35-47.

Fodor, J. A. (1998c). There are No Recognitional Concepts-Not Even RED, Part 2: The Plot Thickens. <u>In Critical Condition</u>. J. A. Fodor. London, England, The MIT Press: 49-62.

Fodor, J. A. (2001). "Language, Thought and Compositionality." <u>Mind and Language</u> **16**(1): 1-15.

Fodor, J. A. and E. Lepore (2002). Geographical Introduction. <u>The Compositionality Papers</u>. J. A. Fodor and E. Lepore. Oxford, Clarendon Press: 1-6.

Fodor, J. A. and E. Lepore ([1991] 2002). Why Meaning (Probably) Isn't Conceptual Role. <u>The Compositionality Papers</u>. J. A. Fodor and E. Lepore. Oxford, Clarenden Press: 9-26.

Fodor, J. A. and E. Lepore ([1996] 2002). The Red Herring and the Pet Fish: Why Concepts Still Can't Be Prototypes. <u>The Compositionality Papers</u>. J. A. Fodor and E. Lepore. Oxford, Clarendon Press: 27-42.

Fodor, J. A. and E. Lepore ([2001] 2002). Why Compositionality Won't go away: Reflections on Horwich's 'Deflationary' Theory. <u>The Compositionality Papers</u>. J. A. Fodor and E. Lepore, Oxford University Press.

Fodor, J. A. and B. McLaughlin (1990). "Connectionism and the problem of systematicity: Why Smolenesky's solution dosn't work." <u>Cognition</u> **35**(2): 183-204.

Fodor, J. A. and Z. Pylyshyn (1988). "Connectionism and Cognitive Architecture: A Critical Analysis." Cognition **28**: 3-71.

Frege, G. (1892). "Über Sinn und Bedeutung." Zeitschrift für Philosophie und philosophische Kritik **100**: 25-50.

Frege, G. ([1892] 1997). On Sinn and Bedeutung. The Frege Reader. M. Beany. Oxford, Blackwell Publishers, Ltd.: 151-171.

Frege, G. ([1914] 1979). Logic in Mathematics. Posthumous Writings. Hermes. Chicago, Chicago University Press.

Frege, G. ([1919] 1979). Notes to Ludwig Darmstaedter. Posthumous Writings. Hermes. Chicago, Chicago University Press: 253-257.

Frege, G. ([1923] 1977). Compound Thoughts. Logical Investigations. Oxford, Basil Blackwell.

Gillon, B. S. (2004). Ambiguity, Indeterminacy, Deixis, and Vagueness. Semantics. S. Davis and B. S. Gillon. Oxford, Oxford University Press: 157-187.

Grandy, R. E. (1990). "Understanding and the Principle of Compositionality." Philosophical Perspectives **4**: 557-572.

Greenberg, M. and G. Harman (2006). Conceptual Role Semantics. Conceptual Role Semantics. E. Lepore and B. Smith. Oxford, Oxford University Press: 295-322.

Grice, H. P. ([1957] 1989). Meaning. Studies in the Way of Words. H. P. Grice. London, Harvard University Press: 377-388.

Grice, H. P. ([1968] 1989). Utterer's Meaning, Sentence Meaning, and Word Meaning. Studies in the Ways of Words. H. P. Grice. London, Harvard University Press.

Grice, H. P. ([1982] 1989). Meaning Revisited. Studies in the Way of Words. H. P. Grice. London, Harvard University Press.

Grätzer, G. (1979). Universal Algebra. New York, Springer-Verlag.

Hadley, R. (1997a). "Cognition, Systematicity, and Nomic Necessity." Mind & Language **16**: 137-153.

Hadley, R. (1997b). "Explaining Systematicity: A reply to Kenneth Aizawa." Minds and Machines **7**: 571-579.

Haegeman, L. (1995). Introduction of Government and Binding Theory. Oxford, Blackwell Publishers Ltd.

Hampton, J. A. (1987). "Inheritance of Attributes in Natural Concept Conjunctions." Memory and Cognition **15**(1): 55-71.

Hampton, J. A. (1988a). "Disjunction of Natural Concepts." Memory and Cognition **16**(6): 579-591.

Hampton, J. A. (1988b). "Overextension of Conjunctive Concepts: Evidence for a Unitary Model of Concept Typicality and Class Inclusion." Journal of Experimental Psychology: Learning, Memory, and Cognition **14**(1): 12-32.

Hampton, J. A.(1991). The Combination of Prototype Concepts. The Psychology of Word Meanings. P. J. Schwanenflugel. Hillsdale, Lawrence Erlbaum Associates**:** 91-116.

Hampton, J. A. (1993). Prototype Models of Concept Representation. Categories and concepts: Theoretical views and inductive data analysis. I. van Mechelen, J. Hampton, R. S. Michalski and P. Theuns. London, Academic Press.**:** 67-95.

Hampton, J. A. (1997). "Conceptual combination:Conjunction and negation of natural concepts." Memory and Cognition **25**(6): 888-909.

Hampton, J. A. (2006a). Concepts as Prototypes. The Psychology of Learning and Motivation: Advances in Research and Theory. B. H. Ross. **46:** 79-113.

Hampton, J. A. (2006b). "Typicality, Graded Membership and Vagueness." Cognitive Science, **In Press**.

Hampton, J. A. and M. L. Jönsson (Forthcoming). Typicality and Compositionality: The Logic of Combining Vague Concepts. The Oxford Handbook of Compositionality. M. Werning, W. Hinzen and E. Machery, Oxford University Press.

Heim, I. and A. Kratzer (1998). Semantics in Generative Grammar. Oxford, Blackwell Publishers.

Hendriks, H. (1993). Studied Flexibility. Categories and Types in Syntax and Semantics, ICCL Dissertation Series, University of Amsterdam.

Hendriks, H. (2001). "Compositionality and Model-Theoretic Interpretations." Journal of Logic, Language, and Information **10**: 29-48.

Higginbotham, J. (1986). Linguistic Theory and Davidson's program in Semantics. <u>The Philosophy of Donald Davidson: Perspectives on Truth and Interpretation</u>. E. Lepore. Oxford, Basil Blackwell.

Higginbotham, J. (2003). "Conditionals and Compositionality." <u>Philosophical Perspectives</u> **17**: 181-194.

Hintikka, J. (1980). Theories of Truth and Learnable Languages. <u>Philosophy and Grammer</u>. S. Kanger and S. Öhman, D. Reidel Publishing Company: 37 - 57.

Hintikka, J. and G. Sandu (1999). Tarski's Guilty Secret: Compositionality. <u>Alfred Tarski and the Vienna Circle</u>. J. Wolenski and E. Köhler, Kluwer Academic Publishers: 217-230.

Hodges, W. (2001). "Formal Features of Compositionality." <u>Journal of Logic, Language, and Information</u> **10**(1): 7-28.

Hodges, W. (forthcomming). Formalising the Relationship between Meaning and Syntax. <u>Handbook of Compositionality</u> M. Werning, W. Hinzen and E. Machery, Oxford University Press, Oxford

Horwich, P. (1997). "The Composition of Meanings." <u>Philosophical Review</u> **106**(4): 503-532.

Horwich, P. (2001). "Deflating Compositionality." <u>Ratio</u> **14**(4): 369-385.

Horwich, P. (2005). <u>Reflections on meaning</u>. Oxford, Oxford University Press.

Janssen, T. M. V. (1986). Foundations and Applications of Montague Grammar, Part 1. <u>CWI Tract no. 19</u>, Center for Mathematics and Computer Science, Amsterdam.

Janssen, T. M. V. (1997). Compositionality. <u>Handbook of Logic and Language</u> J. van Benthem and A. ter Meulen, Elsevier, Amsterdam: 417-464

Janssen, T. M. V. (2001). "Frege, Contextuality and Compositionality." <u>Journal of Logic, Language, and Information</u> **10**: 115-136.

Johnson, K. (2004). "On the Systematicity of Language and Thought." <u>Journal of Philosophy</u> **101**(3): 111-139.

Johnson, K. (2006). "On the Nature of Reverse Compositionality." <u>Erkenntnis</u> **64**: 37-60.

Jönsson, M. L. (2007). The Unity of Quoted Material, Manuscript -

Jönsson, M. L. and J. A. Hampton (2008). "On Prototypes as Defaults." Cognition **106**(2): 913-923.

Kamp, H. and B. Partee (1995). "Prototype Theory and Compositionality." Cognition **57**: 129 - 191.

Kaplan, D. ([1977] 1989). Demonstratives. Themes from Kaplan. J. Almog, J. Perry and H. K. Wettstein. Oxford, Oxford University Press**:** 481-563.

Katz, J. J. and J. A. Fodor (1963). "The structure of a semantic theory." Language **39**: 170-210.

Katz, J. J. and P. M. Postal (1964). An integrated theory of linguistic descriptions. Cambridge. Massachusetts, The M.I.T. Press.

Kayne, R. (1984). Connectedness and Binary Branching. Dordrecht, Foris Publications.

Kazmi, A. and F. J. Pelletier (1998). "Is Compositionality Formally Vacuous?" Linguistics and Philosophy **21**: 629-633.

King, J. C. and J. Stanley (2005). Semantics, Pragmatics, and the Role of Semantic Content. Semantics versus Pragmatics. Z. G. Szabó. Oxford, Oxford University Press.

Laertius, D. (2006). The Lives and Opinions of Eminent Philosophers, London, Kessinger Publishing.

Lahav, R. (1989). "Against Compositionality: The Case of Adjectives." Philosophical Studies **57**: 261-279.

Lappin, S. and W. Zadrozny(2000). Compositionality, Synonymy, and the Systematic Representation of Meaning, http://www.informatik.uni-trier.de/~ley/db/journals/corr/corr0001.html#cs-CL-0001006, downloaded 2007-04-23

Larson, R. and P. Ludlow (1993). "Interpreted Logical Form." Synthese **95**: 305-355.

Larson, R. and G. Segal (1995). Knowledge of Meaning. Cambridge, Massachusetts, M.I.T. Press.

Lepore, E. and B. Loewer (1989). "You Can Say That Again." Midwest Studies in Philosophy **14**: 338 - 356.

Leslie, S.-J. (2007). 'If', 'Unless', and Quantification, Manuscript -

Lewis, D. (1986). Causal Explanation. <u>Philosophical Papers: Vol II</u>. D. Lewis. Oxford, Oxford University Press: 214-240.

Lewis, D. ([1970] 2004). General Semantics. <u>Semantics</u>. S. Davis and B. S. Gillon. Oxford, Oxford University Press: 191-221.

Lewis, D. ([1975] 2001). Languages and Language. <u>The Philosophy of Language</u>. A. P. Martinich. New York, Oxford University Press: 562-580.

Lycan, W. G. (2000). <u>The Philosophy of Language</u>. London, Routledge.

Mates, B. (1952). Synonymity. <u>Semantics and the Philosophy of Language</u>. L. Linsky. Urbana, IL., University of Illinois Press.

McCullagh, M. (2003). "Do inferential roles compose?" <u>Dialectica</u> **57**(4): 431-438.

McLaughlin, B. (1993a). "The classicism/connectionism battle to win souls." <u>Philosophical Studies</u> **70**: 45-72.

McLaughlin, B. (1993b). Systematicity, conceptual truth, and evolution. <u>Philosophy and Cognitive Science</u>. C. Hookway and D. Peterson. Cambridge, Cambridge University Press.

Moffett, M. A. (2002). "A Note on the Relationship Between Mates' Puzzle and Frege's Puzzle." <u>Journal of Semanics</u> **19**: 159-166.

Montague, R. ([1970] 1974a). English as a Formal Language. <u>Formal Philosophy</u>. R. H. Thomason. London, Yale University Press: 188-222.

Montague, R. ([1970] 1974b). Universal Grammar. <u>Formal Semantics</u>. R. H. Thomason. London, Yale University Press: 222-246.

Montague, R. ([1973] 1974). The Proper Treatment of Quantification in Ordinary English. <u>Formal Semantics</u>. R. H. Thomason. London, Yale University Press: 247-270.

Murphy, G. L. (1988). "Comprehending complex concepts." <u>Cognitive Science</u> **12**(529-562).

Murphy, G. L. (2002). <u>The Big Book of Concepts</u>. London, Mit Press.

Niklasson, L. and T. van Gelder (1994). "On being systematically connectionist." <u>Mind & Language</u> **9**: 288-302.

Nunberg, G., I. A. Sag and T. Wasow (1994). "Idioms." Language **70**(3): 491-538.

Orwell, G. ([1946] 2001). Orwell and Politics. Orwell and Politics. P. Davison. London, The Penguin Group.

Osherson, D. N. and E. E. Smith (1981). "On the Adequacy of Prototype Theory as a Theory of Concepts." Cognition 9.

Osherson, D. N. and E. E. Smith (1982). "Gradedness and Conceptual Combination." Cognition **12**: 299-318.

Osherson, D. N. and E. E. Smith (1997). "On Typicality and Vagueness." Cognition **64**: 189-206.

Pagin, P. (2003a). "Communication and Strong Compositionality." Journal of Philosophical Logic **32**: 287-322.

Pagin, P. (2003b). "Schiffer on communication." Facta Philosophica **5**: 25-48.

Pagin, P. (2004). Compositionality and Context. Contextualism in Philosophy. G. Peter and G. Preyer. Oxford, Oxford University Press: 303-348.

Pagin, P. (2007). Compositionality, Computability and Complexity, Manuscript

Pagin, P. and F. J. Pelletier (2007). Content, Context and Composition. Content and Context. Essays on Semantics and Pragmatics. G. Peter and G. Preyer. Oxford, Oxford University Press.

Pagin, P. and D. Westerståhl (forthcoming). Compositionality. Semantics. An International Handbook of Natural Language Meaning. K. von Heusinger, C Maienborn, and P. Portner. Berlin, Mouton de Gruyter.

Partee, B. (1997). Montague Grammar. Handbook of Logic & Language. J. van Benthem and A. ter Meulen. Cambridge, Massachusetts, The Mit Press: 5-91.

Pelletier, F. J. (1994a). On an Argument against Semantic Compositionality. Logic and Philosophy of Science in Uppsala. D. Prawitz and D. Westerståhl, Kluwer Academic Publishers: 599-610.

Pelletier, F. J. (1994b). Semantic Compositionality:The Argument from Synonomy. Philosophy and the Cognitive Sciences. R.

Casati, B. Smith and G. White. Vienna, Höldar-Pichler-Tempsky: 311-317.

Pelletier, F. J. (2000). Semantic Compositionality: Free Algebras and the Argument from Ambiguity. <u>Formalizing the Dynamics of Information</u>. M. Feller, S. Kaufmann and M. Pauly. Stanford, CSLI Publications: 207-218.

Pelletier, F. J. (2001). "Did Frege Believes Frege's Principle?" <u>Journal of Logic, Language, and Information</u> **10**: 87-114.

Pelletier, F. J. (2004). The Principle of Semantic Compositionality. <u>Semantics: A Reader</u>. S. Davis and B. S. Gillon. New York, Oxford University Press.

Perry, J. ([1986] 2000). Thought without Representation. <u>The Essential Indexical and Other Essays</u>. J. Perry. Stanford, California, CSLI Publications: 171-188.

Quine, W. V. (1972). Methodological Reflections on Current Linguistic Theory. <u>Semantics in Natural Language</u>. D. Davidson and G. Harman. Dordrecht, Reidel Publishing Company: 442-454.

Quine, W. V. ([1975] 1981). On the Individuation of Attributes. <u>Theories and Things</u>. W. V. Quine. Cambridge, MA, Harvard University Press.

Recanati, F. (2002). "Unarticulated Constituents." <u>Linguistics and Philosophy</u> **25**: 299-345.

Recanati, F. (2004). <u>Literal Meaning</u>. Cambridge, U.K., Cambridge University Press.

Robbins, P. (2005). "The Myth of Reverse Compositionality." <u>Philosophical Studies</u> **125**: 251-275.

Rosch, E. (1975). "Cognitive representations of semantic categories." <u>Journal of Experimental Psychology: General</u> **104**: 192-232.

Rosch, E. and C. B. Mervis (1975). "Family Resemblances: studies in the internal structure of categories." <u>Cognitive Psychology</u> **7**: 573-605.

Russell, B. (1940). <u>An Inquiry into Meaning and Truth</u>. London, George Allen & Unwin, Ltd.

Russell, B. ([1918] 1956). The Philosophy of Logical Atomism. <u>Logic and Knowledge</u>. R. Marsh. London, Routledge: 177-281.

Russell, B. ([1919] 1997). On Propositions: What they are and how they Mean. <u>Logic and Knowledge</u>. R. C. Marsh. London, Routledge**:** 285-320.

Russell, B. ([1919] 2001). Descriptions. <u>The Philosophy of Language</u>. A. P. Martinich. Oxford, Oxford University Press**:** 221-227.

Ryle, G. (1949). <u>The Concept of Mind</u>. New York, Barnes & Noble.

Schiffer, S. R. (1987). <u>The Remnants of Meaning</u>. London, The MIT Press.

Schiffer, S. R. (2003). <u>The Things we Mean</u>. Oxford, Oxford University Press.

Simons, P. (1987). <u>Parts</u>. Oxford, Clarenden Press.

Smith, E. E., D. N. Osherson, L. J. Rips and M. Keane (1988). "Combining Prototypes: A Selective Modification Model." <u>Cognitive Science</u> **12**.

Staal, J. F. (1969). Sanskrit Philosophy of Language. <u>Current Trends in Linguistics</u>.**5**:499-531. S. (ed.). The Hague, Mouton.

Stalnaker, R. (1997). Reference and Necessity. <u>A Companion to the Philosophy of Language</u>. B. Hale and C. Wright. Cambridge, Blackwell.

Stanley, J. (2000). "Context and Logical Form." <u>Linguistics and Philosophy</u> **23**: 391-434.

Stich, S. (1971). "What Every Speaker Knows." <u>The Philosophical Review</u> **80**(4): 476-496.

Storms, G., P. de Boeck, I. van Mechelen and W. Ruts (1996). "The Dominance in Concept Conjunctions: Generality and Interaction Aspects." <u>Journal of Experimental Psychology: Learning, Memory, and Cognition</u> **22**(5): 1266-1280.

Storms, G., W. Ruts and A. Vandenbroucke (1998). "Dominance, Overextensions, and the Conjunction Effect in Different Syntactic Phrasings of Concept Conjunctions." <u>European Journal of Cognitive Psychology</u> **10**: 337-372.

Szabó, Z. G. (2000a). "Compositionality as Supervenience." <u>Linguistics and Philosophy</u> **23**: 475 - 505.

Szabó, Z. G. (2000b). <u>Problems of Compositionality</u>. New York, Garland.

Szabó, Z. G. (2001). Adjectives in Context. <u>Perspectives on Semantics, Pragmatics, and Discourse</u>. R. Harnish and I. Kenesei. Amsterdam, John Benjamins Publishing Company: 119-46.

Szabó, Z. G. (2004). "The Compositionality Papers." <u>Mind</u> **113**: 340-344.

Tarski, A. ([1933] 1956). The concept of truth in formalised languages. <u>Semantics, Logic and Metamathematics</u>. A. Tarski. Oxford, Clarendon Press: 152-278.

Taylor, K. A. (1998). Propositional Attitude Statements. <u>Routledge Encyclopedia of Philosophy</u>. E. Craig. London, Routledge.

Tesnière, L. (1953). <u>Esquisse d'une Syntaxe Structurale</u>. Paris, Klincksieck.

Thomason, R. H. (1974). Introduction. <u>Formal Philosophy</u>. R. H. Thomason. London, Yale University Press: 1-69.

Titone, D. and C. Connine (1999). "On the Compositional and Noncompositional Nature of idiomatic Expressions." <u>Journal of Pragmatics</u> **31**: 1655-1674.

van Benthem, J. (1984). The Logic of Semantics. <u>Varieties of Formal Semantics</u>. F. Landman and F. Veltman. Dordrecht, Holland, Foris: 55-80.

van Gelder, T. and L. F. Niklasson (1994). <u>Classicism and cognitive architecture</u>. Sixteenth Annual Conference of the Cognitive Science Society, Atlanta.

Wedberg, A. (1958). <u>Filosofins Historia; Antiken och medeltiden</u>. Stockholm, Thales.

Weiskopf, D. A. (2007). "Compound Nominals, Context, and Compositionality." <u>Synthese</u> **156**: 161-204.

Werning, M. (2005). Right and Wrong Reasons for Compositionality. <u>The Compositionality of Meaning and Content. Volume 1: Foundational Issues</u>. M. Werning, E. Machery and G. Schurz. Frankfurt, Ontos Verlag: 285-309.

Westerståhl, D. (1998). "On Mathematical Proofs of the Vacuity of Compositionality." <u>Linguistics and Philosophy</u> **21**: 635-643.

Westerståhl, D. (2002). On the compositionality of idioms; an abstract approach. <u>Words, Proofs, and Diagrams</u>. D. Barker-Plummer, D. Beaver, J. van Benthem and P. Scotto di Luzio. Stanford, CSLI Publications: 241–271.

Westerståhl, D. (2004). "On the Compositional Extension Problem." Journal of Philosophical Logic **33**(6).

Westerståhl, D. (2007). Compositionality and Ambiguity. http://hum.gu.se/institutioner/filosofi/publikationer/web/index_html/CompAmb2007.pdf downloaded 2008-03-05

Wisniewski, E. J. (1997). "When concepts combine." Psyconomic Bulletin & Review **4**: 167-183.

Wisniewski, E. J. and B. C. Love (1998). "Relations versus properties in conceptual combination." Journal of Memory and Language **38**: 177-202.

von Fintel, K. and S. Iatridou (2002). If and When If -Clauses Can Restrict Quantifiers, Manuscript -

Wright, C. (1986). "How Can the Theory of Meaning be a Philosophical Project?" Mind and Language **1**(1): 31-44.

Zadeh, L. A. (1965). "Fuzzy Sets." Information and Control **8**(3): 338-353.

Zadeh, L. A. (1975). "Fuzzy Logic and Approximate Reasoning." Synthese **30**: 407-428.

Zadeh, L. A. (1982). "A Note on Prototype Theory and Fuzzy Sets." Cognition **12**(3): 291-297.

Zadrozny, W. (1994). "From Compositional to Systematic Semantics." Linguistics and Philosophy **17**: 329-342.

# Index