LOGIC AND THE COMPREHENSION OF LANGUAGE

by

Gregory Coppola

Bachelor of Mathematics, University of Waterloo, 2007

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS
in the Department
of
Linguistics

SIMON FRASER UNIVERSITY
Spring 2009

APPROVAL

Name: Gregory Coppola

Degree: Master of Arts

Title of Thesis: Logic and the Comprehension of Language

Examining Committee:

Dr. María Teresa Taboada
Chair

Dr. Francis Jeffry Pelletier, Senior Supervisor
Professor, Philosophy
Professor, Linguistics
Simon Fraser University

Dr. Nancy Hedberg, Supervisor,
Associate Professor, Linguistics
Simon Fraser University

Dr. Fred Popowich, External Examiner,
Professor, Computer Science
Simon Fraser University

Date Approved: March 27, 2009

Abstract

This thesis examines what is necessary to formally model a hearer's comprehension of a natural language sentence. Our theory of comprehension should at least explain how different words within the same grammatical class make different contributions to the meaning of a sentence. And, our theory should explain how the "full propositional form" that a speaker communicates is recovered from the relatively semantically underspecified acoustic signal.

A model is provided which achieves this. A speaker is said to understand an utterance by, first, choosing the maximally "relevant" full propositional semantic enrichment of the underspecified acoustic signal, measured according to a formally defined comparison operator, and, then, computing the inferences that follow from that chosen propositional form in conjunction with their individual word-/world-knowledge.

This model of comprehension apparently makes comprehension relative to an individual's idiosyncratic knowledge. So, I also discuss how conventionalized word-meanings co-ordinate individuals' knowledges to allow successful interpersonal communication.

To the electric car

Acknowledgments

I want to thank my parents and grand-parents, who worked determinedly so that future generations in our family would have the opportunity to do things like go to University, and for making sure I always did my homework (and with a full stomach, at that). I also want to thank my aunt who counseled me, when I had thought about quitting computer science to study psychology, that I should try to find a way to do both because the former might somehow turn out to be useful in studying the latter. Indeed. And, I want to thank all of my family for being such a good family in general.

Academically, Dr. Jeff Pelletier, my senior supervisor, a rad individual by general agreement who taught me the lay of the semantics landscape, was a *sine qua non* of this thesis because he let me write about absolutely whatever I felt like, and also because he was in the unusual position of being able to give very useful comments on chapters about each of semantics, philosophy and computer science. Thanks also to Dr. Nancy Hedberg, who seems to sit on everyone's thesis committee, for finding time to also sit on mine mine, and also for her enjoyable class on pragmatics, which was very influential on this thesis and on my development as a linguist. Thanks to Dr. Maite Taboada for letting me sit in on her class on discourse, where I got a lot of the crucial ideas that you will read about. Also, thanks to the hard-working Dr. Fred Popowich who found time to read my thesis among the million or so other things he has got to do.

I first had an inkling of the general ideas that would become my Master's thesis as an undergraduate student at the University of Waterloo. There, Dr. Mike Ross, Dr. Jonathan Buss, and Dr. Randall Harris, each taught me things about psychology, computer science, and linguistics, respectively, that facilitated the birth of this inkling.

Lastly I would like to thank all of my friends in B. C., and especially in East Vancouver, for good times and shakin' parties. Also, thanks to my friends back in Ontario, especially because sometimes I was down and visiting home made me better.

Contents

Chapter 1

Introduction

1.1 Overview of this Thesis

This is an inquiry into the role of logic in the comprehension of a linguistic utterance, conducted from a scientific perspective[1]. Our concern is with what sort of processes must be said to occur, and what sort of representations must be said to be created, in order that we can say that a linguistic utterance has been "comprehended" by a hearer.

We will begin by trying to figure out what the overall structure of a theory of comprehension should look like. To this end, we will first consider, in §2, the truth-theoretic, compositional semantic program of Montague (1970a, 1970c, 1974) and Davidson (1967). I will conclude that what truth-theoretic, compositional semantics amounts to, from a scientific perspective, is a translation algorithm from a natural language to a logical language (i.e. the "meta-language"). In arguing this, one thing that I will have to do, in §2.2.1, is rebut the suggestions of some philosophers and semanticists that model-theoretic semantics is *more* than translation because it is somehow able to explain a link between language and the "world of non-symbols"[2] (Lewis 1970; Dowty, Wall and Peters 1981).

I then argue that compositional semantics, i.e. translation of natural language into logical language, does not constitute, in and of itself, a theory of comprehension because, if nothing else, it gives us only an account of how the "meanings" of complex expressions are built out of the meanings of their parts. As Thomason (1974) readily admits, it does not give an account of the meanings

[1] On the notion of a "scientific perspective," cf. §1.2.1.

[2] This is an allusion to Lewis's use of this term, cf. Lewis 1970, p. 170.

of the atomic parts, i.e. the words. That is, it cannot distinguish between the contribution to meaning of words in a single grammatical class, such as, e.g., between "walk" and "run." Since it seems rather obvious that a full model of comprehension will have to be sensitive to such differences, I conclude that, at the least, we will need to supplement compositional semantics with a theory that explains the contributions to meaning of lexical items within a grammatical class.

In §3, we will look at some further complications for a theory of comprehension. We will see that the semantic material available directly, and in a context-invariant way, from the acoustic signal is underspecified with respect to the *full propositional form*, a logical language sentence that the speaker is hypothesized (by our theory) to communicate via that acoustic signal, communicated.

We will adopt a Sperber and Wilson (1986, 1995)-style distinction between the *explicature*, which is the full propositional form that the speaker communicates with an utterance, and the *implicature*, the set of other logical sentences that the explicature communicates, in addition to itself. This distinction, in turn, can be seen as a formalization of the Gricean (1975) distinction between "what is said" and "what is meant" by a communicative utterance[3].

We will see that guessing the speaker's intended explicature is a context-sensitive process that seems to require awareness of word-/world-knowledge as well as general reasoning processes. In addition, the speaker must guess the explicature at the same time as, and as part of the same process in which, they resolve the speaker's implicature (Sperber and Wilson 1998). The nature of this pragmatic process will have significant repercussions for the overall structure of our theory of comprehension (Carston 2000).

The conclusion of the first two sections will be as follows. We cannot appeal to the "world of non-symbols" to help us explain comprehension scientifically. Our model must involve only the manipulation of representations. We begin by translating a natural language sentence to an intermediate, semantically underspecified representation. Then, we must explain how pragmatic processes enrich this representation to arrive at the explicature-implicature pair that the speaker is thought to have meant to convey. Then, we must explain how the explicature is "understood" in a way that, at least, is sensitive to the differences in meaning between individual words from the same grammatical class.

In §4.1, I will sketch a model that shows how to accomplish the goals just outlined. An explicature will be said to be understood through the computation of the inferences that follow from it, in conjunction with the set of sentences that represents the durable word-/world-knowledge of the

[3]Given that Grice's notions of "what is said" and "what is meant" are so vague, some purely terminological questions arise in trying to identify them with more formal concepts. On this point, cf., the discussion of Bach on page 38.

hearer. I will explain why the "comprehension" of a propositional form should be equated with the computation of inference following from it.

In §4.2, I will give a formal theory that generates, for one example problem, an explicature-implicature pair on the basis of an underspecified acoustic signal. That is, I will model how to retrieve the fully specified form from an underspecified intermediate representation as a function of the hearer's word-/world-knowledge and the context. The explicature will be resolved by the same pragmatic processes that resolve the speaker's implicature. *This is the first formal model that demonstrates how explicature and an arbitrary implicature can be derived in parallel*[4].

In equating the "comprehension" of an explicature with the computation of inference that follows from it, I will be appealing to what is known in philosophical circles as a *holistic* theory of word-meaning (cf. Block 1986, Fodor and Lepore 1992). A "holistic" theory, in general, is so-called because the "meaning" of each word in a language depends on the inferential relationship between that word and each other word in the language. This is relevant because, in the model of comprehension laid out in §4, a speaker's understanding of a word is relative to the inferential relationships between that word and others in their own (idiosyncratic) word-/world-knowledge. So, formidable questions are raised as to how communication between multiple people can succeed if the "meaning" of each word is so strongly relative to who is interpreting it.

Thus, in §5, I explain how it is that the word-/world-knowledges of various linguistic community members are co-ordinated via the conventionalization of word-meanings, so that interpersonal communication can succeed. This will involve a close examination of Quine's (1951) argument against an "analytic"-"synthetic" distinction, with the conclusion being that neither Quine's arguments nor his apparent intention in his famous paper really preclude the possibility of assuming that conventionalized word-meanings serve a role in interpersonal communication.

Questions about holism and analycity might traditionally be seen as belonging more to the philosophy of language than to the more empirically-minded branch of linguistics. However, this discussion is highly relevant to our discussion here. First of all, the nature of the solution given in §4 requires one to address the problems noted in the literature reviewed in §5, since these concerns are well known, in certain circles. If I were to ignore these concerns, some readers might dismiss my general account on the grounds that I was unaware of, or unable to address, its attendant problems, both of which criticisms would be incorrect.

[4]The only other case of a formal theory that could be said to deal with implicature resolution, that I am aware of, is the resolution of presuppositions in Asher and Lascarides 2003. However, resolution of presupposition is a highly restricted sub-problem compared to the resolution of implicature, in general.

Secondly, the empirically-minded linguist will appreciate the fact that my attitude in §5 is realistic, in the sense that it has as its goal the modeling of simple speech communities approximating real ones, rather than the creation of an abstract theory of meaning, not necessarily intended to explain anything.

In the course of this thesis, we will look at and use the arguments and ideas of: Montague, Asher and Lascarides, Grice, Carston, Sperber and Wilson, Chomsky, Quine, Fodor, Davidson, Tarksi, Lewis, Dowty, Wall and Peters, Karttunen, Heim, Kratzer, Gundel, Hedberg and Zacharski, Partee, Bach, Levinson, Pelletier, Lepore and Davis and Gillon, among others.

1.2 Some Preliminary Remarks

Before moving on, I would like to make a few brief remarks about the nature of this inquiry. First of all, this is intended as a *scientific* inquiry, as explained in §1.2.1. Second, this is primarily a look at what a *hearer*, as opposed to a speaker, does, as discussed in §1.2.2. Lastly, this is an inquiry into the *process* of "comprehension," rather than any *object* called "meaning," as discussed in §1.2.3.

1.2.1 This is a Scientific Inquiry

In certain branches of linguistics, there is, it seems, a distinction to be made between researchers who feel that they are doing mathematics (with the most famous example being, perhaps, Montague), versus those who feel that they are doing psychology (with the most famous example being, perhaps, Chomsky). I actually feel that the distinction to be drawn is between those doing *mathematics* and those doing *science*.

I will define a *mathematical inquiry* as one in which the only valid objections (other than boredom) that might be made against some inquiry are that: i) there have been conclusions drawn which did not follow properly from the premises and the rules of inference, or that ii) a contradiction has been derived. That is, suppose that one assumes that, "All men are immortal," and, "Socrates is a man." Then, it would be mathematically correct to say that, "Socrates is immortal," since this follows from the premises (as well as some standard rules of inference). It would be mathematically *incorrect* to conclude that, "Socrates is mortal," as this does not follow from the premises.

In a *scientific inquiry*, there are additional objections that can be made against a theory. In science, it is fair to object, for example, that, "That is not how things are," or, "Our observations contradict your predictions." So, even if, "Socrates is immortal," follows properly from one's premises,

there is the additional possibility that one's inquiry can be discredited if there is some observational test for immortality that Socrates fails.

So, this is a scientific inquiry. My goal is to model a part of the "world out there," and the ultimate test for whether this inquiry accomplishes this task will be whether it makes, or else at least leads eventually, to predictions about "observables."

So, all discussion and critique of the work of others will take place in this context, regardless of what the bias of the author of that work was. Consider, for example, that much of Montague's work (e.g., Montague 1970c) is unimpeachable and very interesting as an abstract study of formal language and logic.

But, when I discuss his work, I will be evaluating it primarily for its potential to fit in to an overall *scientific* theory of comprehension. I will do this whether he intended his work to be evaluated this way or not, because, regardless of his goals, if there is a suspicion that his work will be useful towards our ends, we should consider it. If the answer is negative, that answer should be recorded along with the attendant reasoning.

1.2.2 This is an Inquiry Focused on the Hearer

Much work in (scientific) linguistics takes place in abstraction from speakers and hearers. That is, many theories neither makes reference to the process of producing an utterance, nor of decoding it, but are meant to apply to both.

Syntax is an excellent exemplar. If one idealizes, say, the English language as a(n infinite) set of sentences, then one can give a transformational grammar (Chomsky 1957) that generates this language, and this is assumed to characterize knowledge that is used by both speaker and hearer. I think that this is a productive way, in certain contexts, to approach syntax, and it is adopted, not only by Chomsky, but even by legitimate modern syntactic theorists, such as Pollard and Sag (1994).

In some cases, however, one cannot abstract away from the processes of speech production and comprehension. One has to either choose to model the hearer or to model the speaker. Comprehension is, perhaps obviously, one of those cases. What we will be looking at are things that only the hearer does (though the speaker may have to take account of his hearer). For example, if Annie says to Bob, "I can fish," only Bob, but not Annie, needs to resolve the syntactic and lexical ambiguity here (there are two readings). In general, there will be much such work that the hearer must undertake to recover the thought from an ambiguous and underspecified signal, which the speaker, as the one who knows the thought beforehand, will not need to undertake.

1.2.3 "Comprehension" versus "Meaning"

Two words which are going to occur repeatedly throughout this thesis are "comprehension" and "meaning." These two terms deserve special attention, right from the outset, for two reasons. First, they are both *pretheoretical terms*. I define a pretheoretical term as a term which does not have a definition which we are prepared to fix once and for all. Such terms are used by thinkers to name objects and phenomena at a time when those object and phenomena are not well understood, in the sense that our models of them are not sufficiently advanced to make predictions that had at an early stage been impossible.

An example of a pretheoretical term would be the grammatical notion of "subject." Most grade-school children will be able to identify, in any given sentence, the expression that should be considered the subject. For example, a Spanish speaking child would probably identify the "subject" as being *los tacos* in the following example:

(1) Llegaron los tacos.
arrived-3RD-PL DET-PL tacos
'The tacos have arrived.'

But, while many children can do this, it is quite unlikely that they will be able to give an explicit rule that explains, without recourse to the tacit knowledge of a competent speaker, how to find the "subject" in an arbitrary sentence. In support of this claim, consider that it is widely agreed among linguists that Chomsky's (1957) advent of *generative grammar* constituted the first time that such a goal could be discussed. Further, consider that, even so many years later, there continue to exist among professional linguists serious disagreements as to the nature of subjecthood[5].

Pretheoretical terms can be extremely useful precisely because we can use them to communicate about a notion long before that notion is well-understood enough to allow the creation of explicit and immutable definitions. Also, we can continue attaching new properties to pretheoretical terms.

Consider the notion of "comprehension." Well, it would seem that one of the aspects of "comprehending" a sentence is somehow noticing that (2) relates $snow$y things (or ideas about them) to $white$ things:

(2) Snow is white.

We can all recognize this as some part of "comprehension."

[5]For example, Chomsky (1965) has suggested, and still maintains, that subjecthood should be considered an epiphenomenon, which our syntactic theory could be stated without reference to. Pollard and Sag (1994) have suggested that subjecthood is a necessary, uneliminable primary notion, which one cannot do syntax without.

But, are we ready to say that this is all that is involved in "comprehension"? Surely not. Another thing that we would probably want a theory of comprehension to do is to explain that (3) is actually making a stronger statement about John that it appears on the "surface" (whatever the "surface" might turn out to be):

(3) John isn't looking his best today, is he now, William?

That is, (3) is saying that John looks rather bad[6].

So, if we had fixed the notion of comprehension, after viewing the example in (2), we would have found it to be insufficient as a model for everything that we feel, on the basis of our notions of the pretheoretic term, that a theory of "comprehension" should explain. It therefore seems to be a bad strategy to use "comprehension" as a fixed concept, and better to use it as an open-ended concept for which we are always ready to add new properties.

But, we we should, in general, exercise caution when using pretheoretical terms. One cannot use them to draw strict conclusions the way that one can use rigorously defined terms for this purpose. Such usage will yield wild conclusions on the basis of plausible-sounding axioms. The reason for this, I think, is that, because reasoning about them is purely intuitive, it is easy to smuggle in faulty assumptions without noticing.

Jerry Fodor is a good example of someone who, I allege, uses pretheoretical terms with entire seriousness, as though they were formal, in a way that frequently leads him to questionable conclusions. His general attitude can be well illustrated by the following Fodorism on the topic of the principle of semantic compositionality: "So not-negotiable is compositionality that I'm not even going to tell you what it is," (Fodor 2001, p. 6). In other words, his expectation is that we can have a profitable discussion without even needing to know what we are talking about.

Clearly, we cannot, I think, seriously discuss any claim Fodor makes involving his notion of "compositionality" because we have not been told "what is is." First of all, the discussion of the (pretheoretical) notion of "compositionality" in §3.2[7] illustrates that the principle in question can be trivially true, or trivially false, depending on how one fleshes out the definition. Thus, without being firmed up, the concept is not "not-negotiable," but vacuous.

For another example of confusion arising in Fodor's work due to his overly-serious usage of pretheoretic terms, cf. the discussion of Fodor and Lepore 1992 in §5.5. There, I argue that his primary "arguments" against the notion of "semantic holism" arise from his equivocating between,

[6] Cf., the discussion of Grice in §3.1.1.

[7] And, cf., especially, p. 42.

and getting confused about, different definitions for important terms such as, predictably, "compositionality."

Thus, my intention is that the reader should be able to, if they like, find that I have attempted to fix the meaning of "comprehension" or "meaning" in any context for which I mean to create a conclusive argument. To the extent that I have not done this, I have erred.

That said, I want to draw the reader's attention to the fact that this inquiry is named, *Logic and the Comprehension of Language*, rather than, for example, *Logic and Meaning*. That is, I have chosen to focus on the pretheoretical notion of "comprehension," instead of that of "meaning." Why have I done this? Obviously it is not that only the latter, but not the former, is pretheoretical (i.e. because both are pretheoretical).

The crucial difference, for me, is that the word "comprehension" names something which we take to be a *process*, whereas "meaning" names an *object that just sits there*. Allow me to formalize these notions. Let us define a *state machine* to be an object which can be in any one, at a time, of a (possibly infinite) set of *states*. A state machine has a *current state*, and a set of *state transitions*, which are functions from certain states to other *later* states, for all states. Here, *state* and *later* are primitive terms, with the addendum that if the state a is *later* than b, and b is *later* than c, then a is *later* than c. Finite state machines and Turing machines are state machines. I also want to view the universe and the brain as state machines.

Now, the concept of a state machine relates to the making of predictions. That is, a *theory* is a set of sentences in a language such that, for sentences in that language, we can say explicitly whether one sentence *follows as an inference* from a set of others[8]. A theory *makes a prediction* when, on the basis of being given some *information* about the states up to state t, one of the inferences, p, that follows from the theory and this information, is that there exists some state, t', later than t, such that p *describes* the information in the state t'.

A *process* is the set of state transitions for a state machine. Something which is neither a state machine nor a process will be said to be an *object that just sits there*. "Meaning" is always, as far as I am aware of the major literature, considered to be an object that just sits there. Take for example, Putnam's (1975) notion of meaning. The main aspects of Putnam's notion of "meaning" for a word are its: i) extension, and ii) intension[9]. I will assume that an *extension* is clearly neither a state machine nor an algorithm for one. An intension we will take to be a function which takes

[8] I will not define this technically, though cf., e.g., Andrews 1986.

[9] Here, the "extension" of "cat" would be the set of all cats "in the world."

a possible world as an argument, and which yields an extension. The function from possible world to extension could arguably be called a transition, but, since we have just agreed that an extension is not a state, the function does not map us from states to states, and so an intension is not a state machine. Similarly, it is not an algorithm for one.

Since I want to eventually make predictions, we need to deal with state machines. Since "meaning," because it just sits there, does not make predictions *on its own*, it is not a suitable *sole* topic for an inquiry that attempts to make predictions. Thus, we instead make our inquiry one into the nature of "comprehension," which is a process.

Chapter 2

Truth-Conditional Semantics

In this chapter I want to briefly overview the *truth-conditional semantics* paradigm popularized by Montague and Davidson and consider how far along a theory like this gets us towards our goal of a scientific model of the comprehension of a linguistic utterance.

§2.1 will overview the works relevant to the topic by Tarski, Montague and Davidson. §2.1.3, in particular, will delineate the research program that I am going to repeatedly refer to as the "Davidsonian program." What I call the Davidsonian program is a very general, and popular—possibly the most popular—method of implementing a truth-conditional, compositional semantic program to analyze natural language. It is based on the ideas of the early Davidson[1]. Montague's theories would count as "Davidsonian" theories, on the definition that I will supply.

In §2.2, I will argue that, at most, a compositional Davidsonian theory amounts to translation from English into a logical language. In particular, it does not allow the explication of an object-language sentence in terms of the "world of non-symbols." I think that the conclusion that one must draw from the inability of a theory of language to exploit a connection between language and its counterparts in the "world of non-symbols" is that a theory of comprehension should focus on explaining how manipulation of representation—either by translation between kinds of language, or otherwise—can be used to model comprehension.

I will then argue that, while translation from natural to logical language *is* an important aspect of a full model of comprehension, it cannot be more than the first step in such a model, and is not a theory of comprehension in and of itself.

[1]The later Davidson (i.e. Davidson 1986), however, would not be one to endorse what I am calling a "Davidsonian theory."

2.1 Truth-Conditional Semantics

In this section, we consider the works by Tarski, Montague and Davidson that have been influential in the history of the semantic theory of the last half of the twentieth century.

2.1.1 Tarski's Formal Concept of Truth

The story of truth-conditional semantics begins with Tarski's (1935) formulation of a definition for the concept $True$ in a rigorous enough way to form part of the foundation of mathematics. Tarski was trying to formulate the notion of "truth" illustrated by the following monologue:

(4) Bob says that, "North Dakota is beautiful." And, this is true, because, North Dakota *is* beautiful.

In other words, Tarski tries to create a single language L such that, for each sentence in L, α, L also contains a *structural-descriptive name*, for α, say α^{Name}. I will create the name, or the structural description, of an expression by putting it in quotes. So, the name of, Snow, will be, "Snow," and the name of, Snow is white, will be "Snow is white". The reason it is called a "structural-descriptive name" is that the size and structure of the name correlates directly with the size and structure of the expression it names. So, "Snow" names a single word, and has a name which is shorter than "Snow is white", which names a whole sentence that contains the word that "Snow" names.

Using this notation, we can say that, if Tarski's desire had been possible, he would have been able to construct, for the language L, a definition for the predicate $True$, in L, such that, for each sentence p in L, we would have (5):

(5) $True(\text{“}p\text{”}) \leftrightarrow p$

But, for reasons related to the "liar paradox," this does not work. Suppose that "$\neg True(a)$" is the name of the sentence $\neg True(a)$. A problem arises if a = "$\neg True(a)$". In that case:

$$True(\text{“}\neg True(a)\text{”}) \leftrightarrow \neg True(a)$$

And, the above is equal to:

$$True(\text{“}\neg True(a)\text{”}) \leftrightarrow \neg True(\text{“}\neg True(a)\text{”})$$

So, if one derives $True(\text{“}\neg True(a)\text{”})$, then they also derive its negation, which would constitute a contradiction. Thus, $True(\text{“}\neg True(a)\text{”})$ can neither be true nor untrue.

For the applications he had in mind, Tarski wanted each statement to either be true or not true (but not neither), so he needed to solve this problem. His well-known solution was to realize that $True$ can be defined in a way that produces no paradox if one is speaking *in* another language than one is speaking *about*. Adopting terms which were already around at the time, he called the language that we speak *in* the *meta-language* and he called the language that we speak *about* the *object-language*.

Then, Tarski was able to formulate the following condition on an "adequate" meta-language definition of $True$ for each sentence in an object-language:

(6) **Tarski's Convention T**:
A definition of the symbol $True$, formulated in the meta-language, will be called adequate if it yields all sentences obtainable from the template "x is $True$ if and only if p", where for x is substituted the structural descriptive name of an object-language sentence, and for p is substituted its translation into the meta-language.

The reason that we can guarantee that we will not run into a paradox, as far as I understand, is that there is no word in the object language that translates to $True$. So, if we could imagine a pesky object-language speaker, who is trying to catch us, the meta-language speakers, in a liar paradox, he will not be able to, because nothing he says will translate to a sentence that uses our word $True$, and only sentences that use our word $True$ can lead to a liar paradox.

I think it would be worthwhile to go through a small Tarskian truth-definition here, to illustrate some of its properties. Instead of giving a truth-definition for the predicate calculus, as Tarski does, though, I will give a truth-predicate for the propositional calculus, which is much simpler. What I want to illustrate is not the technical tricks that Tarski was forced and able to employ, but rather the *nature* of this kind of "truth," which, we will see, differs considerably from the notion of "truth" that links language to the "world of non-symbols."

So, we are going to define a language of a propositional calculus, PC, and then define a truth-predicate for it. Suppose we have a set, **A**, of variable symbols. These are called the *propositional variables*. Then:

(7) Syntax of PC:
The set of well-formed formulae of PC are defined as follows:
   a. "a" is a well-formed formula of PC for each $a \in \mathbf{A}$
   b. If "α" and "β" are well-formed formulae of PC, then "$(\alpha \wedge \beta)$" is a well-formed formula of PC

c. If "α" is a well-formed formula of PC, then "$\neg\alpha$" is a well-formed formula of PC

d. Nothing else is a well-formed formula of PC

So, if the set of variables **A** includes the variables A and B, then some well-formed formulae of PC would be "A", "B", "$(A \wedge B)$", "$\neg B$", and "$\neg(A \wedge B)$".

Now, I will define a Tarskian truth-predicate for the language PC. The definition of a truth-predicate for a language is often otherwise referred to as a "semantics" for that language. Suppose that we have a function F such that, for each $a \in \mathbf{A}$, either $F(a) = 1$ or else $F(a) = 0$. Then, we can define $True_F$, or in other words, "truth" relative to the function F, as follows:

(8) Semantics of PC:

a. If α is a string consisting of the single variable a, then "α" is $True_F$ if and only if $F(a) = 1$

b. If "α" and "β" are well-formed formulae of PC, then "$(\alpha \wedge \beta)$" is $True_F$ if and only if "α" is $True_F$ and "β" is $True_F$

c. If "α" is a well-formed formula of PC, then "$\neg\alpha$" is $True_F$ if and only if "α" is not $True_F$

(Note that I have switched to a notation that says things like "x is $True_F$". This is as opposed to the use of the "$True_F(x)$" notation that I used in discussing the liar paradox. I think it allows a cleaner exposition.)

Let us, now, illustrate the truth-definition with some examples:

(9) "A" is $True_F$ if and only if $F(A) = 1$

(10) "$(A \wedge B)$"[2] is $True_F$ if and only if "A" is $True_F$ and "B" is $True_F$ (which in turn holds if and only if $F(A) = 1$ and $F(B) = 1$)

(11) "$\neg B$"[3] is $True_F$ if and only if "B" is not $True_F$

First of all, the material that occurs to the right hand side of "if and only if" in the truth-definition of each formula is referred to as the *truth-conditions*, or the *model-theoretic truth-conditions*, of that formula. I will continue to use this terminology throughout the thesis.

Now, the first thing to note about the Tarskian truth-definition for PC that I have just given is that, while many semanticists and philosophers of language will appeal to this concept of "truth" as,

[2]Note that $\wedge$ is the logical language symbol for the word "and". So, $A \wedge B$ is read "A and B".

[3]Note that $\neg$ is the logical language symbol for "it is not the case that". So, $\neg B$ is read as "it is not the case that B".

ostensibly, a way of explicating a connection between language and the "real world" that language describes, there is actually no mention of the "real world," "reality," or the "world of non-symbols" in (8). Whether or not "A" is true depends only on the value of $F(A)$. That is, it only depends on the function F itself.

So, whatever may come of attempts to use Tarski's concept for the study of "truth" in the philosophical sense of the word, we should note that discussion of "reality" is not in any way intrinsic to Tarski's concept. Also, talk of reality is never present in the mathematical applications of this concept.

Also, and in connection with the truth-definition for PC specifically, I would like to address the question as to why there is such a large fuss made about such an apparently vapid concept. That is, what good does it do to know that "$A \wedge B$" is $True$ if and only if "A" is $True$ and "B" is $True$? Does it not seem that we are simply shuffling words around?

Well, one place where this concept is rigorously employed to great benefit is the field of mathematics. Suppose we have a rule of inference, R, which says that if one assumes "$(A \wedge B)$", then one is justified to conclude "A". That is, if we are assuming that, "Tom is tall and Tom likes sports," then we are justified in inferring that "Tom is tall". Now, intuitively, we want to know that the rules of inference that we employ are not leading us from "true" assumptions to "false" conclusions. But, we need some way to formalize this notion of "truth," in a way that contains no mystery, metaphysical or otherwise. For this, we use, for a given language like PC, a truth-concept such as that of $True_F$ given in (8).

(8) tells us that "$(A \wedge B)$" is $True_F$ if and only if "A" is $True_F$ and "B" is $True_F$. So, if "$(A \wedge B)$" is $True_F$, then "A" is $True_F$. Thus, our inference rule which that, from "$A \wedge B$", we can infer "A", can only lead from truth to truth (i.e. from $True_F$'th to $True_F$'th). Thus, we can be sure that our inference rule R, because it leads us from truth only to truth, is trustworthy. This is the sort of application that Tarski's concept is given among mathematicians which is, from the point of view of metaphysics, rather mundane.

To recapitulate, we have looked at what Tarski's truth-concept is, how it avoids the liar paradox, and what use it has. I have wanted to go through Tarski's concept of truth in some detail because in §2.2.1 I will consider proposals by various semanticists and philosophers of language that attempt to imbue this concept with a mystical ability to connect language to the world that, I argue, it simply does not have. That discussion would have been less clear if the nature of Tarski's definition had not actually been discussed.

2.1.2 Montague's Semantics for Natural Languages

The story of how Tarski's concept of truth came to form the basis for the dominant paradigm in semantics heavily features the work of Davidson and Montague. Davidson's (1967) important paper on the topic antedates Montague's but I will consider Montague's work first, as it is arguably more faithful to the Tarskian ideas it is based on. Montague essentially showed how one could give Tarskian truth-conditions for natural language sentences.

Tarski (1935) had said that one prerequisite for the definition of a truth-predicate like his was that one should be able to tell, based on "purely structural properties" (1935/1956, p. 166), for the language, L, for which "truth" is being defined, whether or not some string is in L. Now, if the language in question is English, it would not have been possible in 1935 to give a formal theory which could determine whether a string was an English sentence, because, at that time, the only formal grammars were context-free grammars, and English is not a context-free language.

The argument for this, and the idea that an arbitrary computable language could be modeled using an unrestricted rewrite grammar is precisely Chomsky's (1956, 1957) seminal contribution to linguistics. Montague (1974) was able to exploit Chomsky's grammatical tools to create a method that could assign model-theoretic truth-conditions to, in principle, any sentence in English.

Suppose we have an unrestricted rewrite grammar for English. Then, what we can do is we can take some given English sentence, s, and show how s can be translated into a logical language sentence, l. Then, we show how to give meta-language truth conditions for l. Call these m. Then, the truth-conditions for s are m.

For example, the English (12) would be translated via a set of translation rules to the logical language (13). Then, Montague would give a set of (Tarski-style) semantic rules that would apply to (13) to yield the meta-language truth-conditions (14):

(12) Annie gives Bob money

(13) $gives(\textsc{Annie}, \textsc{Bob}, \textsc{Money})$

(14) $\langle [\![\textsc{Annie}]\!], [\![\textsc{Bob}]\!], [\![\textsc{Money}]\!] \rangle \in [\![gives]\!]$

In other words, (12) is assigned truth-conditions as follows:

(15) "Annie gives Bob money" is $True_M$ if and only if
$\langle [\![\textsc{Annie}]\!], [\![\textsc{Bob}]\!], [\![\textsc{Money}]\!] \rangle \in [\![gives]\!]$

In "Universal Grammar," Montague (1970c) proves that this process can work generally as a method to assign truth-conditions to arbitrary languages.

Montague does not wax philosophical in "Universal Grammar," but Dowty, Wall and Peters (1981) explain that Montague expressed, presumably off-the-record, that *the creation of the model-theoretic interpretation was the whole purpose of doing semantics*. The logical language that served as the intermediary was only a means to an end, and was not of any particular interest in itself.

Now, in mathematics, "semantics" *is* the giving of truth-conditions for a language. That is, one will speak of the syntax and semantics, such as that of PC in §2.1.1, where the "semantics" are the truth-conditions. This seems to be the way that Tarski used the term. And, Montague, it would seem, was following suit.

However, in §4, I am going to suggest that, in a scientific theory of comprehension, the level of logical language that Montague viewed as an uninteresting intermediary is actually more important than the model-theoretic truth-conditions. That is, my model of comprehension will include a substantial, and crucial, linguistic "level"[4] of representation consisting of the sort of logical language translation that Montague pioneered. But, I will not have particular use for any Tarskian concept of truth in my model of comprehension.

2.1.3 "Davidsonian Theories"

Montague's application of Tarskian truth-conditions to natural language is quite plausibly the most natural extension of Tarski's idea in this direction. There is really no need for philosophical pause on Montague's part. Once Chomsky made unrestricted rewrite grammars available, it was only a technical question as to how to use this new tool for the purpose of assigning truth-conditions to an arbitrary language. If one's only aim is to give truth-conditions to arbitrary languages, as Montague's was, then there is no need to ask how or whether this kind of project fits in with theories of how people understand language.

Davidson (1967), on the other hand, starts from a perspective of interest in how language is used and understood by humans. He gives a very elegant philosophical argument—which I will reject in §2.2.1—to the effect that a theory that can assign truth-conditions to each sentence in a natural language will give us everything that we had wanted from a theory of natural language *meaning*. Davidson's ideas have remained popular with linguists of the opinion that semantics is a branch of psychology, such as Heim and Kratzer (1998) and Davis and Gillon (2004)[5].

[4] In the sense of Chomsky 1957.

[5] The former endorse Davidson's approach whole-heartedly. The latter list it as an important one out of several ways to think of semantics.

Davidson assumes that a theory of how the "meanings" of wholes are made up of the meanings of parts is required to explain the infinite use of finite means[6] in language. And, Davidson apparently assumes that the purpose of the "meaning" of an expression, in turn, is to pick out the *referent* of that expression.

Davidson argues as follows. Suppose we have a theory in which the meaning of "the father of Annette" is made up of the meaning of "the father of x" as well as the meaning of "Annette." Then, the meaning of "the father of Annette" is used to pick out the referent of the whole thing. Well, what is the use of discussing meaning? Why not just make a theory that picks out the referent of "the father of Annette" directly?

Also, for Frege, the referent of each sentence is a truth-value, i.e. either *true* or *false*. And, Davidson adopts this view. So, says Davidson, we should create a theory that pairs each structural description of a sentence in a language with its "extension"—either true or false.

That is, we need a theory which will yield, for each structural description, s, of a sentence in a language L, a sentence of the form:

(16) s is T if and only if p

Here, "is T" is a predicate which we do not necessarily assume anything about. This bears a striking resemblance, of course, to the Tarskian schema, from (6), repeated here:

(17) s is $True$ if and only if p

Indeed, Davidson notes that predicate T satisfies Tarski's Convention T, which we saw in (6), which specifies the condition for calling something an adequate definition of truth. So, it *is* an adequate definition of truth, in the Tarskian sense, says Davidson. So, says Davidson, what he had sought in a theory of meaning has turned out to be a Tarskian truth-definition for natural language.

I am going to refer to any semantic theory of the form (17) as a "Davidsonian theory" and to the research program that aims to construct Davidsonian theories as the "Davidsonian Program." I use this terminology because, first of all, Davidson was the first, it seems, to propose that semantic theories should be organized in this way.

Also, even though the concept of using model theory to analyze language might be better associated with Montague, Davidson's concept is more general than Montague's. In the Montagovian concept, the right side expression, p, is necessarily a formula in a formal logical meta-language. In a Davidsonian theory, in contrast, p can be a sentence in a natural language and, even, the same

[6]To borrow from Chomsky (1995), who borrows from von Humboldt.

natural language being explicated. It will prove useful to refer to this more general concept, and so I speak of Davidsonian, rather than Montagovian, theories[7].

There have been two, perhaps conflicting, principal ways that the relevance of Davidson's theoretical paradigm has been viewed. Davidson's style seems intentionally cryptic, and both interpretations have remarks that seem to support them.

One popular interpretation of the relevance of this this kind of theory is that, if we can give a Tarskian definition of "truth" for a language like English, what we are somehow showing is that a person understands the "conditions under which each sentence is true." This comfortable-sounding view is especially popular among philosophers of language, whatever it might turn out to mean (cf. the lengthy discussion of §2.2.1).

In support of this view, we find Davidson explaining that his theory "works by giving the necessary and sufficient conditions for the truth of every sentence." And, also, "[t]o know the semantic concept of truth for a language is to know what it is for a sentence—any sentence—to be true, and this amounts, in one good sense we can give to the phrase, to understanding the language," (1967, p. 226). I will delay any elaboration until §2.2.1.

The second interpretation of this sort of theory is more concerned with being able to explain how words systematically contribute to the truth-conditions for the various sentences that they are a part of. In this second interpretation, a Davidsonian theory forces us to explicate how the meaning of English sentences can be given in terms of the meanings of their parts. This view has been popular among many strictly empirically minded linguists, who profess to prefer to largely ignore metaphysical issues, such as Heim, Kratzer and Partee.

What is intended here can be illustrated by looking at Davidson's (1967) own discussion of Bridgitte Bardot. He notes that, depending what one thinks of truth, "Bardot is good," may not have a truth value, because it is a normative statement. But, this does not stop us from building a theory in which:

(18) "Bardot is good" is $True$ if and only if Bardot is good.

Here whatever mystery we may have about how something might be objectively "good," says Davidson, is just transported over into a meta-language mystery. And, of course, it would seem that, if this theory does not solve the mystery, we cannot adopt an overly strong interpretation of what it

[7] I should note, however, that this usage might be slightly misleading in that in Davidson 1986, Davidson takes the position that he no longer thinks that a theory organized along these lines is of any use. This is the same conclusion we will draw in this and the next chapter.

means to know "what it is for a sentence to be true."

So, on this line of thinking, Davidson's theory is not intended to illuminate knowledge about what it means to be Bardot, or what it means to be good. What the paradigm *is* doing is forcing us to explain how, for example, the truth-conditions of "Bardot is a good actress," can be stated in terms of "good" and "actress."

The key point is that we can accept having the word "good" as an atomic element (i.e. one that does not decompose into others) in our compositional theory. And, we can accept having "actress" as an atomic element. But, we cannot accept having "good actress" as an atomic element because if all such compounds are treated as atomic, such as "good friend," "good movie," "bad actress," "tall building," etc., our theory will have infinite size.

So, on this interpretation, the merit of being able to give a finitely stated Davidsonian theory is that it forces us to explain the infinite use of finite means in language. Thus, adherents to this view can remain agnostic, and perhaps even totally indifferent, about the metaphysical implications of such a theory, as Heim and Krazter (1998) and Partee (1996) do.

But, as we will see in §2.2.1 and §2.2.2, a Davidsonian theory can only force us to create a meaningfully compositional theory if the language that the truth-conditions are given in is different than the language being explicated.

2.1.4 Conclusion

We began, in §2.1.1, with Tarski's definition of the concept $True$, which has been of recurring importance in the semantics.

We then saw, in §2.1.2, how Montague extended it to any language for which one could give an unrestricted rewrite grammar, but for which one might have been unable to give a context-free grammar. And, we saw that, though Montague invented a method by which each natural language sentence is translated to its logical language counterpart, he felt that the logical language translation was only an uninteresting and theoretically eliminable step on the way to a model-theoretic interpretation.

Lastly, in §2.1.3, I introduced the notion of a Davidsonian theory, which is a semantic theory in which the goal is to define a Tarski-like truth-predicate for a natural language that fits the template of (17). I said that a Montagovian theory, which assigns formal, model-theoretic truth-conditions to a natural language sentence, was a particular *kind* of Davidsonian theory, but that a Davidsonian theory is a more general concept, for which the truth-conditions in a Davidsonian theory need not

be given in a formal language, as they are in a Montagovian theory, but can instead by given in any kind of language at all.

2.2 The Davidsonian Program is Translation

In this section, I want to argue that, at most, a Davidsonian theory amounts to translation from English into a logical language. And, then, that, at most, this translation is only the first step in a theory of comprehension, rather than a theory of comprehension in and of itself.

Further to this, in §2.2.1, I will examine and rebut some interpretations of a Davidsonian theory that either claim it is theory of comprehension in and of itself, or else that it is above the realm of "mere translation" between different kinds of language. In §2.2.2, I will explain what the merits of the Davidsonian translation project are, but why it still does not suffice as a full theory of comprehension.

Now, in this chapter, we are going to speak as if it makes sense to speak, from a scientific perspective, of a deterministic mapping between the kind of natural language sentence that gets spoken aloud in actual conversation and the corresponding truth-conditions of the "full proposition" that the speaker intends to convey. In §3, we are going to see that, because the information on the acoustic signal is underspecified, this is not actually a valid assumption.

But, I think that this idealization is worthwhile, because it will make it easier to focus on one particular aspect of what will be necessary in a model of comprehension in its full complexity—namely, the ability to account for the difference in significance between different lexical items.

2.2.1 The Davidsonian Program is at most Translation

2.2.1.1 The Vacuity of English-for-English Truth Conditions

One interpretation of the Davidsonian program, especially popular among philosophers of language, is that knowledge of an infinite set of statements of the form (19)—i.e. one for each sentence in the language—constitutes knowledge of the "conditions under which each sentence would be true":

(19) "Snow is white" is $True$ if and only if Snow is white

Wiggins (1997) traces this view back to Frege, but notes that Wittgenstein was the one who focused on it explicitly, cf.:

(20) To understand a sentence in use means to know what is the case if it is true, (Wittgenstein 1921, 4.024).

It is a platitude that is as apparently self-evident as it is vacuous. What does it mean to know what is the case if something is true?

Here is Wiggins' view. Consider the sentence, "The sun is out." Many things are "true," says Wiggins, when, "The sun is out." One is that the sun is out. Another is that things are seen in daylight. Another is that it is daytime. But, understanding a language involves being able to pick out the "intended, privileged, or designated condition," (p. 6), among these. That is, "The sun is out," is true if and only if The sun is out. And, "Snow is white," if and only if Snow is white. And, to know all of this is to know a language.

So, Wiggins evidently feels that a sentence can be its own truth-conditions[8]. I will refer to any theory which uses a sentence of English as the truth-conditions for an English object-language sentence one that gives *English-for-English* truth-conditions. Well, if Wiggins' interpretation is correct, an English-for-English truth-conditional theory constitutes some sort of semantic theory, then the result is that the only thing one actually needs in order to construct a theory of meaning for English is a *syntactic grammar* of English, which I will consider absurd.

That is, if one has a generative grammar of English, which can automatically enumerate all sentences in English, then one can build a Davidsonian theory. That is, for each sentence of English, *s*, we would create the following sentence:

(21) "*s*" is $True$ if and only if *s*

Here, "*s*" is the name of the object-language sentence for which s is the meta-language translation. Here, we would require that no symbol in the object language translate to $True$, otherwise we might be caught in a liar's paradox[9]. Also, for technical reasons, we would require that the quotation symbols in (21) never occur on the right side of the $True$ predicate[10].

But, something must be amiss! That is, we wanted to know about the *meaning* of English sentences, and though we ostensibly have a theory that tells us the conditions under which each

[8] Actually, the sentence does not literally form its own truth-conditions, because, as we said in §2.1.1, we must have a meta-language giving the truth-conditions for an object-language. What I mean here is that the material that appears to the right of the "is $True$ if and only if" has the same syntax as the object-language being explicated.

[9] There are numerous ways to ensure this. For example, we could translate the object-language "true" as true_1, instead of $True$. Recall that I mentioned in §2.1.1 that one of the things that can be used to prevent a liar's paradox from occurring is that no object-language symbol translates to the symbol we are using for our truth-predicate, which in this case is $True$.

[10] Similarly to ff. 9, we can choose some other symbols for any quotes we need on the right side of $True$.

sentence is true, we have been able to make this English-for-English theory while all we started out with was the *grammar* of English. This cannot be right. Thus, it must either be that a language does not amount to knowing its truth-conditions, or else truth-conditions cannot be defined in this way.

But, maybe the reader feels as though I have swindled them. It seems to have been suspiciously easy to give a truth-predicate for English, a complex natural language. I seem to have dealt with the manner in a few lines. Well, one reason that the treatment was so brief was that I omitted *the actual grammar of English*. Obviously, if I had written down the grammar, the theory that generated (21) would have been considerably more complicated.

But, the reader might still say, there is another problem. You have not built up this truth-definition recursively. You have merely built up a single sentence, *s*, first and then stuck two copies of it into a sentence containing $True$. A Tarskian truth-definition, like that given for the language PC in (8), is supposed to be built recursively. Building the truth-definition recursively forces us to explain how meanings of parts are combined to make meanings of wholes.

Actually, if one can build (21) in the manner that we have done, then one can build the same thing recursively. This is easily proven using some of Montague's old tricks. Consider the following mini-grammar of English, where ME abbreviates "meaningful expression," and where α, β, and γ are strings:

(22) a. If "α" is a ME of type N, then "α" is a ME of type NP.

b. If "α" is a ME of type V, then "α" is a ME of type VP.

c. If "α" is a ME of type NP, and "β" is a ME of type VP, then "$\alpha\ \beta$" is a ME of type S

d. "George" is a ME of type N

e. "fishes" is a ME of type V

(Here, $\alpha\ \beta$ is the concatenation of the string α with the string β, with a space in between.)

Then, we define the predicate Ext as follows:

(23) a. If "α" is a ME of type NP and "α" is also a ME of type N, then Ext("α") = Ext("α")[11]

b. If "α" is a ME of type VP and "α" is also a ME of type V, then Ext("α") = Ext("α")

c. If "γ" is a ME of type S such that "γ" = "$\alpha\ \beta$", where "α" is of type NP and "β" is of type VP, then Ext("γ") = Ext("α") Ext("β") [12]

d. Ext("George") = George

e. Ext("fishes") = fishes

(Note that each rule in (23) corresponds to the analogously lettered syntactic rule in (22).)

We can then define $True$ in terms of Ext:

(24) If "α" is of type S, then:
"α" is $True$ if and only if Ext("α")

Now, let us see how to derive the truth-conditions of "George fishes" recursively. "George fishes" is a ME of type S, such that it is equal to "α β", where "α" is a ME of type NP and "β" is a ME of type VP. And, "α" is, in turn, equal to "George", a ME of type of type N. And, "β" is equal to "fishes", a ME of type V.

So, if "γ" = "George fishes", then Ext("George fishes") = Ext("γ") = Ext("α") Ext("β") = Ext("α") Ext("β") = Ext("George") Ext("fishes") = George fishes. So, by (24), "George fishes" is $True$ if and only if Ext("George fishes"). Thus, "George fishes" is $True$ if and only if George fishes. This is as required.

Now, the recursive definition of "truth" supplied in (22) and (23) is only a dressed up version of (21). Because we knew we could give a English-for-English truth-predicate like (21), there was little worry that we could do it in this sort of recursive fashion. Notice that, even though (22) and (23) are more complicated than (21), we still have no knowledge of English besides its grammar. Thus, all one needs to know in order to construct an English-for-English truth predicate is its grammar. And, again, it would seem that knowledge of a language's grammar does not constitute knowledge of the "meaning" of that language.

The reader might now complain that my truth-definition is still rather un-Tarskian because at no point do we explain how to get the value of a sentence by functional application. This is not a valid criticism of my argument because *it is precisely my point!*. English is not a language for which one can get the value of an expression by functional application of its parts.

This is why Frege (1879) is thought to have advanced human knowledge—for he showed us how to create a language for which values of expressions could be gotten by functional application. And, that is why we get nowhere in terms of constructing a semantic theory unless the target language is a logical language. I merely note this point here for the reader making the aforementioned objection. I will not discuss the point in detail now, nor even define here "functional application," because the point will be taken up in §2.2.2.

---

[11] This rule is vacuous because, in this grammar, NP can only be made of a single N. The VP rule is vacuous for the same reason. I have included these rules to better illustrate that this grammar can handle recursivity without trouble.

[12] Note that there is a space between Ext("α") and Ext("β").

To recapitulate, I have argued here that a theory which gives English language truth-conditions to illuminate the English language demonstrates no more knowledge than that of the grammar of English. Assuming knowledge of syntax is not knowledge of "meaning," such a theory contributes nothing to a study of how sentences are understood.

2.2.1.2 Why the "World of Non-Symbols" Cannot Save a Vacuous Theory

Continuing with the discussion of the last section, in the introduction to their textbook on Montague grammar, Dowty, Wall and Peters (1981) explain that the nature of the meta-language, as they conceive it, is such that an English language sentence *can* serve as its own truth-conditions because the meta-language sentence that appears on the right side the "is $True$ if and only if" is more than a mere sentence, it is a stand-in for an extra-linguistic "state-of-affairs," which is a configuration of things "in the world." That is, in explaining that, "Snow is white" if and only if Snow is white, one is explaining "Snow is white" in terms of the state-of-affairs *in the world* where Snow is white.

To elaborate, they say we must learn to "observe that sentences (linguistic entities) and states-of-affairs (configurations of objects in the world) are altogether different things," (p. 5). The false sense that an English-for-English theory is vacuous "comes from the fact that we have used *a sentence of English* to describe the state-of-affairs," (p. 6, emphasis theirs). (Here, they mean that the English meta-language sentence describes the state-of-affairs. The English object-language is just a plain old sentence, despite its superficial similarity to the untrained eye.)

So, the English meta-language sentence represents the state-of-affairs. But, the state-of-affairs cannot be put down on the page—only a representation of it can be put down on the page. So, we use the English meta-language to represent this extra-linguistic state-of-affairs. And, they can use the English meta-language to represent the state-of-affairs because they can "rel[y] on the fact that you, the reader, understand English in order to indicate to you just which state-of-affairs [they] inten[d]," (p. 6).

But, at this point, the circularity of this sort of theory is laid completely bare. Dowty et al. are explicating an English object-language sentence in terms of a state-of-affairs. But, they have admittedly relied on the reader's ability to "understand English" in order to indicate which state-of-affairs is intended. But, it is precisely the ability to "understand English" that we set out to model!, at least if we are operating from a scientific point of view. So, this claim that somehow the extra-symbolic world will turn a trivial, vacuous theory into a serious one is plainly seen to be erroneous.

Now, I should note that, in the body of the book, which is about Montague grammar, Dowty et al. do not actually explicate English sentences in terms of English. They explicate English in terms of Montague's formal logical languages. So, the criticisms I am making do not apply to the technical tools presented in their textbook, but only to the way that they feel that those tools should fit in to a theory of language use.

Now, the discussion of §2.2.1.1 made clear that an English-for-English semantic theory is vacuous. So, it should have been immediately obvious that if Dowty et al. were suggesting that somehow merely holding some point of view about some magical capability of the meta-language could turn an English-for-English semantic theory into a theory of comprehension, they must have been mistaken. But, in fact, because they were so explicit about their position, we were able to go even deeper, and expose the flaw in their logic directly.

Now, what I would like to do is completely discredit the idea that appealing to a link between language the world of non-symbols can save a scientific theory of comprehension that is otherwise insufficient. Now, I have discredited the meta-language interpretation of Dowty et al. to my own satisfaction by looking at the particulars of their argument. But, I would like to now discredit the idea completely without having to review the arguments of each author who has held such a view.

To this end, I would like to consider the views of Lewis, who is perhaps the most famous philosopher to (aggressively) espouse the view that one can supplement the power of an analysis of language by appealing to the link between language and the extra-linguistic. Consider the following passage:

> My proposals regarding the nature of meanings will not conform to the expectations of those linguists who conceive of semantic interpretation as the assignment to sentences and their constituent of compounds of 'semantic markers' or the like... Semantic markers are *symbols*: items in the vocabulary of an artificial language we may call *Semantic Markerese*. Semantic interpretation by means of them amounts merely to a translation algorithm from the object language to the auxiliary language Markerese. But we can know the Markerese translation of an English sentence without knowing the first thing about the meaning of the English sentence: namely, the conditions under which it would be true. Semantics without treatment of truth conditions is not semantics, (Lewis 1970, p. 169, emphasis in original).

(His talk of translation into "Semantic Markerese" is almost self-explanatory. He has in mind, in particular, the theory of Katz and Postal (1964). These researchers felt, as I do, that language

use should be analyzed by positing that an acoustic signal is translated into some code that can be manipulated by the brain. They also felt that there was no need to talk about the "world of non-symbols" when stating a theory which they viewed to be one about how the brain works.)

The first thing to note is Lewis' ardent insistence that merely translating one language to another does *not*, in and of itself, constitute a theory of comprehension. (I would agree. But, I feel that translation is a necessary *part* of such a theory.)

The second is that a truth-conditional theory, like a Davidsonian theory—which is a theory stated *in language*, and which ostensibly relates object-language representations to meta-*language* representations—transcends the linguistic limits of "mere translation." Lewis goes on to suggest that his theory explains "the relations between symbols and the world of non-symbols," (1970, p. 170).

Now, Dowty et al. are notable for at least trying to explain *how* it is that a theory about language written *in* language can transcend language, and their explicitness opened them up to a criticism based on a flaw in their reasoning. Lewis has not explained how it is that his theory does this, otherwise I would rebut his arguments as well. Lewis' attitude seems to be that, not only is a theory which deals with the world of non-symbols superior to one that does not *because he says so*, but he even feels that the very fact that his theory does this should be true because he says so.

As far as I can tell, Lewis thinks that if one were to write, "the translation of 'snow' into semantic markerese," one has put but dull words on a page. But, if one writes, "the extension of 'snow,"' one has transcended the realm of language—and reached right out and grabbed onto some *real* thing in the world of non-symbols. My complaint is that, in both cases, all I see are symbols. And, he has not explained, as far as I can see, why I should think otherwise.

So, contra Lewis, recall that we have already agreed that knowing a grammar of English should be not be a sufficient basis for saying that one understands English. But, what if we were to modify our absurd Davidsonian English-for-English theory above, which we agreed could not constitute an understanding of the meaning of English, so that, instead of printing sentences of the form (21), we will instead print sentences of the form, (25):

(25) "*s*" is $True$ if and only if *t*

Where, *t* is a sentence just like *s*, except that: i) each noun N is instead written as "the extension of N"; ii) each verb V is written as "stands in the extension of the relation V to"; and iii) each adjective A is instead written as, "the extension of A."

Then, instead of (19), we would have (26):

(26) "Snow is white" is $True$ if and only if the extension of Snow stands in the extension of the is relation to the extension of white

Has this talk of extensions now let us get a hold of "the conditions under which this sentence would be true"?

Clearly not, because we have assumed that knowing a grammar of a language is not sufficient to know its meaning. And, since this theory, like the last, was built with only knowledge of the grammar of English—along with some extensional window-dressing that can hardly be called knowledge—it cannot be an adequate account of what it takes to understand the sentence. Thus, merely *believing* in the world of non-symbols is not enough to turn a vacuous theory into a significant one.

But, perhaps I am being a bit unfair to Lewis, who does warn us that his (1970) work concerns only, "possible languages or grammars as abstract semantic systems whereby symbols are associated with aspects of the world," and that this sort of study should not be confused with the "psychology and sociology of language users," (1970, p. 170). The point is this. If Lewis's theory is openly one which deals with an abstract system, with no pretense to be making any empirical claim or modeling the world at all, then that is fine. He is perhaps building an aesthetically pleasing theory of "truth", a picture with words, that evokes in certain kinds of people positive emotions. That is fine.

But, in that case, his claims, such as the one in the block quote above, to be discussing what a person knows when they know the meaning of a sentence are highly misleading, because his is not a theory of what people know (i.e. of "psychology"), but is rather an abstract, unfalsifiable theory of no observable thing in particular. If he wants to discuss what is involved when a person understands language, then he is in the empirical domain. And, in that domain, he ought to give some empirical substance to his claim that his theory manages to transcend language. That is, this claim ought to be involved in making a prediction that could, in principle, be falsified. Otherwise, it is empirically meaningless.

So, in conclusion of our discussion of Lewis, I think we must decide that Lewis' claims to be accessing the "world of non-symbols" are, in the first place, suspicious on quasi-philosophical grounds. That is because his views suggest that by knowing only the grammar of English, as well as how to place the word "extension" in the right places, we can know how to understand English, which we have agreed is absurd. In the second place, Lewis' claims to be accessing the world of non-symbols are empirically meaningless, because he is making no prediction at all.

To recapitulate, the idea that a Davidsonian theory somehow allows us to access the realm of

non-symbols, and transcend the realm of translation, is, in general, highly suspicious and empirically meaningless.

If this is indeed the case, and a connection between language and the world is not to be appealed to in the construction of a scientific theory of comprehension, then it would seem all that *can* be appealed to is representations in the mind, and the processes that manipulate them. So, I think that the natural conclusion is that a theory of comprehension ought to concern itself with the modeling of the manipulation of representations in the mind of the hearer—whether translations between different languages, or manipulation of representations in a single language—in such a way that the result is something that can be said to be a model of "comprehension."

2.2.2 The Merits and Limits of Translation

Now, if the pretensions to be connecting to the world of non-symbols are stripped from the Davidsonian theory, what does a theory of the form of (27) amount to?

(27) x is T if and only if p

What it amounts to is a *translation* from the object-language to the meta-language[13].

Now, the use of English as a meta-language for English demonstrates nothing more than a knowledge of the syntax of English, as we have seen. However, translation of English into French, or English into a logical language is not nearly so trivial.

In the case of using English as a meta-language for English, the only problem which the infinity of language posed for us was that we were required to give a grammar capable of generating the infinite set of English sentences on the basis of a finite description.

But, when translating English to French, one has to give, in addition to this, a finite basis for the translation of an infinite number of English sentences. That is, the problem is kind of doubly as hard, and cannot be achieved with the knowledge of a grammar of English alone, nor even with the knowledge of both the grammar of French and that of English.

[13]Cf. Tarski, who referred to p in a schema like (27) as the "translation of [x] into the meta-language":

> A formally correct definition of the symbol 'Tr,' formalized in the metalanguage, will be be called an *adequate definition of truth* if it has the following consequences[...one group of which is are—g.c.] all sentences which are obtained from the expression '$x \in Tr$ if and only if p' by substituting for the symbol 'x' a structural-descriptive name of any sentence of the language in question and for the symbol 'p' the expression *which forms the translation of this sentence into the metalanguage*, (Tarski 1935, pp. 187—188, emphasis rearranged by myself)

But, while we would call translation of English into French "translation," most of us would call translation of English into a logical language "semantics."[14] The reason is that logical languages are created specifically for the purpose of elucidating natural ones.

In particular, with a logical language, one can give the value of some complex expression by *functional application* of the values of its parts. That is, one can get $[\![\alpha(\beta)]\!]$, called the *value* of $\alpha(\beta)$, by computing $[\![\alpha]\!]\big([\![\beta]\!]\big)$, which is the value of the function α applied to the value of β.

Of course, this is an advance not allowed by translation of English into French because natural languages do not have this property[15]. At least, no well-known semantics program tries to give a value to an English sentence by functional application directly. Instead, most follow Montague in translating English to a logical language for which Tarskian truth-conditions are trivial.

That is, rather than give three different rules that describe how *slowly* would be either a function or an argument in the creation of a value for the three sentences in (28), we would instead translate each of the three to the single Thomason and Stalnaker (1973)-style logical representation in (29), for which one needs only give a single rule:

(28) a. Slowly, Annie kissed Bob.

b. Annie slowly kissed Bob.

c. Annie kissed Bob slowly.

(29) $\big(slowly(kissed)\big)(\textsc{Annie}, \textsc{Bob})$

So, if one is willing to accept, contra Lewis, Dowty, etc., that a Davidsonian theory *is* a translation algorithm, then the benefit of this sort of theory is that it forces the theorists to give, in addition to a grammar of English, a finite set of rules that map English sentences into logical language sentences, for which a definition of truth can be given by functional application. And, in practice this is what most semanticists spend their time doing.

Now, it should be clear that, as I have been categorizing translation, I would consider every aspect of Montague's semantic program to be translation. That is, Montague's method—whereby he would take a natural language sentence, n, and give it a logical language translation, l, for which

[14] Obviously, Lewis is to be excepted here.

[15] Note that to say that the values of natural language sentences cannot be gotten from the functional application of the values of their parts is not necessarily to take a position on compositionality, i.e. on whether or not the "meaning" of an English sentence depends on the meanings of its parts. (Although I will also reject this thesis, in §3.) That is, it is possible that the "meaning" of a natural language is entirely contained within it, but in a way that is more complicated than mere functional application.

he could then give the meta-language truth-conditions, m—is sometimes referred to as "translation semantics."

Someone using this paradigm would refer to m as the "truth-conditions" of n but not a "translation." I would refer to m both as the "truth-conditions" and as a "translation." As discussed in the last section, some semanticists have considered l a mundane representation but m a magical string that grabs on to the world. Since I consider them both to be mundane representations, I will choose to work with l, which is, by the nature of the fact that it is the one that people prefer to translate n into directly, easier to work with.

In addition, with all this emphasis among semanticists on Frege and Tarski, the ability to give a value to a sentence (i.e. either true or false) as a result of functional application is widely appreciated as a, and perhaps the, principal merit of a logical language. But, I would submit, what will turn out to be more important is that translation of English into a logical language is translation into a language for which we have explicitly defined inference rules.

That is, with a given logical language, it is possible to give explicit rules that can say, for some set of premises Γ, and some conclusion ϕ, whether or not $\Gamma \vdash \phi$. That is, whether or not ϕ follows according to rigorous rules of inference from the premises Γ.

To see why this is so important, the question we ask now is, if a Davidsonian theory amounts, and amounts at most, to translation from a natural into a logical language, do we then have a sufficient theory of comprehension? That is, the ability to translate natural language to logical language constitutes *some* knowledge. Does it constitute sufficient knowledge for comprehension?

The answer is no. A compositional theory such as Davidson's, by its design as Davidson 1967 openly discussed, is a theory of how "meanings" of parts come together to create meanings for wholes. The theory does not model the meanings of the atomic (i.e. indivisible) parts[16]. That is, a Davidsonian theory explains what is common between sentences like, "Socrates is a man," and, "Socrates is a Greek," on the one hand, and, "Socrates is a man," and, "Descartes is a man," on the other. That is, in the first pair, we have the common contribution of the subject "Socrates." In the second case, we have the common contribution of the predicate "is a man."

But, the Davidsonian theory does not explain what difference it makes to Socrates whether he is a man, or a Greek, or both. In §4, I am going to propose that significance can be given to the atomic elements by encoding knowledge using inference in a set of statements in a logical language. For example, (30) encodes a mock-up of what it means to be a "man" and to be a "Greek":

[16]Cf. "[T]he task was to give the meaning of all expressions in a certain infinite set on the basis of the meaning of the parts; it was not in the bargain also to give the meanings of the atomic parts," (Davidson 1967, p. 223).

$$(30) \quad \mathbf{K} = \left\{ \begin{array}{c} \forall x\ man(x) \rightarrow \big(mortal(x) \wedge two\text{-}legged(x) \wedge animal(x)\big), \\ \forall x\ animal(x) \rightarrow \neg plant(x), \\ \forall x\ greek(x) \rightarrow \big(citizen\text{-}of\text{-}greece(x) \wedge probably\text{-}tanned(x)\big) \end{array} \right\}$$

So, suppose we have a Davidsonian theory that maps, "Socrates is a man," to a logical language sentence $man(\text{SOCRATES})$. Then, the logical language output of the Davidsonian theory would be combined with **K** in an inferential process, by which the hearer would conclude $mortal(\text{SOCRATES})$, $animal(\text{SOCRATES})$, $\neg plant(\text{SOCRATES})$, etc. This is different than the set of inferences licensed by, "Socrates is a Greek."

This drawing of inference, I claim, is a major aspect of what it we informally consider the "comprehension" of a sentence. That is, the atomic parts of complex expressions—i.e. the individual words or symbols—get their significance via the computation of inference. So, one cannot identify two different phases, one in which the "meaning" of a sentence ϕ is grasped, and another in which the inferences following from ϕ are computed. These two tasks are exactly the same, and are indivisible.

So, the compositional aspect of a semantic theory, however we ultimately construe this, explains how symbols of the same syntactic category (e.g. verb) make the same kind of contribution to a sentence. It helps explain the infinity of language in this way. But, it cannot explain how words of the same syntactic category differ in meaning.

Certain semanticists have attempted to marginalize the importance of accounting for differences in meaning between words within a syntactic category, considering the matter a trivial side-note to real semantic theory. A paradigm example of this attitude is the view espoused by Thomason in his introduction to Montague's posthumous anthology:

> [T]he problems of semantic theory should be distinguished from those of lexicography.... *[W]e should not expect a semantic theory to furnish an account of how any two expressions belonging to the same syntactic category differ in meaning.* 'Walk' and 'run', for instance, ... certainly do differ in meaning, and we require a dictionary of English to tell us how. But the making of a dictionary demands considerable knowledge of the world, (Thomason 1974, p. 48).

Thomason is entitled to circumscribe his domain of inquiry however he likes, I suppose, especially given the ambivalent attitude of the Montagovians as whether they were actually doing anything empirical. But, as for us, our avowed project laid out in §1 is to model comprehension right through its post-syntactic stages and it seems more than obvious that, at some point in the

process of comprehension, the difference between "walk" and "run" will have to be accounted for.

Thomason may feel that giving an exhaustive list of atomic word meanings is not theoretically interesting, and perhaps it is not. But, certainly we need to at least know the *structure* of the system by which word-meanings are encoded. And, we are going to need to encode at least some word-meanings to test this theory out. It does no good to merely dismiss the project as "lexicography." Actually, as we will see in §5, our attempt to model word meanings will actually embroil us in highly interesting and non-trivial theoretical questions.

2.2.3 Conclusion

To recapitulate, in this section, I argued that any Davidsonian theory is at most a translation into another language. To achieve this conclusion, in §2.2.1.1, I showed that an English-for-English semantic theory constitutes only a knowledge of the grammar of English.

I used this fact, in conjunction with other points, to argue, in §2.2.1.2, that any theory that claims to be exploiting a link between language and the world of non-symbols to explain comprehension is mistaken, at least if viewed from a scientific perspective. Thus, a Davidsonian theory is at most a translation from one language to another. From this, I concluded that a theory of comprehension ought to figure out how to use the manipulation of representation in such a way that what results is something that we would call a process of "comprehension."

In §2.2.2, I argued that translation is of merit in the construction of a scientific semantic theory, if the target language is a logical one. But, I then argued that this process of translation could not be equated with comprehension itself, for, if nothing else, translation does not give an account of word-meanings.

Chapter 3

Explicature

In §2, we examined the notion of what I called a "Davidsonian theory." The general form of such a theory is that a natural language sentence is given truth-conditions in a meta-language. I argued that this sort of theory amounted, from a scientific perspective to, at most, a translation from one kind of language to another. I also suggested that, for a translation to be of any value in a theory of comprehension, the target language would have to be a logical language.

So, a Davidsonian theory turns out to be a theory that pairs, in a formal way, natural language sentences with their logical language translations. Let us assume that, when a speaker utters a sentence, they intend to convey an "idea," which we will model as a logical language sentence. Let us call the idea that the speaker means to convey the *full propositional form* of their utterance.

Thus, one task of the hearer in comprehension is to recover the speaker's full propositional form. In this section, I would like to ask the following question. Does it make sense, in creating a model of a hearer, to try to give a translation algorithm that translates the natural language sentence carried on the acoustic signal to the full propositional form it communicates?

The answer will be no and the reason for this, as is much discussed in the literature, is that the information carried on the acoustic signal is underspecified with respect to the full propositional form conveyed. For example, consider the sentence (31):

(31) Everyone passed.

(32) Everyone [who took the test] passed [the test].

A speaker might utter (31), when the full proposition that they intend to convey is actually the one expressed in (32). That is, they would not intend to convey that every single person in the universe had just "passed." Nor are they saying that whoever it was they are talking about "passed"

simpliciter. There is a particular thing that was passed. So, to interpret (31), it seems that the hearer must add the bracketed material of (32) to the logical form for (31).

Thus, it is obvious that a Davidsonian theory is not going to work for the purpose of determining the speaker's full propositional form without at least some modification. This section will look at what sort of modification this will require, and whether it would be profitable, in modeling comprehension, to scrap the general organization of a Davidsonian theory and instead organize our theory in another way.

The ultimate conclusion will indeed be that the idea of a deterministic mapping between acoustic signal and full propositional form (i.e. a Davidsonian theory) should be abandoned.

In §3.1, we are going to review some of the ideas of Grice and the Relevance Theorists that will factor in to the later discussions. In §3.2, I will briefly discuss the semantic "principle" of compositionality. In §3.3, we will consider the concept of indexical words, like "I" and "she," and consider the implication of these for a theory of the derivation of the speaker's intended full propositional form. In §3.4 we will consider the question of how an underspecified acoustic signal like that of (31) is "enriched" to arrive at the full propositional form that it is thought to convey.

We will then conclude in §3.5 with a discussion of what all of this means for the organization of a theory of comprehension, commenting on the nature of the "semantics-pragmatics distinction." To foreshadow the conclusion very quickly for the benefit of the reader who might already be familiar with the requisite terms, I will suggest that the recovery of a fully complete semantic representation from a semantically underspecified acoustic signal will require a two-stage process. First, an *intermediate* logical form will be created carrying all of the semantic information available on the acoustic signal. Second, *pragmatic processes* will enrich the intermediate logical form to create a *fully specified* semantic representation.

3.1 Some Notes from Pragmatics

In this section, we review some work by Grice, Levinson and Sperber and Wilson that will be of relevance later in the chapter.

3.1.1 Grice's Conversational Maxims

In what is certainly one of the foundational papers of the field of pragmatics, Grice (1975) proposed to consider, "talking as a special case or variety of purposive, indeed rational, behavior," (p. 28). Talk is a "cooperative effort" and so communicators should obey, and assume that others are

obeying, certain conversational "maxims." For example, conversational participants should only say what they know to be true, and that for which they have appropriate evidence. They should not say more or less than is required. And, they should make contributions that are relevant.

Supposing that Annie will interpret Professor Bob's communications in terms that assume he is obeying the maxims, consider what she would do if Bob were to say (33):

(33) Here are two of my students, Connie and David. Connie is one of my good students.

Well, the addendum about Connie's being a good student cannot be seen as being superfluous, since Annie assumes Bob is not saying more than he needs to in order to get his point across. But, if there is some reason to point out which of Bob's students are his good ones, then, if David were a good student, Bob ought to have also mentioned *him* in the list, since otherwise he would be providing less information than he had.

So, although Professor Bob has not explicitly said so, he must mean to convey that David is *not* one of his good students. Grice referred to whatever it was that was "implied, suggested, [or] meant" by some statement, as opposed to what was more literally "said," as the *implicature* (Grice 1975, p. 24) of that statement. Thus, we say that Bob *implicates* that David is not a good student, even if he does not *say* so.

This example touches on two points that will be of recurring importance in this chapter and the next. The first is that, Bob, by exploiting "conversational maxims," can communicate a greater range of ideas than can be directly observed on the acoustic signal. In this case, Bob communicated that David is not among his good students, without actually "saying" so.

The second point is that our resolution of what Bob must have meant involved common sense reasoning. So, one might wonder just how deeply intertwined with general reasoning processes the resolution of these implicatures is going to turn out to be. Are a limited number of fixed principles going to be enough? Or, is modeling the resolution of implicature going to turn into the modeling of common-sense reasoning itself? We will address this question further in the next two sections.

Of course, this say-mean distinction is obviously rather fuzzy, at this point. But, Sperber and Wilson will sharpen it in an important way (cf. §3.1.3), and §4.2.2 will provide a completely formal, mathematical formalization of Sperber and Wilson's taxonomy.

3.1.2 Particularized Conversational Implicature

Consider, with respect to the question as to just how general a reasoning process is involved in inferring a speaker's communicative intention, Levinson's (2000) formulation of a Gricean distinction

between *particularized* and *generalized* implicatures:

(34) *The distinction between GCIs and PCIs*

a. An implicature i from utterance U is *particularized* iff U implicates i only by virtue of specific contextual assumptions that would not be invariable or even normally obtain.

b. An implicature i is *generalized* iff U implicates i *unless* there are unusual specific contextual assumptions that defeat it,
(Levinson 2000, p. 16, emphasis his).

I mention Levinson because he has spent much of his career discussing generalized conversational implicatures—i.e. those kinds of implicature that are relatively insensitive to context, and which are almost essentially built in to the lexicon. For an example of this kind of implicature, suppose I were to say that, "Some of our friends are coming to the party." Well, I would have flouted the maxim dictating that I should say as much as I can had I not said that, "All of our friends are coming to the party." So, I must have meant, in addition, that "Not all of our friends are coming to the party."

Levinson would argue that, by and large (i.e. barring "unusual specific contextual assumptions"), one is always going to make an inference from a use of "some" to a use of "some but not all." This implicature would then essentially be as context invariant as the lexicon itself.

What I find to be particularly interesting in this regard is that Levinson, who might be considered to be the researcher most interested in the resolution of the more context-insensitive generalized conversational implicitures—a man who we might quite possibly call *Mr.* Generalized Implicature—is readily willing to admit that there are a significant number of examples in which reasoning patterns are dependent upon "specific contextual assumptions that would not be invariable or even normally obtain." Or, in other words, the resolution of these implicatures depend upon highly particular reasoning. I take this as evidence of significant agreement among researchers that general human reasoning is playing a significant role in the analysis of a communication.

3.1.3 Relevance Theory

Wilson and Sperber's (1986, 1995) "Relevance Theory" consolidates the sort of thinking being discussed here, in which general reasoning processes are crucial in interpreting a speaker's communicative intention, into a "view." Communication, they say, involves a mix of coding-decoding (i.e. linguistic) and inferential processes where the full weight of human reasoning is employed to

understand an utterance in the typical case, and where inferential mechanisms can be used to keep the transmitted (linguistic-acoustic) signal shorter.

Wilson and Sperber make several proposals that will be of recurring importance for us. The first is that inferential processes might be necessary not only to resolve what was *meant* but also what was *said*. For example, consider the dialog in (35), adapted from Grice:

(35) a. Annie: What do you think of David's capabilities as a student?

b. Bob: Well, he does have pretty good hand-writing.

Of course, Bob is once again insinuating (and so implicating, in the terminology we are using) that David is not a good student. But, to even figure this out, one has to know who Bob is referring to with his use of the word "he." In terms of a distinction between what is "said" and what is "implicated," the resolution of "he" to "David" feels like it should be classified as figuring out what is said.

So, Sperber and Wilson propose to carve up the matter as follows. Assume that, underlying Bob's statement, there is a *full propositional form*, which we would model as a sentence in a logical language. In the full propositional form that Bob means to convey, there would be a reference to "Dave" himself. This is as opposed to the actual statement he speaks aloud, which contains a pronoun, which is underspecified in comparison, and could refer to any number of people. So, on the basis of (35b), Annie has some work to do to resolve "he" to "Dave" to recover Bob's intended full propositional form. Sperber and Wilson call the speaker's intended full propositional form the speaker's *explicature*.

Then, the *implicature*, following Grice, is the set of conclusions that follow from the explicature, as well as the set of assumptions which, if they held, would make the explicature a "relevant" one to communicate[1].

The theoretical contribution that, I feel, Sperber and Wilson are making here is the carving up of what the speaker must recover on the basis of an acoustic signal into concepts which can be characterized to a large extent using the tools of formal logic. That is, the explicature (i.e. full propositional form) is a sentence in a logical language. The implicature is a set of sentences, which, if assumed or inferred, make the explicature a "relevant" communication.

This seems to be an advance over Grice's more intuitive distinction between what is "said"

[1]While Sperber and Wilson want to focus on "relevance," one could alternatively say that the implicature is what is assumed in order that the hearer can view the speaker as obeying the Gricean conversational maxims in communicating the given explicature.

and what is "meant," with implicature defined, also intuitively, as what is "implied, suggested, [or] meant." The only, but perhaps rather serious, source of fuzziness with Sperber and Wilson's notion is the question as to what makes a "relevant" communication, a point that would draw Sperber and Wilson much criticism, as we will see. The hardening up of the notion of "relevance" will be addressed in §4.2.2.

Note that there is a purely terminological issue lurking here, which Bach (1994, 2005, 2006), for one, insists on belaboring. He says that he prefers to identify the material on the acoustic signal itself with the "what is said," and to call the full propositional form expressed on the acoustic signal part of the "what is meant." He also objects to calling the speaker's full propositional form the "explicature," because this term suggests that this form is "explicitly" represented on the acoustic signal, which it is not, it is implicit. Bach prefers the term *impliciture* for this purpose, instead.

Ultimately, this terminological debate has no impact on what our theory can predict, and consequently does not seem to be one of much importance to me. Still, contra Bach, I prefer to adopt Sperber and Wilson's terminology because we will inevitably need to make a distinction between what they are calling an explicature and what they are calling an implicature, so I see no reason to not use their terminology. Personally, it seems to me that Bach is trying to draw attention to himself and his work without actually making a theoretical contribution.

Returning to the actual notions of explicature and implicature, another important point that Sperber and Wilson raise is that the explicature and implicature must be resolved in parallel, as part of a single process, which they refer to as a process of "mutual parallel adjustment," (Sperber and Wilson 1998).

That is, if we assume, for the moment, that "he" in (35b) is resolved via an inferential process that employs word-/world-knowledge, then it would seem that the reason that "he" should resolve to "Dave" is that, if it does, then Bob has answered the question about Dave through insinuation. If "he" resolves to someone else, like "Harry," then Bob will presumably not have answered the question.

Now, I have not argued in any way conclusively yet that Annie, the hearer in (35), actually *does* need to use any complicated reasoning to resolve "he." If (35) is one's only evidence, one could propose a simple rule in which Annie simply takes "he" to be the last male referred to. This will not work in general (but, cf., p. 44 for an example that proves this). The point, thus far, is that, if one assumes pragmatic processes are required to resolve the explicature, then it seems that this process must be tightly wound up with, if not identical to, the process that resolves an implicature.

Now, of course, the need to resolve indexicals was recognized before Sperber and Wilson. But,

the indexical resolution process seems to have been viewed as more of an uninteresting blemish on the otherwise beautiful compositional semantic program, rather than as a topic of inquiry in and of itself, as Sperber and Wilson were suggesting it should be.

Moreover, Sperber and Wilson pointed out that the problem of explicature resolution goes beyond the resolution of indexicals. In the context of (37a), (36) is interpreted as though it were (37b):

(36) It will get cold.

(37) a. Bob, come and have your dinner. It will get cold.

b. It will get cold [soon, if you do not come and eat it].

Thus, it seems that whole semantic constituents must be added to the information available on the acoustic signal in order to determine the explicature. Carston (2000, 2004, 1999) would go on to make this point repeatedly, and with a variety of examples convincing enough to draw the attention of most semanticists. For example, in the introduction to their reader on otherwise traditional semantic topics, Davis and Gillon (2004) conclude that, somehow, this data suggesting a process of semantic enrichment will have to be incorporated into a mature semantic theory.

The process of semantic enrichment will form one of the main concerns of this chapter and the next. We will return to discuss the problem at length in §3.4 and then to provide much in the way of modeling it in §4.2.2.

One other noteworthy part of Sperber and Wilson's program is their attempt to Grice's nine conversational maxim, to a single maxim, that of *relevance*, which is assumed to be a general cognitive principle, similar to Grice's position that conversation is just another rational behavior.

In general, they define the cognitive relevance of a sensory phenomenon as follows:

(38) Relevance of a phenomenon (comparative)

a. Extent condition 1: a phenomenon is relevant to an individual to the extent that the contextual effects achieved when it is optimally processed are large.

b. Extent condition 2: a phenomenon is relevant to an individual to the extent that the effort required to process it optimally is small,
(Sperber and Wilson 1986, p. 153).

Then, in the case of ostensive communication, the hearer is essentially to assume that, "[t]he ostensive stimulus is the most relevant one the communicator could have used to communicate [the information they wanted to convey]," (p. 158). Thus, the hearer should assume that the speaker

is communicating a message with as great cognitive "effects" as possible, in a form which should require as little processing as possible.

It would seem that this account, as it is, is not adequate. Sperber and Wilson repeatedly stress that all their talk of relevance "would not be truly explanatory until the notion of relevance had itself been explicitly characterised," (p. 155). They evidently feel that they have accomplished this in (38) but it is hard to agree. Their claim now relies for "explicit characterization" on a measure of the size of "contextual effects," which is not forthcoming. Indeed, the predictive capacity of this theory has drawn criticism from researchers whose views we will consider later, such as Bach (2005) and Asher and Lascarides (2003).

What *is* interesting about this account is the effort to reduce Grice's nine maxims (grouped into four categories) to one or two kind of *super* maxims. In other words, given the particularized nature of particularized conversational implicatures, a list of all of the concerns that go in to resolving ex-/implicatures would go on growing without end. Maybe what we need then—rather than nine axioms grouped in four categories, or attempts to reduce the nine to eight, or else arguments that there really ought to be ten—is some super axiom that somehow acts as a generalization of them all. We will return to this idea in §4.

3.1.4 Conclusion

Grice (§3.1.1) pioneered the idea that conversation might be just another instance of rational human behavior, in which human reasoning could be exploited by the speaker to communicate more than he actually puts on the acoustic signal.

Levinson (§3.1.2) was of interest as a marker of how widely regarded has become the idea that highly particular reasoning patterns might be necessary for the resolution of the speaker's meaning.

Sperber and Wilson (§3.1.3) were of particular interest for suggesting that general inferential processes were required, not only to discover the speakers' implicature, but also the full propositional form that the speaker intended to communicate, which is also referred to as the speaker's explicature. Also, of note is that Sperber and Wilson suggest that the resolution of explicature must take place in parallel with the resolution of the speaker's implicature.

§3.3 and §3.4 will argue that Sperber and Wilson are correct about the pragmatic nature of the resolution of the speaker's explicature. This conclusion will have significant ramifications for the structure of semantic theory.

3.2 The Principle of Semantic Compositionality

The "principle" of semantic compositionality, recounted in (39), is tightly tied up in the question of the adequacy of the Davidsonian program:

(39) The "meaning" of an "expression" is a "function" of, and only of, the meanings of its parts and their mode of combination.

Now, as Pelletier (1994) notes in his discussion of the topic, this slogan is essentially meaningless, because it does not tell us what is meant by "meaning," nor what sort of "function" is allowed.

Thus, there is no point, it seems to me, in discussing (39) in isolation from definitions of these terms. But, there are some very natural ways to define them, and plenty of profitable discussion can take place by considering instances of (39) once we have done so.

Before we do this, however, I would like to note that one major distinction which is, as far as I know, never made is as to whether or not the "expressions" we are talking about are natural language expressions or the logical language representations we often use to represent them. That is, suppose that the translation of (40), as said by Annie, into the logical language we are employing is (41):

(40) I fish.

(41) $fishes(\text{ANNIE})$

Then, do we intend that (39) should apply to (40), or to (41)?

These are entirely different questions. If we are asking whether the model-theoretic truth conditions of (41) are a function of only the parts, then the answer should trivially be *yes!*. The Fregean sorts of logical languages that we use to translate natural language are specifically chosen for this purpose. That is, we purposefully choose to use languages that we know we can give Tarskian truth-conditions for—i.e. those which do not contain indexicals, which are fully specified, and for which the values of complex expressions can be gotten by functional application from the values of their parts (cf. §2.2.2).

The question that everyone is *really* talking about when they discuss compositionality is whether or not Tarskian truth-conditions can be given to a *natural language sentence*, like (40), on the basis only of its parts. But, obviously, if we identify "meaning" in (39) with "model-theoretic truth-conditions," of the sort discussed in §2.1.1, and "expression" with "natural language sentence," we will get (42), a version of the principle which is not only falsifiable, but also false:

(42) The model-theoretic truth-conditions of a natural language sentence are a function of, and only of, the meanings of its parts and their mode of combination.

To see that (42) is false, consider (43):

(43) Every boy loves a girl.

This sentence is structurally ambiguous, as there might either be one girl that all boys love, or else one girl per boy. This is not news. Pelletier (1994) raises this example and notes that people who feel warmly about the principle of compositionality have a response. That is, one way to save (39) is to say that the "meaning" is not the truth-conditions of the sentence, but instead a *set* of propositions that (43) could be expressing. Here, the presumption that some pragmatic process downstream will pick out the one that the speaker is thought to be expressing.

So, the point is not that (43) disproves the principle of compositionality (39). It seems as nothing could do this as (39) is not making any falsifiable claim. The point is that the strongest version of (39), i.e. (42), is false.

Also note that it is possible to define the terms of (39) so that it is trivially true. That is, we can define the "meaning" of an expression to be the expression itself. In that case, (39) is, of course, true.

So, the lesson we have learned is this. The "principle" of semantic compositionality is but a meaningless slogan unless some crucial terms are defined. When applied to natural language sentences, the only kind of expressions for which any discussion could be of interest, the strongest version of the thesis is false. But, the weakest version is true. So, the question becomes, is there any meaningful version of the principle which is both tenable and strong enough to be of interest?

3.3 Indexicals

The reason that any version of the compositionality principle (39) would be false is that natural language expressions—whether sentences, noun phrases, or whatever—occur in contexts, broadly construed[2], and, depending on how one defines "meaning," the meaning of an expression is liable to vary with context.

Indexical expressions—such as "I," "you," "here," "now," "she," etc.—in model-theoretic terms, are expressions that "refer" to different individuals in the universe of entities, depending on the context of utterance. They are one of the first kinds of expressions in which the context-sensitivity

[2]In a narrow sense, things like, who the speaker is, and what time it is are part of the context. But, in a broader sense, we could say that, for example, the pronoun "she" occurs in a different (syntactic) context in, "She is my sister," than it does in, "She is my mother." The difference in syntactic context may affect the hearer's guess as to who is being talked about.

of natural language sentences was realized. They are interesting to consider because, if "meaning" is defined weakly enough, they can be accommodated by the principle of compositionality.

But, the cost of this is that "meaning" cannot be equated with the speaker's full propositional form, i.e. the speaker's explicature. Thus, even if the compositionality thesis can be saved, the role of pragmatic inference in the resolution of explicature is undeniable.

Bar-Hillel (1954) discussed indexicals at length before modern linguistics had even really begun in earnest. He essentially identified the study of "pragmatics" with "the investigation of indexical languages," (p. 78). This view found an influential ideological backer in Montague (1968, 1970b), who not only did the same, but also showed how to give truth-conditions for an expression containing indexicals.

Montague's (1968) first treatment of indexicals was extensional. An extensional Montagovian solution to the problem of indexicals would work like this. Translate (44) into (45):

(44) I fish.

(45) $fish(\text{I})$

Note that this translation is context-insensitive, because "I" is just translated as I. Similarly, we would translate "you" as YOU and "she" as SHE, etc.

Now, I is not a constant. It is an *indexical*. This means that it is a function that takes *context of use* argument and yields an element in the universe of the model being used to interpret the statement. So, maybe in context c_1, the value of I is (the value of) ANNIE, while in context c_2, the value of I is (the value of) BOB. The value of the context is part of the interpretation (along with the universe and the valuation function)[3]. So, with the value of the context as an argument, I yields an individual in the universe, which gives $fish(\text{I})$ a truth-value. Thus, we can indeed say whether or not a sentence with an indexical is true or false.

An indexical works a lot like an intension, except that an indexical takes as its argument a context of use, and an intension takes as its argument a possible world index. To see the difference, consider that we could be in a possible world where Bob fishes but Annie does not. So, even within the same "world," (44) can be true or false depending on who is uttering it, which is reflected in the context. So, in an intensional framework (Montague 1970b), an indexical is a function that takes the context as its argument and yields an intension as its value.

[3]The interpretation can be seen as a function from truth-conditions (i.e. the string on the right side of "is $True$ if and only if") to truth (i.e. either the value *true* or the value *false*).

Kaplan (1977) takes a similar view, distinguishing between the *character* and the *content* of an indexical. An indexical's character is something that it keeps across contexts. The character of a sentence is not the sort of thing that can be true or false. But, a character is a function from a context to a content. And, the content of a sentence *can* be true or false. So, the content of "I" in a given context might be Annie. And, in that context, the content of, "I fish," will be false (because Annie does not fish).

In this parlance, we find a concept that we can substitute for "meaning," and still preserve the compositionality principle of (39):

(46) The character of the whole is a function of the character [*sic*] of the parts, (Kaplan 1977, p. 760).

Now, Montague's avowed goal (according to his exegetors, such as Dowty, Wall and Peters [1981]) was not empirical[4], but mathematical, or philosophical. That is, his goal was only to show how a language with an unrestricted rewrite grammar, even one with indexicals, could actually be given meta-language truth-conditions in a rigorous way. If this is one's goal, then, with respect to the challenge posed by indexicals, this goal has been met.

Now, what about if one's concern is scientific, and therefore psychological? What relevance does this sort of solution have in that case? The obvious way to use this concept in a theory of comprehension is to have Davidsonian theory that maps a natural language sentence to a *character*. Then, we have some way to represent the context, and from this context, the character can be mapped to a content. The content is a full propositional form, so we would have what we are after.

But, it is not hard to see that, as a general scientific theory of how indexicals and pronouns are resolved—one concerned with how comprehension actually takes place—this approach is going to run in to problems. Consider the difference between (47) and (48):

(47) If Connie keeps flirting with Emily's boyfriend, she is going to be furious.

(48) If Connie keeps flirting with Emily's boyfriend, she is going to enrage her best friend.

I will assume we all would resolve "she" in (47) as referring to Emily[5], and "she" in (48) as referring to Connie. That is to say, the explicature for (47) would contain a *discourse referent* (Karttunen 1968, 1976, for elaboration on the notion of a discourse referent, cf. ff. 6) for Emily

[4]Though, I am yet to figure out what coherent non-empirical sense could possibly be given to a claim to have found the "proper" treatment of "ordinary" English (cf., Montague 1974).

[5]It would be more precise to say that "she" refers to "what 'Emily' refers to," but I will be sloppy for the sake of exposition.

in place of "she," while the explicature for (48) would contain the discourse referent for Connie in place of "she."

It furthermore seems obvious that the reason that "she" must refer to Emily in (47) is because getting furious is a stereotypical reaction to having one's boyfriend flirted with, and not of flirting. Similarly, (48) must refer to "Connie," because enraging one's best friend is a stereotypical result of flirting, and not of having one's boyfriend flirted with. In other words, the reasoning involved relies crucially on word-/world-knowledge. Certainly one cannot give some simple syntactic rule, such as one that says that "she" should refer to the last mentioned female, that can handle the resolution of this discourse referent (this is the example that was promised, on page 38, to support the idea that pragmatic processes are indeed necessary to resolve indexicals).

Also, it would be ridiculous to suggest, in the spirit of Montague's idea, that, in understanding one of these statements, one computes the index of the context of use, say 2343123345 in this case, and then uses that index as an argument to the indexical SHE, so that SHE(2343123345) = EMILY.

Even if we break the context into an array of values, such as one for the speaker, one for the hearer, one for the location of utterance, etc., as Kaplan does, and as Lewis (1970) does, we are still going to get nowhere, because the difference in who "she" refers to is not a function of *this* kind of context, but instead is resolved on the basis the syntactic context in which "she" occurs, in conjunction with the hearer's world-knowledge.

And, this problem is not limited to "indexicals" in the traditional sense of the word. As Carston (1999) points out, even proper names function essentially like highly specific indexicals. For example, consider (51):

(51) George Bush is the worst president I have ever seen.

Most speakers in 2008 would have immediately assumed that "George Bush" in (51) referred to the discourse referent for "George *W.* Bush," rather than "George *H. W.* Bush," because at that time there was so much discussion about how intensely disliked the former, rather than the latter, had become. So, world-knowledge must be used even in resolving the discourse referent for a proper name.

---

[6] A discourse referent is a constant symbol that is stored and associated with a given entity. For example, the discourse (49) might be represented as (50), where d_{john} is a persistent constant is used to represent everything that a person thinks or knows about the entity "John":

(49) John likes Mary. John likes Sue.

(50) $likes(d_{john}, d_{mary}) \wedge likes(d_{john}, d_{sue})$

To take yet another case, consider the theory of referring expressions proposed by Gundel, Hedberg and Zacharski (1993). In this theory, a noun phrase with a definite article—e.g., "the cat," "the black dog"—is one that informs the hearer that they should be able to uniquely identify the speaker's intended discourse referent on the basis of being told the nominal alone.

So, for example, suppose I were asked (52):

(52) Have you read the paper?

If the context of this utterance were that I was just arriving to a semantics class in which we had been assigned a paper to read, I would create a full propositional form for (52) which asked about the discourse referent associated with that paper. If the context of utterance were that I was at my parents' house, where there is a newspaper delivered daily, I would probably create a full propositional form that included the discourse referent of that newspaper. This is another way in which pragmatic inference is required to generate the full propositional form.

Now, one could extend the Montagovian method of dealing with indexicals to also deal with this challenge from proper names as well as definite noun phrases. Such an analysis might be perfectly adequate from an abstract standpoint. But, it tells us nothing about what the process of resolving an indexical or pronoun involves in the course of actual human language use.

Note that, if the problem of the resolution of "indexical"-type items actually incorporates all proper names as well as all definite noun phrases, then the influence of context in the resolution of a discourse referent for a given noun phrase seems more the norm than the exception.

I will assume that one must concede, on the basis of the discussion that has just transpired, that *pragmatic processes*—i.e. processes that that are a function of word- and world-knowledge—are indeed crucially involved in the resolution of discourse referents. The question then is, what is the minimum amount of revision to a Davidsonian theory required to accommodate this fact? It would seem that all that would be required is that Kaplan-style "characters" be assigned to, perhaps, all noun phrases. Then, some process of pragmatic enrichment could be used to create contents out of these characters.

This would be a concession that pragmatic processes are crucial in the resolution of the speaker's full propositional form. However, the Davidsonian program is arguably little changed by all of this. Instead of translating (53) to (54), where d_{gwb} is a constant discourse referent, we would instead translate it as (55), where $?_{gb}$ is perhaps a placeholder, for a discourse referent to be filled in by the context:

(53) George Bush fishes.

(54) $fishes(d_{gwb})$

(55) $fishes(?_{gb})$

So, maybe some pragmatic process would be needed to resolve $?_{gb}$ to d_{gwb} in 2008. But, as for the practical effect on the day-to-day business of the semanticist, this seems to be a technicality that can rather be ignored. That is, (54) and (55) have exactly the same syntax and so are essentially from the same language.

But, as we are going to see in the next section, not only is pragmatic inference necessary for the derivation of the full propositional form, there is good reason to think that, in a fully complete theory of comprehension, the sort of form that can be created on the basis of a syntactic parse will be an "intermediate representation" of a totally different kind—i.e. will be a statement from a language with a very different syntax—than that of the full propositional form. In such a case, there will be little point in trying to maintain any version of the compositionality principle (39), as we will see.

3.4 Free Enrichment

Consider the following mock-up of a faulty argument, which is adapted from Carston (2000):

(56) a. If it is raining, we cannot play tennis.

b. It is raining in Vancouver.

c. We cannot play tennis.

d. We cannot play tennis in Arizona.

e. Because it is raining in Vancouver, we cannot play tennis in Arizona.

Obviously, (56e) is absurd, at least if this is meant as a "common sense" argument. (Here I have tacitly used the assumption, which I find reasonable and do not consider to be the source of the derivation of the absurdity, that if something is happening in a particular place, then it is happening simpliciter. And, if something cannot happen simpliciter, then it cannot happen in any particular place.)

Now, Annie might well utter (56a) in natural conversation, perhaps to a newcomer to the sport who does not know about the appropriate conditions for play. But, would Annie, an Arizona resident, cancel a game of tennis upon hearing there were rain, as there almost always is, in Vancouver? Of course not.

The problem with this argument seems to be that we do not interpret the natural language (56a) literally. We seem to actually interpret it as though it said (57):

(57) If it is raining *where we want to play tennis (and when we want to play tennis)*, then we cannot play tennis.

So, if this analysis of this faulty argument is correct, then it seems that the hearer is adding logical constituents to the full propositional form that is taken to have been communicated which do not correspond to any part of the acoustic signal. Further, these additions on the part of the hearer are clearly dependent on world-knowledge, and so pragmatic processes.

Carston barrages us with examples intended to illustrate that this process of semantic enrichment—which she calls "free enrichment"—is rather widespread. Mixing some of hers with some of my own:

(58) a. i. It will take [a significant amount of] time for your knee to heal.
ii. Emily has a[n unusually high] temperature.
iii. Something [out of the ordinary] has happened.
iv. What time are you done [work] on Sunday?

b. i. She walked right up and kissed me. That was pretty cool [of her], considering she was only eighteen.
ii. She walked right up and kissed me. That was pretty cool [for me], considering she was only eighteen.

c. i. There must be [roughly] fifty people in here.
ii. John has [exactly] four children.

d. i. Everything [being served at this meal] tastes so good!
ii. Don't worry, everyone [who took the test] passed [the test].

(59) a. Sue finished her drink and [then] left the bar.

b. Mary left Paul and [as a consequence] he became depressed.

c. Bill beat Marnie 10-0 and [i.e. even though] he hasn't practiced in years.

I have divided these examples into two major groupings. In the examples of (58), the semantic constituent or constituents that the hearer contributes are subclausal. That is, the semantic material added is within a single sentence—modifying nouns, verbs and adjectives. In those in (59), the

added material is between full clauses. Relationships between full clauses are called *discourse relations* (cf., e.g., Mann and Thompson 1987).

This distinction is relevant given the young history of the study of this sort of topic. The latter kind of explicature has been given at least one very thorough treatment—in Asher and Lascarides 2003[7], which we will consider in §4—while the former, to my knowledge, has not.

I will assume that these examples are conclusive in the sense that there is no need to debate whether there is semantic enrichment going on. The main debate, for those who discuss the issue, seems to be what the implications of this kind of enrichment are for the overall structure of a theory of comprehension. And, this is the debate that I will review.

Bach (2005) argues that the Relevance Theorists have overblown the significance of these examples. He makes a Gricean distinction between what is "said" and what is "meant", and concludes that no substantial change to our theoretical apparatus is required because the enriched aspects of underspecified sentences are part of what is "meant" and not what is "said."

I think that Bach is, as before, merely playing a game with words. The idea being set up by Sperber, Wilson and Carston is that linguistic theory ought to posit that an utterance communicates a "full propositional form," and that this full form is used to determine the implicature. Saying that the full propositional form is only "meant" and not "said" gets us no closer to recovering it. These smoke and mirrors only serve to distract from the real question at hand: what is the nature of the logical form recoverable directly from the acoustic signal, and how does this relate to the full form?

Stanley (2000) does make an attempt to answer this question. His arguments against Carston's position are based on an erroneous interpretation (whether deliberate or otherwise) of her proposal. He says that Relevance Theorists propose the following sort of theory. In (60), there is a hidden restriction on the domain of quantification, as it presumably does not mean that *every* bottle in the universe is green.

(60) Every bottle is green.

This restriction, says Stanley, is supplied by the context. So, for example, the denotation of bottle, relative to context c, would be:

(61) $Denotation$("bottle") relative to a context c = the set of bottles that are in the domain salient in the context c.

[7]On a terminological point, Asher and Lascarides do not use the term "explicature." They instead speak about recovering the "what was said" (p. 77), a notion essentially synonymous with "explicature." This is perhaps an attempt to distance themselves from Relevance Theory, which they allege, as I did in §3.1.3, has "problems" in its "predictive power" (p. 75).

But, this straw theory, Stanley points out, runs afoul in the following example:

(62) In every room in John's house, he keeps every bottle in the corner.

So, the set of bottles quantified over is dependent on the choice of room, for each choice of room. Thus, a theory including a statement like (61), which posits the domain of quantification to be a single, contextually supplied set of bottles fails because it only supplies one way to restrict the set of bottles, and so cannot accommodate there being more than one room quantified over.

The problem with Stanley's argument is that no one, aside from himself, is suggesting that "bottle" would have one denotation per context of utterance. What is being suggested is simply that the material that is available from the acoustic signal is *enriched as a function of context*. But, that function has not been proposed, at least not by the Relevance Theorists. Thus, the enrichment of (62), in the actual Relevance Theory-type solution, might be:

(63) In every room[, say r,] in John's house, he keeps every bottle [in r] in the corner [of r].

Instead of the straw theory, which he proudly demonstrates does not work, Stanley, who does accept the fact that *some* enrichment is going on, proposes that the solution to the problem is to place hidden "operators" in logical form. These operators can be filled in with the sorts of values we have seen in enriched forms. That is, corresponding to each aspect of square-bracketed content (i.e. content understood but not present on the acoustic signal) in (58) and (59), there must at one time have been a hidden operator variable, placed there by the parser.

For example, in (58c-ii), repeated in (64), we might have begun with some form containing an operator, like (65):

(64) John has ⌈ [exactly] four ⌉ children.

(65) John has ⌈ ? four ⌉ children.

(Here the ? is an operator whose value can be, and is, set to the logical language symbol "exactly" by some pragmatic process.)

Stanley sees this as way to unify Carston-type enrichment with a treatment of the indexical phenomena that we saw in the last section. That is, these operators are essentially unseen indexicals, whose values can be filled in by the same process of indexical resolution that we decided was necessary, at the least, in the last section.

The problem here is that the bracketed content of each example of (58)—or, for example, in (64)—represents only what content had to be filled in to make sense of the statement in *in that case*. The major question that arises with Stanley's approach is: in how many places *could* there have

been additional structure added, even though there had not been in the particular cases in question? How many of these operators would this require to be present in a given logical form? If it is, in principle, infinite, then Stanley cannot be right, because it seems we cannot say that the parser is creating semantic representations of infinite size.

If the number of sites at which content can be added in a given sentence is finite but large, then his solution is very awkward. But, if the number of sites is finite and small, then Stanley's proposal could well be made to work on these examples. And, I think Stanley has raised a valuable point in calling in to question just how "free" Carston's process of "free enrichment" really is. As far as I see, Carston has not demonstrated in any serious way that this enrichment is completely unconstrained.

It would certainly be a welcome result if all of the examples of explicature that can be found can be reduced to a few points of variation. Carston seems to have been facing a field of researchers uninterested in countenancing explicature as a worthwhile topic of discussion, and, in the process of getting their attention, may have aggrandized the case a bit. I think that, for those who accept, as I do, that Carston is pointing out an interesting and formidable phenomenon, the task immediately becomes to get a handle on it, and not to remain in a state of self-induced bewilderment.

But, I think that this is all that can be said in favor of Stanley. His list of examples is extremely cursory and does not even pretend to attempt to be comprehensive. He certainly has not demonstrated anything like an upper bound on the variation or the number of kinds of intraclausal enrichment.

To briefly recapitulate, what Stanley is proposing as an overall structure of linguistic theory is one in which syntax is mapped to semantics in a fairly deterministic fashion—i.e. in the Davidsonian style—where the only input of pragmatics is the filling in of operator variables. These can either be due to indexicals or else can be due to hidden operators, which function like indexicals.

If Stanley is suggesting that the syntactic parser can supply a *single* one of these operator-filled semantic forms, he is surely wrong, as is shown by phenomena of scope ambiguity:

(66) Every boy loves a girl.

(67) a. Please, do not sleep and pay attention. (Adapted from Bos 1996.)

b. Please, do not sleep and waste your whole day.

(66) is an example we reviewed above. It could be saying that there is a particular girl loved by all, or else that each boy has some, but not necessarily the same, girl that he loves.

(67) demonstrates convincingly, if it is perhaps unclear in the contrived (66), that the resolution of scope ambiguity relies heavily on word-/world-knowledge. Note that in what I take to be the

natural reading of (67a), *not* only has scope over *sleep*, whereas in (67b), *not* has scope over *sleep and waste your day*. The correct scopal configuration is not evident from the syntax but rather is resolved in the basis of the knowledge that, for one thing, one does not sleep and pay attention at the same time.

The representation of scope ambiguity cannot be accomplished by the insertion of operators. I am going to assume that Stanley would not suggest that the syntactic parser, which is assumed to not have access to world-knowledge or the context, would pick only one of the parses to pass on itself. He seems to be acknowledging, with his hidden variables left for pragmatic filling, that he agrees that this sort of decision is not to be built into the parser itself.

So, it would seem, if he were to insist on believing that what the parser produces is a form that is of essentially the same kind as the full propositional form, that he would have to propose that the syntactic parser provides the downstream system with a *set* of logical forms. That is, for, "Every boy loves a girl," which has two parses, the parser would have to pass on two different forms. If there were six different scopal combinations, the parser would create and pass on all six.

Computationally minded linguists note that this sort of approach does not seem to scale up particularly well. Consider, for example, (68), adapted from Bos 1996:

(68) A few politicians can fool many voters on some of the issues all of the time.

(68) has a nesting of 4 quantifiers and so $4! = 24$ possible ways to arrange the quantifier scope. So, the parser would need to pass on 24 different possible parses if we adopt the Stanley-inspired solution we are considering where the parser passes *sets* of logical forms. Unhappy with this, the aforementioned computationally minded linguists have sought a way for the parser to pass on a single representation, whose size grows only linearly on the number of nested quantifiers.

That is, a major trend of research has been the development of *underspecified logical forms*, which can *pass on* scopal ambiguities, without resolving them, to some pragmatic system downstream. That is, if the scopal configuration is not completely specified on the acoustic signal, then an underspecified logical form allows the syntactic parser to create a logical form in which the scopal configuration is precisely as underspecified as it was in the acoustic signal, so that the matter of settling the issue can be left for the system with access to context and word-/world-knowledge. This line of research began with Alshawi and Crouch 1992, Reyle 1993 and Bos 1996, and is now getting pretty large.

I think it would be worthwhile to demonstrate briefly how this sort of underspecified form works in the case of scopal ambiguities. I will do this because it will demonstrate that the underspecified

languages being developed are an altogether different kind of logical language than the languages that semanticists typically work on. That is, if one accepts the need for this kind of intermediate form, then one must admit a whole new kind of study into the field of linguistics (or at least semantics-pragmatics).

Consider again the sentence:

(69) Every boy loves a girl.

Now, (69) can be translated to logical language as either (70a) or (70b). I am going to switch to a kind of generalized quantifier notation for this example only, because it will simplify our discussion:

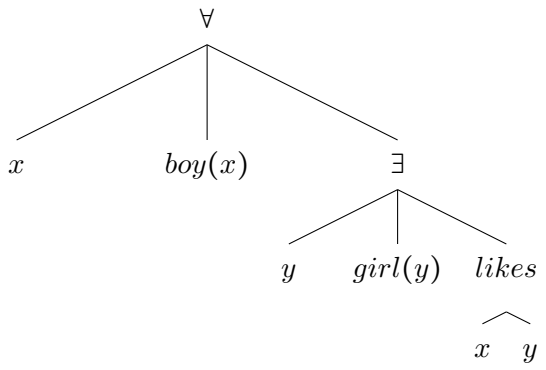
(70) a. $\forall\big(x, boy(x), \exists\big(y, girl(y), likes(x,y)\big)\big)$

b. $\exists\big(y, girl(y), \forall\big(x, boy(x), likes(x,y)\big)\big)$

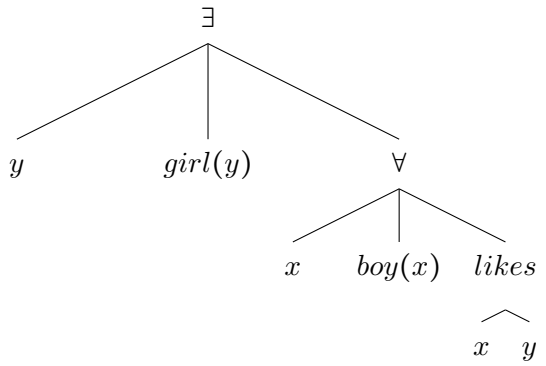
((71a), for example, says that, for all x, such that $boy(x)$, there exists a y, such that $girl(y)$, such that $likes(x,y)$.)

Now, it turns out to be easier to work with the parse trees for the forms of (70), rather than the logical formulae themselves. So, note that the parse trees corresponding to (70a), and (70b), are (71a), and (71b), respectively:

(71) a.



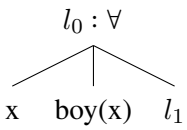
b.



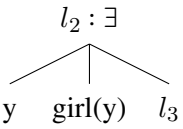
So, the idea is that we do not want the parser to have to pick between (71a) and (71b), but to pass on a single form that tells another system downstream to pick one or the other of them. So, we need the parser to give a set of constraints, u, on trees such that there are precisely two trees in the set of all trees with labeled nodes that satisfy u, namely (71a) and (71b).

So, we proceed as follows. Note that the common elements between (71a) and (71b) are those shown in (72):

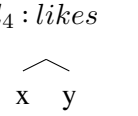
(72) a.



b.



c.



The idea is that these three little trees can be put together to make either (71a) or (71b). l_1 and l_3 are like attachment sites[8]. We can attach other trees to these. That is, if we assemble the pieces of (72) such that l_2 is attached at the site l_1, and l_4 is attached at the site l_3, we get (71a). If we assemble them such that l_0 is attached at l_3 and l_4 is attached at l_1, we get (71b).

Thus, the underspecified logical language form that generates either (71a) or (71b) is:

(73) $$\exists l_1, l_3 \begin{pmatrix} l_0 : \forall(x, boy(x), l_1) \wedge \\ l_2 : \exists(y, girl(y), l_3) \wedge \\ l_4 : likes(x, y) \wedge \\ outscopes(l_0, l_4) \wedge \\ outscopes(l_2, l_4) \end{pmatrix}$$

Here $outscopes(l, l')$ means that, in the final tree, it must be that l' is a descendent of l. If we assume that there are only three nodes that the labels l_i can refer to[9], then there are only two trees which satisfy (73), namely (71a) and (71b). Thus, there are only two formulae which satisfy (73), namely (70a) and (70b), as desired.

---

[8]The reader may recognize attachment sites from Tree Adjunct Grammars (e.g., Joshi, Levy and Takahashi 1975).

[9]The reader may be wondering what principle this assumption follows from. A discussion of the answer would require a level of detail that we otherwise do not need to get into.

Now, what do we make of this? Well, note that the language that the underspecified (73) form is written in is different than the language that the fully specified (70) forms were written in. Stanley's hope seems to be that the parser can produce a form which is alike in kind to the full propositional form, but which differs only in that it contains various indexical-like operators waiting to be filled. But, if we adopt the course of underspecified logical forms, then Stanley is clearly wrong because we have just agreed that the underspecified language is qualitatively different from the language it describes.

To recapitulate, this section began by giving Carston-type examples of cases in which the hearer must "enrich" the semantic information available on the acoustic signal to arrive at the full propositional form. I took these examples to be conclusive in that there must be *some* kind of enrichment going on with the only question up for debate being as to what the implications of this kind of enrichment would be for the structure of a theory of comprehension.

I concluded that the Stanley-esque solution of placing many hidden operators in logical form was either wrong (in case the number of operators needed was somehow infinite) or else a (perhaps bulky) notational variant of a solution involving "underspecified logical forms." For those who accept the desirability (or, perhaps, necessity) of underspecified forms, then our conclusion must be that the intermediate form created by the parser must be of a qualitatively different kind of language than the one that the full propositional form is written in.

Furthermore, if we accept the desirability (or necessity) of underspecified forms, then the principle of compositionality (39) that we are left with is the following:

(74) The *underspecified logical form* for a natural language expression is a function of, and only of, its parts and their mode of combination.

But, I think that one has to feel that (74) is so far removed from the original intention and spirit of the principle that there is hardly a point in maintaining it.

3.5 Conclusion (The Semantics-Pragmatics Distinction)

The question as to how to delineate the province of semantics, as opposed to that of pragmatics, has been the topic of much discussion. In this section, I would like to apply the preceding discussion of this chapter towards an answer to this question.

One can identify two potentially different debates. That is, some, such as Gazdar (1979) in

the introduction to a monograph on pragmatics, seem to try to define what "semanticists," as *researchers*, as opposed to "pragmaticists," as researchers, should spend their time doing. Others, such as Stanley and Carston, try to specify what a "semantic" *system* should be, and how it should relate to a "pragmatic" system.

Though they sometimes result in merely terminological debates, I think that these are important questions. The persistent interest in this debate, which has been going on for over thirty years, attests to the fact that the carving up of the pie of a theory of comprehension is a crucial step to moving forward with theory construction. And, unlike with ditch-digging, in this case, a profitable division of labor requires a fairly advanced theory of what is being worked on.

The discussion of this chapter suggests the following division between systems. We begin with a syntactic parse of the material on the acoustic signal into lexical items with syntactic structure. We then need one system, which I will tentatively call the "first" system, which takes this syntactic parse and yields an underspecified logical form, which includes all of the semantic content that is recoverable from the acoustic signal, and in which all information left ambiguous or underspecified on the acoustic signal is left ambiguous or underspecified in this preliminary logical form.

Then, another system, which I will call the "second" system, and which has access to the context and word-/world-knowledge, will take that underspecified form and, based on it, create the full propositional form that the speaker is hypothesized to have wanted to communicate. And, as the Relevance Theorists have pointed out, this process of enrichment, which yields the explicature, will have to take place *at the same time* as the hearer hypothesizes about the speaker's implicature.

Both Carston and Stanley refer to what I have called the first, and second, systems as the "semantic," and "pragmatic," systems respectively. The question is essentially terminological. But, I still think this usage should be branded as inappropriate.

If we consider works like Montague 1970a, 1974, and Kamp and Reyle 1993, to be paradigm examples of what we might call "semantics, traditionally construed," then what the first system is yielding is *not* the kind of form that semantics, traditionally construed, typically produces. This is because the first system produces underspecified logical forms. But, what semantics, traditionally construed, creates are *full propositional forms*.

In terms of how the division of labor between researchers should go, with regards to the explicature problem specifically, I think we can say the following. There must be people who devise the language that full propositional forms are written in. And, work in semantics, traditionally construed, is basically the state-of-the-art in this.

There must also be people who devise the language of the corresponding underspecified logical

forms that goes with the language for fully specified forms. The work by Alshawi, Crouch, Reyle and Bos mentioned earlier is the foundational work in this vein.

And, there must be people who work on the mapping from the underspecified to the full propositional form. It is the topic of this mapping that we will turn to in the next section.

Chapter 4

Two Models of Comprehension

In this section I am going to present two models of comprehension.

The first, given in §4.1, models the comprehension of a sentence under the idealizing assumption that a natural language sentence can be mapped in a context-invariant way to its logical language translation. In that section, I argue that the "comprehension" of a natural language sentence should be identified with the computation of the inferences that follow from its logical language translation, along with the word-/world-knowledge of the hearer.

Of course, I spent §3 arguing that the aforementioned (idealizing) assumption does not hold, and, moreover that its not holding has serious ramifications for the organization of a theory of comprehension. So, in §4.2, I will present a more complex model of comprehension that can resolve the speaker's full propositional form, or explicature, from an underspecified acoustic signal.

4.1 A Simple Model of Comprehension

4.1.1 The Simple Model of Comprehension

We saw in §2 that the Davidsonian program amounts to a translation from natural language to a logical meta-language. But, we also saw that this translation would not be sufficient to model the full process of comprehension because compositional semantics cannot model word-meanings and word-meanings, we decided, must be taken account of at *some point* in the process of modeling comprehension.

In this section we will see how this is done. The solution that I propose heavily resembles ideas from the literature on semantic "holism" coming from the philosophical and cognitive science

community[1], which we will discuss at great length in §5. It also resembles the structure of the method of representing semantic knowledge employed by the commercial Artificial Intelligence software developed by Cycorp Inc., which is headed by Lenat[2].

Let us make the simplifying assumption, to begin with, that the acoustic signal provided to the hearer is neither ambiguous nor underspecified. That is, make the simplifying assumption that the derivation of the full propositional form is not context-sensitive, even though I spent §3 arguing that it is. We will see in §4.2 how the discussion of this section can be expanded to accommodate the aforementioned problems caused by underspecification on the acoustic signal.

Recall that the problem discussed in §2.2.2 was that compositional semantics cannot distinguish between the differences in contribution to the "meaning" of a complex expression of component words within the same grammatical class. That is, compositional semantics, as we saw that Thomason explained, does not distinguish between, for example, the meanings of "walk" and "run." Thus, a Davidsonian theory, which at most explains how to give a compositional translation from one language to another cannot distinguish between these.

To put the matter another way, compositional semantics is what tells us that "Socrates" is making a similar contribution in, "Socrates is a man," as it is in, "Socrates is a Greek." But, it does not explain what difference it makes to Socrates whether he is a man or a Greek.

To model this distinction, we are going to model the *word-/world-knowledge* of a hearer, Henry, as a set of sentences in a logical language, LL. Basically, our primary requirement on whatever logical language we choose to use as our LL is that it allows the definition of a derivability relation (e.g. $\vdash$) that is sound. I.e., one that does not lead us from "true" premises to "false" conclusions.

We will assume that Henry understands the natural language NL, which has an unrestricted rewrite grammar. And, each sentence in NL, we are assuming, can be translated to a sentence in LL by Henry's *parser* in a context-insensitive way. Unlike Montague (1970c), the translation to LL here is done because LL is an important "level" of linguistic representation. It is not done for the purpose of yielding meta-language truth-conditions for the original NL sentence. In fact, we will not be using a meta-language at all.

Now, (75) depicts a set, **K**, of sentences in LL, which represent Henry's word-/world-knowledge about what it means to be a man, an $animal$ and a $greek$:

---

[1] Cf., e.g., Harman 1982, Sellars 1974, Field 1977, Block 1986, 1993 and Fodor and Lepore 1992.

[2] See http://www.cycorp.com. And, cf. Lenat 1997.

$$(75) \quad \mathbf{K} = \left\{ \begin{array}{c} \forall x\ man(x) \rightarrow \big(mortal(x) \wedge two\text{-}legged(x) \wedge animal(x)\big), \\ \forall x\ animal(x) \rightarrow \neg plant(x), \\ \forall x\ greek(x) \rightarrow \big(citizen\text{-}of\text{-}greece(x) \wedge wears\text{-}toga(x)\big) \end{array} \right\}$$

That is, "men" are "mortal, two-legged animals;" "animals" are not "plants;" and, "Greeks" are "citizens of Greece" and "toga-wearers."

I will always use **K** as the symbol to represent a speaker's or hearer's word-/world-knowledge, and will sometimes refer to **K** as the person's *knowledge-set*, or *belief-set*.

There is no distinction being made, here, as to whether the fact, encoded in **K**, that all men are mortal is knowledge about how to use the *word* man, or whether it is an element of *world-knowledge* about men. That is, there is no distinction between word- and world-knowledge being made here, nor is a distinction being made between beliefs and knowledge. (But, cf., §5, where we will make a distinction between statements that represent conventionalized word-meanings, and those which do not.)

Now, when Henry hear a sentence, U, of NL, the first thing he does to understand it is that he parses it to yield the LL form ϕ_U. He then computes the set of all inferences that follow from ϕ_U, in conjunction with $\mathbf{K}$[3].

To discuss this further, we must introduce some notation. Let,

$$Cn(X) = \{\phi : X \vdash \phi\}$$

Here, X is a set of sentences in LL. ϕ is a single sentence of LL. $X \vdash \phi$ is read as "X derives ϕ." $\vdash$ is associated with a fixed set of inference rules. So, $X \vdash \phi$ means that, if one assumes the set of sentences X, then one is allowed, by associated set of inferences rules, to conclude ϕ.

$Cn(X)$ is read as "the closure with respect to inference of the set X." So, $Cn(X)$ is the set of all things that one can infer from the set of sentences X, according to the inference rules associated with $\vdash$. In a formal logical setting, such as a textbook on logic, one would have to actually specify which inferences $\vdash$ licenses. But, I would prefer to work on an intuitive level.

Suppose, for example, that we have the following set L,

$$L = \left\{ \begin{array}{c} \forall x\ woman(x) \rightarrow mortal(x) \\ woman(\textsc{Helene}) \end{array} \right\}$$

[3] This is not literally possible, but I will speak as though it is as an idealization. A discussion of computational plausibility will occur at the end of this section.

Well, since L posits that all women are mortal, and HELENE is a woman, by the rules of logic that we intuitively know, we should conclude that HELENE is a mortal. We express this in our Cn notation as follows:

$$mortal(\text{HELENE}) \in Cn(L)$$

This is read as saying that, "The LL sentence $mortal(\text{HELENE})$ is in the set $Cn(L)$, which is constitutes the closure with respect to inference of the set L." Or, in other words, $mortal(\text{HELENE})$ is an inference that follows from the set L.

Then let,

$$I(X, \phi) = Cn(X \cup \{\phi\})$$

So, $I(X, \phi)$ is the set of all sentences that follow as logical conclusions when one assumes both every sentence in the set X, as well as the sentence ϕ.

Now, we are ready to discuss the second step in the model of comprehension that we are building. The first step, again, is for the parser to translate the natural language $U \in$ NL to the logical language form $\phi_U \in$ LL. The second step will be to compute $I(\mathbf{K}, \phi_U)$. That is, the second step is to compute all of the inferences that one draws when one assumes everything in **K** as well as ϕ_U.

So, again, Henry's knowledge is **K**, as described in (75). Now, suppose someone says to him, "Socrates is a man." We assume that the parser translates this to LL as $man(\text{SOCRATES})$. So, Henry's next step in understanding this utterance is to compute $I(\mathbf{K}, man(\text{SOCRATES}))$, which is such that:

$$(76) \quad I(\mathbf{K}, man(\text{SOCRATES})) \supset \left\{ \begin{array}{c} mortal(\text{SOCRATES}) \\ two\text{-}legged(\text{SOCRATES}) \\ animal(\text{SOCRATES}) \\ \neg plant(\text{SOCRATES}) \end{array} \right\}$$

Here, $A \supset B$ is read as, "A is a superset of B." This means that everything in the set B is also in the set A. So, the set of inferences that are induced by $greek(\text{SOCRATES})$, in conjunction with **K**, *includes* (but is not limited to) those listed in (76).

In contrast, if someone says to Henry, "Socrates is a Greek," which his parser translates as $greek(\text{SOCRATES})$, he will instead compute:

$$(77) \quad I(\mathbf{K}, greek(\text{SOCRATES})) \supset \left\{ \begin{array}{c} citizen\text{-}of\text{-}greece(\text{SOCRATES}) \\ wears\text{-}toga(\text{SOCRATES}) \end{array} \right\}$$

To simplify notation, let us now let,

$$I_{\mathbf{K}}(\phi) = I(\mathbf{K}, \phi)$$

Then, it seems, intuitively, as though, in $I_{\mathbf{K}}(man(\textsc{Socrates}))$ and $I_{\mathbf{K}}(greek(\textsc{Socrates}))$, we have got two different values that somehow illuminate a difference between being a man and being a $greek$. But, how are we to formalize this intuition?

I want to consider this question from two points of view. First, how do we conceptualize the process of comprehension in a way that makes use of the concept of $I(\mathbf{K}, \phi_U)$ that we have just discussed? And, second, what kind of predictions does this kind of theory make?

We might start by saying that, because $I_{\mathbf{K}}(man(\textsc{Socrates}))$ and $I_{\mathbf{K}}(greek(\textsc{Socrates}))$ have different values (i.e. they are different sets that are not equal to one another), we have found a way to tease apart the difference between being a man and being a $greek$. But, this cannot be quite right because, $man(\textsc{Socrates})$ and $greek(\textsc{Socrates})$, considered as *sentences*, were different entities to begin with, even before the application of $I_{\mathbf{K}}$.

But, still, intuitively, it seems to me that we have made progress because, while $man(\textsc{Socrates})$ and $greek(\textsc{Socrates})$ differ on only one lexical item, $I_{\mathbf{K}}(man(\textsc{Socrates}))$ and $I_{\mathbf{K}}(greek(\textsc{Socrates}))$ differ in that they contain different statements, and these differences include many different lexical items. The values of $man(\textsc{Socrates})$ and $greek(\textsc{Socrates})$ under $I_{\mathbf{K}}$ are "more different" than they were to begin with.

To see how we can make use of this fact, suppose that Henry had observed the following conversation:

(78) a. Annie: Is Socrates a citizen of Greece?

b. Bob: Socrates is a Greek.

Here, Henry recognizes that there is a salient question posed by Annie as to whether or not $wears\text{-}toga(\textsc{Socrates})$. Now, because it is in Henry's knowledge that Greeks are citizens of Greece, he is going to recognize that Bob's response in (78b) is probably intended to answer the question affirmatively. And, crucially, if Bob's response had been that, "Socrates is a man," he would not have answered Annie's question. Compositional semantics alone does not predict this.

Now, one is likely to be skeptical of this theory if they associate $I_{\mathbf{K}}(man(\textsc{Socrates}))$ too closely with the "meaning" of $man(\textsc{Socrates})$, in some way that inherits all of the demands that have been placed on a theory of "meaning" over the years. That is, it is easy to imagine the following retort:

> You were supposed to give a theory of the "meaning" of a sentence. But, for you, the meaning of $man(\text{SOCRATES})$ is $I_{\mathbf{K}}(man(\text{SOCRATES}))$. But, $I_{\mathbf{K}}(man(\text{SOCRATES}))$ is a set which contains sentences. So, we need to know the meanings of *those* sentences, if we are going to be able to get the meaning of the original sentence. But, it is precisely the matter of how to give meanings to sentences that you are supposed to explain. So, it seems as though you are setting up a recursive definition of "meaning," in which the meaning of one sentence is given in terms of others. But, you have not provided a "base case," i.e. a sentence whose meaning can be gotten without looking at others. So, you are setting up a situation of infinite regress; our search for the meaning of $man(\text{SOCRATES})$ will never cease[4].

The first thing to note is that I am only ever using the term "meaning" informally. I have not defined "meaning" as a technical term. $I_{\mathbf{K}}$ is the technical term we are employing. What I am proposing is a theory of comprehension that focuses on sentences themselves, rather than their "meanings." That is, a sentence is of value in and of itself, without having to ask what its meaning is.

The important thing for Henry, as the hearer of the dialog (78), is that he understands that an answer to the salient question as to whether or not $wears\text{-}toga(\text{SOCRATES})$ has been given, and that it has been in the affirmative. He does not need to know the "meaning" of this sentence to understand that. He only needs to know which sentences are in $I_{\mathbf{K}}(greek(\text{SOCRATES}))$.

Similarly, consider the case of an early human like Tarzan, whose word-/world-knowledge is as in (79):

$$\text{(79)}\quad \mathbf{K} = \left\{ \begin{array}{c} \forall x\ caught\text{-}food(x) \rightarrow will\text{-}eat(x) \\ \forall x\ will\text{-}eat(x) \rightarrow will\text{-}feel\text{-}full(x) \\ \\ \forall x\ lost\text{-}food(x) \rightarrow will\text{-}not\text{-}eat(x) \\ \forall x\ will\text{-}not\text{-}eat(x) \rightarrow will\text{-}feel\text{-}hungry(x) \end{array} \right\}$$

Then, we will have,

$$\text{(80)}\quad I(\mathbf{K}, caught\text{-}food(\text{TARZAN})) \supset \left\{ \begin{array}{c} caught\text{-}food(\text{TARZAN}) \\ will\text{-}eat(\text{TARZAN}) \\ will\text{-}feel\text{-}full(\text{TARZAN}) \end{array} \right\}$$

[4]We will see further arguments like this in §5.

$$(81) \quad I(\mathbf{K}, lost\text{-}food(\textsc{Tarzan})) \supset \left\{ \begin{array}{c} lost\text{-}food(\textsc{Tarzan}) \\ will\text{-}not\text{-}eat(\textsc{Tarzan}) \\ will\text{-}feel\text{-}hungry(\textsc{Tarzan}) \end{array} \right\}$$

Now, suppose we assume that $will\text{-}feel\text{-}full(\textsc{Tarzan})$ is associated with positive affect in Tarzan's brain. And, suppose that $will\text{-}feel\text{-}hungry(\textsc{Tarzan})$ is connected to negative affect. Then, even though (80) and (81) are mirror images of each other, Tarzan's being informed that he has caught food will differ from his being informed that he has lost food, because one statement will lead to a positive affect in Tarzan, and the other will lead to a negative affect. Again, the "meanings" of these statements are not required to achieve this effect.

So, I have been trying to explain how $I_{\mathbf{K}}$ causes us to conceptualize language. But, science, unlike post-modernist sociology, is not in the business of choosing arbitrary conceptualizations for the world. We evaluate competing conceptualizations by comparing the predictions that they make.

So, here is a prediction. Tell some experimental participant that, "All blargs are nargs." Then, tell them that, "Park is a blarg." We would assume, as a bridge hypothesis connecting observables to abstract entities, that the logical language translation of these two statements will be added to the **K** of our participant. Then ask them, "Is Park a narg?" They are predicted to say yes, because the answer to the question, i.e. that "Park is a narg," will be in the closure with respect to inference of their **K**.

Of course, this is a mundane prediction. It is one which we already know is probably true. But, the requirement that science puts on us is not necessarily to make, in each case, fantastic predictions that require mountaineering expeditions during solar eclipses to test. We only need to make a superset of the predictions of the competing theories, however mundane these may be.

A compositional semantic theory makes no predictions. Well, Davidson 1967 argues that it does make one kind of prediction, which he admits is a "perverse" kind of prediction (1967/2004, p. 226). He says that we can test a compositional theory by showing it to people and asking whether they think it is correct. But, asking people whether they like a theory is not a prediction in the true sense of the word. Otherwise, we could conduct plenty of experiments to confirm the Creation hypothesis in the Southern United States. Further, if our intuitions were so clear about semantic representation, it is a wonder that there is a need to still do semantics at all, or that there have ever been mistakes or debates.

Thus, I conclude that a compositional theory alone makes no predictions, while mine does. Further, I think if anyone tries to make a theory that predicts the same thing by appeal to meanings,

we will have a theory from which the word meaning can be eliminated, a conclusion paralleling that made by Davidson in 1967.

4.1.2 Some Notes on the Simple Model

One very important conceptual note to make, here, is that on the model of comprehension that I have just presented, there is not a distinction between understanding a sentence, on the one hand, and drawing the conclusions that follow from it, on the other. That is, *it is not as though one can first understand a sentence, then compute its conclusions—these are the same process*.

In another note, the reader might well be wondering about the relationship between the model of knowledge and comprehension just proposed, on the one hand, and the system of *meaning postulates* employed by Carnap (1956) and Montague (1974). Meaning postulates, essentially, are universally quantified sentences that "place constraints on admissible models," and, in particular, "on the notion of 'admissible model for English"' (Chierchia and McConnell-Ginnet 2000, p. 449).

To illustrate, suppose we consider (82) to be a "meaning postulate":

(82) $\forall x\ man(x) \rightarrow \big(mortal(x) \wedge two\text{-}legged(x) \wedge animal(x)\big)$

So, then any model in which some individual is a man but is not $mortal$ would be an "inadmissible model" for a speaker of English. The notion has, obviously, a range of intuitive interpretations.

The question we then want to ask is, what is the difference between a set of sentences, such as **K**, representing an individual speakers *knowledge*, on the one hand, and a set of sentences, say MP, called *meaning postulates*, on the other? It is important to first note that both objects are sets of sentences. They are not different kinds of objects, the only difference between these notions is in how they have been proposed to fit into the broader picture of language use.

One difference is that what I have proposed is a model of *comprehension*, which is a process. A knowledge set, **K**, is a set of sentences used to draw inferences. I have specified a particular operation that happens (i.e. the computation of inference) at a particular time (i.e. after a sentence has been parsed), and **K** is an input to this process.

In contrast, a set of meaning postulates, MP, does not necessarily participate in any computational process. That is, it does not seem as though anyone is suggesting that a hearer actually *computes* the set of "admissible models" at any time, which would seem to be both impossible as well as pointless. That is, it would seem to be pointless to enumerate such facts as that there is an admissible model for English in which Timmy is a mortal man, and that there is one where Ulrich is a mortal man, and that there is one where Victor is a mortal man, etc. In general, no one, to

my knowledge, has proposed what sorts of computations meaning postulates might be used for, nor when, nor to what end.

A second important difference between a knowledge set and a set of meaning postulates is that the former is individualistic, while the latter would seem to not be. That is, two speakers might have different knowledges[5], whereas it would seem that there would be one set of meaning postulates for all speakers of English. In Chomsky's (1986) terms then, a knowledge set is an *I-Language* notion (i.e. internalist and relative to the individual), while a set of meaning postulates is an *E-Language* notion (i.e. externalist and independent of any individual).

These are the two principal differences between knowledge sets and meaning postulates and the name "knowledge set" has been chosen to highlight this. That is, the name "knowledge set" seems to connote that it denotes something internal to a given speaker, which is drawn upon in by particular cognitive processes[6].

One last important note to be made with respect to the model introduced in this section, as alluded to in ff. 3, is that it is an obvious but still formidable criticism of this model that I have modeled the speaker by supposing that he can compute all of the inferences that follow from a given set of sentences. This is, of course, not possible, and probably computationally intractable to the extent that it is. In fact, evidence that we do not draw all possible inferences is, obviously, not hard to find. For example, while the reader probably knows the rules of arithmetic, they have probably never computed $(123 * 88) + ((12 - 5) * 90)$.

That is, even though this is a conclusion that would have followed from your knowledge upon your being taught the rules of arithmetic, the answer is probably useless to you, and so you have left it uncomputed along with the vast majority of conclusions that you might have but have not actually drawn, generally.

One might choose to say that, instead of computing *all* inferences following from their knowledge, the hearer computes the first n, or else computes as many as they can in time t. In that case, one runs into the question of whether or not they will compute the inferences that our theory needs them to compute in the allotted time. This is especially important when inference is used, as it will be in the next section, to recover the speaker's full propositional form—i.e. the speaker's "what was said." This is why Asher and Lascarides (2003) use a limited version of predicate calculus, for

[5]This fact raises questions about whether or not interpersonal communication can succeed, which are introduced and addressed in §5.

[6]Cf., of course, the fact that Chomsky (1986), in advocating a focus on I-Language, also chose to use the name "knowledge of language" for what he saw as the principal topic of linguistic inquiry.

which they can guarantee that all inferences that will be made will be made in finite time.

I prefer not to adopt this course but leave the matter open for future research. (Cf., in this regard, the discussion of possibilities for future research in §6.2.) I would note that the same problem comes up in the computer scientific field of *belief revision* (Alchourrón, Gärdenfors and Makinson 1985), which studies which beliefs rational agents will drop when they are given inconsistent information. The attitude in that case seems to be that the ultimate goal is indeed to incorporate computational feasibility, but that useful work can be done initially by ignoring computational constraints. Similarly, in this case, the idealization by which we ignore computational constraints in the short-term does not imply that the goal is not to incorporate them in a way that allows us to make meaningful predictions in the long-term.

I think that the important question to ask is, while we all recognize that this idealization is not realistic, is the discrepancy with reality such that a mature model would not use logic at all, and perhaps use some completely different system of knowledge representation, like semantic markers (Katz and Postal 1964), so that our work with the idealizing assumptions becomes completely useless? I think not. I think that, when the theory is at an advanced enough level that computational concerns become unavoidable, we are going to be able to get computationally plausible models from the idealized ones by simply restricting them.

Furthermore, Lenat claims that many useful inference patterns are only a few inferences long, and so the inferences needed to answer many questions can actually be carried out fairly quickly (Lenat 1997). But, the data that he has on this are not public information.

4.2 A More Complex Model

4.2.1 Asher and Lascarides' Realization of Discourse Relations

Recall that in §3.4, I said that one aspect of the explicature problem had been addressed in some detail by Asher and Lascarides (2003). This aspect is the resolution of the potentially unmarked discourse relations in a text between clausal elements. I want to briefly consider this solution because the one that I will employ for semantic enrichment generally will be highly similar.

Consider the following discourse, adapted from Asher and Lascarides:

(83) a. Max had a lovely evening last night. π_1

b. He went to the new Italian restaurant. π_2

c. He had lasagna. π_3

Now, Asher and Lascarides posit an $Elaboration$ relation. Let us say that $Elaboration(\beta, \alpha)$ holds between two sentences α and β if β "elaborates" α[7]. Now, one crucial fact about elaboration is that:

(84) Transitivity of elaboration:
If $Elaboration(\gamma, \beta)$ and $Elaboration(\beta, \alpha)$, then $Elaboration(\gamma, \alpha)$.

In other words, if γ elaborates β and β elaborates α then γ elaborates α. To keep matters simple, $Elaboration$ is the only discourse relation that we are going to consider here.

Now, Asher and Lascarides' theory would assign the following discourse structure to the discourse in (83):

(85) a. $Elaboration(\pi_2, \pi_1)$
b. $Elaboration(\pi_3, \pi_2)$

That is, π_2 elaborates π_1, and π_3 elaborates π_2. In other words, being told that Max went to the Italian restaurant elaborates on his having a nice evening. And, being told that he ate lasagna elaborates on his trip to the Italian restaurant, which, by (84), in turn elaborates on Max's nice evening.

Now, Asher and Lascarides are rather ambiguous about how this is supposed to be empirical. They consider their theory successful to the extent that the theory can pick out the discourse structure that matches their intuitions about what the right discourse structure is. But, the theory itself is a formalization of their intuitions. But, most work in semantics is essentially like this, and their solution is non-trivial, so let us consider how it works.

The question is, how is the structure (85) picked out for the discourse (83)? Well, Asher and Lascarides do not give a deterministic algorithm for finding the correct discourse structure. What they give instead is a comparative relation that says whether one discourse structure is better than another. Dropping some detail, let us call this relation $\geq_{\text{disc}}$, where $d \geq_{\text{disc}} d'$ is read as saying that "d is a more coherent discourse structure than d'."

Then, in the fashion of Optimality Theory (Prince and Smolensky 1993 [2004]), they say, let us speak as though *all* discourse structures can be generated, and then we will pick as the "winner," the discourse structure d that is $\geq_{\text{disc}}$-maximal. That is, we are going to pick the discourse structure d such that $d \geq_{\text{disc}} d'$ for all other possible discourse structures d'. That is, we are going to pick the "most coherent discourse structure" out of all possible, according to the relation $\geq_{\text{disc}}$.

[7] They give a more substantive definition but we can deal on an intuitive level.

So, we are considering the discourse about Max, (83). Now, we are only considering one discourse relation, i.e. $Elaboration$, and there are only three sentences, so there are not many candidate discourse structures to compare. Let us compare (85) with the other most plausible candidate, (86):

(86) a. $Elaboration(\pi_2, \pi_1)$

b. $Elaboration(\pi_3, \pi_1)$

Note that the difference is that (86) has π_3 elaborating on π_1, instead of elaborating π_2. That is, Max's eating the lasagna elaborates his good night, and not his trip to the Italian restaurant.

Well, let us say that part of the definition of $\geq_{\text{disc}}$ is that $d \geq_{\text{disc}} d'$ if, all other things being equal, d contains more discourse relations than d'. Now, (86) has two discourse relations; one between π_2 and π_1 and one between π_3 and π_1. But, (85) actually has *three* discourse relations, because π_2 elaborates on π_1, π_3 elaborates on π_1, and, by (84), there is also an elaboration relation between π_3 and π_1.

Thus, (85) $\geq_{\text{disc}}$ (86), and this is chosen as the discourse structure for the text (83). Note that, like a typical Optimality Theory solution, we do not give an algorithm that actually tells us which candidates are going to be compared, or in what order. As the size of the discourse and the number of allowable discourse relations grows, the number of candidates grows very (exponentially) fast.

Just as in the last section, the computational concerns are recognized. The attitude seems to be, again, that, while computational reality is indeed important, it is perhaps most profitable to ignore computational constraints in these early days of theory making.

4.2.2 Resolving Explicature-Implicature

While Asher and Lascarides are concerned primarily with discovering discourse relations, I would like to use their method to resolve the speaker's explicature-implicature pair from the underspecified acoustic signal. This is, as far as I am aware, the first formal general model of how an explicature-implicature pair is resolved on the basis of an underspecified intermediate representation. (Asher and Lascarides model the resolution of presuppositions, which one might like to call a kind of implicature, but which is a less general phenomenon.)

Consider the following Gricean discussion:

(87) a. i. Annie: Jones and Smith are hilarious.

ii. Bob: Yeah. Hey, does Jones have a girlfriend?

b. i. Annie: Well, he$_?$ *is* visiting New York a lot.

Here, our goal is going to be to model Bob's resolution, as the hearer, of the discourse referent he$_?$ that he should use to understand Annie's statement (87b-i), along with any implicature that might serve to make his choice of discourse referent more relevant. I have included the statement (87a-i) in this dialogue only so that more than one discourse referent is introduced—otherwise, resolving he$_?$ would be trivial—and it is of no further interest.

Let us make the idealizing assumption that there is no trouble in resolving the discourse referents for proper names. Let d_{Jones}, and d_{Smith}, be the discourse referents for "Jones," and "Smith," respectively.

We assume Bob parses (87) as (88). Consider (88) to be the transcription of a *dialogue*, **D**. A dialogue is a sequence of two element sequences[8].

(88) a. i. $hilarious(d_{\mathrm{Jones}}) \wedge hilarious(d_{\mathrm{Smith}})$

ii. $question(\lceil \exists x\,(girlfriend\text{-}of(d_{\mathrm{Jones}}, x))\rceil)$

b. i. $visits\text{-}a\text{-}lot(?_{\mathrm{he}}, d_{\mathrm{NYC}})$

What we can ask of a dialogue is, "What is Annie's first statement?," "What is Bob's first statement?," "What is Annie's second statement?," etc. Annie's sentences are the (i) examples and Bob's are the (ii) examples. (The notation is unfortunate.)

So, our focus is on what Bob does during Annie's second statement. Let us suppose that Bob is maintaining a set DR of the discourse referents that are coming up in the dialogue (cf. Karttunen 1968, 1976, Heim 1982, Kamp 1981). DR contains, before the beginning of any statement, each of the discourse referents used in all "prior" statements in the dialogue[9]. Now, the problem before us is to choose some $d \in DR$ and let $?_{\mathrm{he}} = d$.

We have seen that we are maintaining a set of discourse referents. We will also keep track of the *salient questions*. The salient questions are the sentences of the form $question(s)$ that have occurred in prior statements in the dialogue **D**. We will add each salient question to the context, **C**.

The $\lceil\rceil$ brackets in (88a-ii) indicate an intension. I am using an intensional first-order predicate logic. I will not go into all of the details but $\lceil \exists x\,(girlfriend\text{-}of(d_{\mathrm{Jones}}, x))\rceil$ is the structural-descriptive *name* of the sentence $\exists x\,(girlfriend\text{-}of(d_{\mathrm{Jones}}, x))$.

[8] A sequence is something for which one can ask how many items there are and what the i'th item is.

[9] Annie's i'th statement is "prior" to Bob's i'th statement. Bob's i'th statement is "prior" to Annie's $i+1$'th statement.

I cannot actually cite a reference to such a logic, but if the reader might wonder whether such a logic could exist, they may note that if we were to simply pretend that the intensional brackets are not there, I would be using Gilmore's (2005) intensional type theory, which he has proven to be sound.

So, by the time of Annie's second statement, we are in the following state:

(89) $DR = \{d_{\text{Jones}}, d_{\text{Smith}}\}$

(90) $\mathbf{C} = \{question(\lceil \exists x\,(girlfriend\text{-}of(d_{\text{Jones}}, x))\rceil)\}$

Note that Bob's salient question is represented in **C**.

Now, let us assume that Bob has the following knowledge:

(91) $$\mathbf{K} \supset \left\{ \forall x, y, z \left[\begin{array}{c} \left[girlfriend\text{-}of(x, y) \wedge lives\text{-}in(y, z)\right] \rightarrow \\ visits\text{-}a\text{-}lot(x, z) \end{array} \right] \right\}$$

So, one thing that Bob knows is that people with girlfriends in cities visit those cities a lot (this is a monotonic inference, unfortunately[10]).

With this set up, let us digress to devise some general framework for the resolution of explicature-implicature. We assume that we are modeling a system, which, following §3, we might as well call the *pragmatic system*, for we are now modeling the "pragmatic processes" that I referred to in that section.

Then, we will say that the *parser* takes as input a string of symbols, U, and, according to rules which we can, for our purposes, assume are not context-sensitive, creates and passes to the pragmatic system an *underspecified logical form*, u. The underspecified form is in a language, $\mathcal{L}_{ulf}$. Each sentence in $\mathcal{L}_{ulf}$ is "satisfied" by a set of sentences in the language of the full propositional forms, LL. As discussed in §3.4, any underspecification in the input to the parser is not resolved, but passed on as underspecification in u.

Note that (88b-i), repeated here,

$$visits\text{-}a\text{-}lot(?_{\text{he}}, d_{\text{NYC}})$$

would be considered an underspecified logical form, because it specifies a range of full propositional

[10] A "monotonic" inference contrasts with a "nonmonotonic" inference. The latter would allow us to say that guys with girlfriends in cities will *probably*, rather than certainly, visit them. I would prefer to encode this inference as being nonmonotonic, but this would have required a more complicated logical language. I am not exactly sure how to mix intensions with nonmonotonicity, and the solution is complicated enough as it is. But, the overall idea should be clear in any case.

forms but is not one itself. That is, the full propositional forms that satisfy u = $visits\text{-}a\text{-}lot(?_{\rm he}, d_{\rm NYC})$ are those gotten by replacing $?_{\rm he}$ with a discourse referent, such as $d_{\rm Jones}$ or $d_{\rm Smith}$.

In this case, the only piece of underspecification is an indexical, so we can write the underspecified form in a language with the same syntax as the fully specified form, even if this is not the case in general (cf. §3.4). That is, $\mathcal{L}_{ulf}$ and LL have essentially the same syntax, except that $\mathcal{L}_{ulf}$ statements can include constants of the form $?_x$, while LL sentences cannot.

We are going to compute an explicature-implicature pair for u as a function of three items: i) u, the underspecified form; ii) **K**, Bob's durable word-/world-knowledge; and iii) **C**, the context. The *explicature* for u, will be what the hearer hypothesizes to be the speaker's intended full propositional form, and will be denoted $\mathfrak{l}_u$. The *implicature* associated with $\mathfrak{l}_u$will be called $\mathfrak{i}_u$, and will constitute a set of hypotheses which, if assumed, make $\mathfrak{l}_u$ "more relevant" as an explicature.

In a similar fashion to Asher and Lascarides, we are going to define a comparative relation, $\geq_{\mathbf{K},\mathbf{C}}$. We can informally read $\langle l, i\rangle \geq_{\mathbf{K},\mathbf{C}} \langle l', i'\rangle$ as saying that the sentence l and the set of hypotheses i are a "more relevant" explicature-implicature given **K** and **C**, than l' and i'.

We will then pick out, in Optimality Theory-style, the $\langle l, i\rangle$ that is $\geq_{\mathbf{K},\mathbf{C}}$-maximal (i.e. such that $\langle l, i\rangle \geq_{\mathbf{K},\mathbf{C}} \langle l', i'\rangle$ for all $\langle l', i'\rangle$) and choose this as the explicature-implicature pair for u. That is, we will set $\langle \mathfrak{l}_u, \mathfrak{i}_u\rangle$ = $\langle l, i\rangle$. The possible candidates, for $\mathfrak{l}_u$ will, in general, consist of the set of all statements in LL that "satisfy" u. In this case, this means those that are exactly like u except that $?_{\rm he}$ is replaced with an discourse referent, as discussed. A candidate for $\mathfrak{i}_u$ will be any set of well-formed sentences in LL.

The reader may be noticing that I keep deferring a hard definition of the notion "more relevant." Well, I am now going to define the symbol $\geq_{\mathbf{K},\mathbf{C}}$ in a formal, if only partial, way[11]. I will give a technical definition in (92), and then elaborate immediately after:

(92) *Indirect Question Answering Principle*:
Call α the answer to a question in **C** if either $question(\lceil\alpha\rceil) \in \mathbf{C}$ or $question(\lceil\neg\alpha\rceil) \in \mathbf{C}$.
Say that the pair $\langle l_0, i_0\rangle$ answers a question in **C** through explicature if l_0 is the answer to a question in **C**.
Say that the pair $\langle l_0, i_0\rangle$ answers a question indirectly through implicature if i_0 contains an answer to a question in **C** and $l_0 \in Cn(i_0 \cup \mathbf{K} \cup \mathbf{C})$ but $l_0 \notin Cn(\mathbf{K} \cup \mathbf{C})$.[12]

[11] This is a "partial" definition in the sense that I will only give one "principle" of relevance, whereas a more advanced model would have many. And, an even more advanced model might replace the concept of principles altogether, cf. the discussion at the end of this section.

> Then $\langle l, i\rangle \geq_{\mathbf{K},\mathbf{C}} \langle l', i'\rangle$ if, other things being equal[13], $\langle l, i\rangle$ answers a superset of the questions that $\langle l', i'\rangle$ answers.

Now, the idea that some sentence α answers a question is rather intuitive. And, so is the idea that the explicature answers a question. The part that might require elaboration is the indirect answering of a question through implicature.

To put this idea in concrete terms, consider the case of Bob trying to figure out who $?_{\mathrm{he}}$ is in (88b-i). Suppose he takes Annie to be saying that it is Jones who has been visiting New York. Now, if Bob were to *hypothesize* that Jones has a girlfriend in New York, then that would predict, or cause him to infer in conjunction with **K**, that Jones visits New York. That is, it would predict the explicature. So, this is an indirect answer to a question through implicature because, by assuming an answer to the question, in conjunction with his knowledge, Bob infers the explicature.

Then, an explicature-implicature pair, $\langle l, i\rangle$, is more relevant than another pair, $\langle l', i'\rangle$, if $\langle l, i\rangle$ answers all of the questions that that $\langle l', i'\rangle$ answers and more. Note that, in this case, it is not possible that Annie's explicature answers the question as to whether or not Jones has a girlfriend, because an answer such as $\exists x\ girlfriend\text{-}of(d_{\mathrm{Jones}}, x)$ would not satisfy the constraints on the form of $\mathfrak{l}_u$ imposed by the constraint that it must have the form u = $visits\text{-}a\text{-}lot(?_{\mathrm{he}}, d_{\mathrm{NYC}})$.

Let us start by considering the candidate that I think should be the winner:

(93) $\langle l_w, i_w\rangle$, such that, l_w = $visits\text{-}a\text{-}lot(d_{\mathrm{Jones}}, d_{\mathrm{NYC}})$ and
$i_w = \left\{ \begin{array}{c} girlfriend\text{-}of(d_{\mathrm{Jones}}, d_{\mathrm{Jones'Girl}}), \\ lives\text{-}in(d_{\mathrm{Jones'Girl}}, d_{\mathrm{NYC}}) \end{array} \right\}$

Here, the pronoun $?_{\mathrm{he}}$ is resolved to d_{Jones}, the discourse referent for Jones, in the explicature. The implicature is the hypothesis that Jones has a girlfriend, $d_{\mathrm{Jones'Girl}}$, who lives in New York. Note, á la Heim, that $girlfriend\text{-}of(d_{\mathrm{Jones}}, d_{\mathrm{Jones'Girl}})$ implies $\exists x\ girlfriend\text{-}of(d_{\mathrm{Jones}}, x)$. Now, the former does not actually fit the template of answer to a question about the latter as defined in (92). But, let us just fudge the matter and pretend it is, to save adding technical detail that is otherwise unnecessary.

So, the pair $\langle l_w, i_w\rangle$ of (93) answers $question(\lceil \exists x\,(girlfriend\text{-}of(d_{\mathrm{Jones}}, x))\rceil)$ through implicature. Thus, $\langle l_w, i_w\rangle$ is a fairly relevant explicature-implicature pair. Now, let us compare it to some other candidates, to see whether any others are as relevant (where "relevance" is measured by

[12] We might also define a notion of answering a question directly through implicature, but it would only lead to clutter that we will not have use of.

[13] Let us not go through the trouble of trying to formulate a definition of "other things being equal."

[92]).

It is always difficult to select a good set of candidates in testing an Optimality Theory solution. The situation is especially acute in this case since candidates for the implicature set, $\mathfrak{i}_u$, could have any size and could contain anything at all. I think that the methodological matter of how theories like this are to be tested will be something of a topic of discussion in and of itself as theories like this become popular. Suffice it to say, for now, that I will admittedly only be picking out the candidates that strike me as being pertinent.

Let us first consider a competitor in which $?_{\mathrm{he}}$ is resolved as the other possible discourse referent, i.e. d_{Smith}, the other element of DR:

(94) $\langle l, i\rangle$, such that l = $visits\text{-}a\text{-}lot(d_{\mathrm{Smith}}, d_{\mathrm{NYC}})$ and $i = \varnothing$

Clearly, in (94), the explicature does not answer the question about whether Jones has a girlfriend. Also, there is nothing in the implicature, so this pair also does not answer the salient question through implicature. Thus, this $\langle l, i\rangle$ is less relevant than the currently leading candidate, $\langle l_w, l_i\rangle$, as it answers a strict subset of the questions that $\langle l_w, i_w\rangle$ does (in fact it answers no questions).

Let us continue looking at competitors that resolve $?_{\mathrm{he}}$ to d_{Smith}—incorrectly, according to our stated intuitions—which contain material in their implicature that might make this an answer to the question. For example:

(95) $\langle l, i\rangle$ such that l = $visits\text{-}a\text{-}lot(d_{\mathrm{Smith}}, d_{\mathrm{NYC}})$ and
$i = \{girlfriend\text{-}of(d_{\mathrm{Jones}}, d_{\mathrm{Jones'Girl}})\}$

(95) has the pronoun resolved to Smith with the totally unrelated hypothesis that Jones has a girlfriend as the implicature. Obviously, we need to make sure an example like this is ruled out because, intuitively, the implicature has nothing to do with the explicature.

In fact, (95) is ruled out. Recall the following piece of (92):

(96) Say that the pair $\langle l_0, i_0\rangle$ answers a question through implicature if i_0 contains an answer to a question in $\mathbf{C}$ and $l_0 \in Cn(i_0 \cup \mathbf{K} \cup \mathbf{C})$ but $l_0 \notin Cn(\mathbf{K} \cup \mathbf{C})$.

Here, l = $visits\text{-}a\text{-}lot(d_{\mathrm{Smith}}, d_{\mathrm{NYC}})$ is *not* in $Cn(i \cup \mathbf{K} \cup \mathbf{C})$ and so this does not constitute an answer through implicature.

Next, consider the following suspicious solution:

(97) $\langle l, i\rangle$ such that l = $visits\text{-}a\text{-}lot(d_{\mathrm{Smith}}, d_{\mathrm{NYC}})$ and
$i = \left\{ \begin{array}{c} girlfriend\text{-}of(d_{\mathrm{Jones}}, d_{\mathrm{Jones'Girl}}), \\ girlfriend\text{-}of(d_{\mathrm{Jones}}, d_{\mathrm{Jones'Girl}}) \rightarrow visits\text{-}a\text{-}lot(d_{\mathrm{Smith}}, d_{\mathrm{NYC}}) \end{array} \right\}$

This solution involves Bob positing, with no real reason to do so, that if Jones has a girlfriend, then Smith will visit New York a lot. Here, the explicature $visits\text{-}a\text{-}lot(d_{\mathrm{Smith}}, d_{\mathrm{NYC}})$ *is* in the set $Cn(i \cup \mathbf{K} \cup \mathbf{C})$, but is not in $Cn(\mathbf{K} \cup \mathbf{C})$. Thus, (97) constitutes an answer to the salient question through implicature.

The apparent problem here is that there is nothing in Bob's knowledge that would make this a natural hypothesis. There is no particular reason to think that if Jones has a girlfriend, Smith will visit New York. Clearly, if this sort of inference is allowed to stand, it would be possible to see any explicature as the answer to any question.

My first thought was to prevent this kind of example by saying that, in order to constitute an answer to a question through implicature, we would require that $l_0 \in Cn(i_0 \cup \mathbf{K} \cup \mathbf{C})$ but that $l_0 \notin Cn(i_0)$. That is, the link between the explicature and the answer to the question must involve at least *some* of the hearer's antecedent world-knowledge.

But, it would be possible for an example to circumvent even this new rule. To see why let,

$$\mathbf{\Psi} = \forall x, y, z \left[\begin{array}{c} \left[girlfriend\text{-}of(x, y) \wedge lives\text{-}in(y, z) \right] \rightarrow \\ visits\text{-}a\text{-}lot(x, z) \end{array} \right]$$

Recall that (91) stated that $\mathbf{K} \supset \{\mathbf{\Psi}\}$. That is, $\mathbf{\Psi} \in \mathbf{K}$. That is, $\mathbf{\Psi}$ is in Bob's word-/world-knowledge. Now, consider the following silly implicature-explicature combination:

(98) $\langle l, i \rangle$ such that $l = visits\text{-}a\text{-}lot(d_{\mathrm{Smith}}, d_{\mathrm{NYC}})$ and

$$i = \left\{ \begin{array}{c} girlfriend\text{-}of(d_{\mathrm{Jones}}, d_{\mathrm{Jones'Girl}}), \\ \left[\begin{array}{c} \left(girlfriend\text{-}of(d_{\mathrm{Jones}}, d_{\mathrm{Jones'Girl}}) \wedge \mathbf{\Psi} \right) \\ \rightarrow visits\text{-}a\text{-}lot(d_{\mathrm{Smith}}, d_{\mathrm{NYC}}) \end{array} \right] \end{array} \right\}$$

Note that, now, $l \notin Cn(i)$ because one needs to assume $\mathbf{\Psi}$ in order to conclude l. And, $\mathbf{\Psi}$ is not in i but it is in $\mathbf{K}$. Thus, we will have $l \in Cn(i \cup \mathbf{K} \cup \mathbf{C})$, but neither $l \in Cn(\mathbf{K} \cup \mathbf{C})$ nor $l \in Cn(i)$. So, this explicature-implicature pair answers the question as to whether or not Jones has a girlfriend. But, this is the most ridiculous explicature-implicature pair we have seen yet.

Clearly, what we need is another principle characterizing $\geq_{\mathbf{K},\mathbf{C}}$ that stipulates that, other things equal, a more "complicated" or "unnatural" implicature is less relevant than a less complicated one. Suggesting and testing various ideas would constitute a thesis in and of itself. But, we can easily imagine avenues along which to go.

For example, one rule which would work here is to suppose that a pair $\langle l, i \rangle$ is, other things equal, more relevant than the pair $\langle l', i' \rangle$ if i' contains statements that have implication (i.e. contain

the symbol $\rightarrow$) while i does not. This would rule out both (97) and (98), while retaining our desired winner (93).

I would note, in connection with this point, that Elio and Pelletier (1994) report on a study inspired by the belief revision literature (cf., Alchourròn et al. 1985) in which subjects are presented with a generalization about the world, and some statements about individuals, such as the following:

(99) a. When it rains, everyone stays inside.

b. It is raining right now.

c. There are people outside.

(Note that (99a) is, for our purposes, a generalization over times. In other words, it essentially says that, for all times, t, such that it is raining at t, everyone will be inside at t. Meanwhile, (99b–99c) are both statements about an individual time, say t_{now}.)

They report that people will more often give up the generalization than the facts about individuals. This parallels the aforementioned speculation that we would prefer an implicature that makes statements about individuals than one that makes generalizations. So, one might speculate on a relationship between the kinds of statements that people prefer to hold to, and the kinds of statements that people will be willing to attribute to a speaker's implicature. I will not pursue this idea, but merely leave it as an interesting speculation.

Returning to the broader themes, note that our derivation of the explicature-implicature pair, $\langle \mathfrak{l}_u, \mathfrak{i}_u \rangle$, for u essentially happens, just as Sperber and Wilson (1998) suggested, in a process of "mutual parallel adjustment." That is, both elements of the pair are chosen as part of a single process. Also, note that the notion $\langle \mathfrak{l}_u, \mathfrak{i}_u \rangle$ would seem to be the ultimate formal characterization of the distinction between the "what is said" and the "what is meant" that Grice made over forty years ago.

In terms of the relationship between the model of this section and that in §4.1, one can view matters in either of two ways. First, one might suppose that, once the speaker's explicature, $\mathfrak{l}_u$, is resolved, the hearer then computes $I_{\mathbf{K}}(\mathfrak{l}_u)$. Or, one might suppose that $I_{\mathbf{K}}(\mathfrak{l}_u)$ is computed as part of the process of resolving the explicature. The model is too rudimentary, at this point, to bother distinguishing, I think.

Also, note that, in parsing (87) as (88), I ignored the fact that there was focus on the auxiliary *is*. To me, the dialogue sounds much more natural if the *is* is emphasised, so I included this emphasis. I usually take emphasis on this kind of *is* to indicate insinuation. Perhaps the reader will agree. This is interesting because we did, indeed, predict that this statement involved an implicature. An advanced

model of explicature resolution might be able to incorporate this kind of emphasis to prejudice the choice of an explicature-implicature pair that involves an implied answer to a question.

Also, recall the discussion of §3.1.3, in which I discussed the fact that Wilson and Sperber (1986) had sought to reduce all of Grice's maxims to essentially, what I called, one "super" maxim, that of relevance. They call one explicature-implicature pair better than another if it is more "relevant." But, I noted that, a single general principle has no predictive validity, i.e. it cannot actually pick out the explicature in any concrete case. For this, one needs more specific axioms.

Here, I have kind of created an analog of this super maxim: $\geq_{\mathbf{K},\mathbf{C}}$. That is, one explicature-implicature pair is better than another if it is $\geq_{\mathbf{K},\mathbf{C}}$ than the other. Thus, the general framework for picking an explicature-implicature—i.e. that we pick the $\geq_{\mathbf{K},\mathbf{C}}$-maximal pair—can be stated without any need to know exactly how $\geq_{\mathbf{K},\mathbf{C}}$ is being defined. But, in order to actually pick a winner in a concrete case, $\geq_{\mathbf{K},\mathbf{C}}$ needs some further characterization, such as that given in the *Indirect Question Answering Principle* (92). Obviously, (92) is not, in and of itself, a complete definition of relevance. There will need to be more principles added.

As the number of sub-principles of $\geq_{\mathbf{K},\mathbf{C}}$ grows, it is may start to contain particularized, context-rigid, means of choosing between candidates. The solution to this will be to push our pragmatic theory, which we are now stating in our meta-language, into the language of the hearer's thoughts, so that they are seen to be his knowledge. Then, we will have the hearer try to solve the problem:

(100) By u, the speaker probably meant $\langle l_0, i_0 \rangle$.

Such a solution would have largely the same flavor. In this case, we would still use an Optimality Theory-style solution, with the ordering relation ordering pairs $\langle l, i \rangle$ according to how strongly the hearer believes (100), when $\langle l, i \rangle$ is substituted for $\langle l_0, i_0 \rangle$.

4.3 Conclusion

In §2 and §3 we brought out two short-comings with using a traditional, compositional semantic theory as a theory of comprehension. The first was that differences in word-meaning between words within the same grammatical class are not accounted for. The second was that the acoustic signal was semantically underspecified. The speaker's intended full propositional form, or explicature, had to be hypothesized by the hearer in a parallel process to determining what the speaker meant to insinuate.

Correspondingly, in this section, I have sketched a solution to each problem. In §4.1, I showed

how word-meanings could be modeled assuming that the mapping from natural language form to logical language form was deterministic.

In §4.2, we saw how an underspecified logical form can be enriched to yield the speaker's full propositional form, at the same time as the speaker's implicature was guessed at, just as Sperber and Wilson had suggested.

Chapter 5

As a Holistic Model of Knowledge and "Meaning"

The model of comprehension that I presented in §4, and especially §4.1, would be referred to in a large philosophy literature as a "holistic" theory of "meaning." Though this literature may not be as well-known among linguists, there are well-known problems facing holistic theories, which await such a model of comprehension as I have given in the last section.

In §5.1, I would like to introduce the concept of a holistic theory and explain what a "holistic theory of meaning" is. We will also see what kinds of challenges a holistic theory of meaning faces. And, I will also explain why the model of §4.1 constitutes a holistic theory.

In §5.2, I will review Quine's (1951) "Two Dogmas of Empiricism," in which he militates against any kind of distinction between "analytic" and "synthetic" statements, i.e. a distinction between statements that hold in virtue of what a word "means" versus those that hold contingently. This will be necessary for, later in the chapter, I will want to in fact adopt some notion of a conventionalized distinction between statements that hold by virtue of "meaning" and those that hold contingently.

In §5.3, I will explain what benefits a holistic theory affords. That is, we have already said that our model of comprehension in §4.1 is a holistic one. So, all of the arguments given in favor of that model are arguments in favor of holism. But, we will see that there are some important others. In particular, a holistic semantic theory easily affords an explanation as to what innate basis is required for the ultimate acquisition of mature verbal concepts, as well as an explanation of how this innate basis is transformed into an adult competence.

In §5.4, I will show how to address the legitimate and most important criticisms of meaning holism that will have been reviewed in §5.1. This will involve an appeal to a notion in of conventionalized word-meanings, which will make use of the careful treatment of Quine given in §5.2.

In §5.5, after having addressed some legitimate criticisms of holism, I will also discuss some due to Fodor and Lepore (1992), which have become somewhat popular.

5.1 Introduction

§5.1.1 explains what a holistic theory of meaning *is*. §5.1.2 overviews some of the problems that confront a holistic theory of meaning, or anything resembling one. §5.1.3 explains why the model of comprehension given in §4.1 would be called holistic.

5.1.1 What is "Meaning Holism"?

I am now going to be discussing works mainly from the philosophical literature. So, I am going to start using "meaning" as a quasi-technical term, because the authors in question use it this way and I must do so as well to evaluate their proposals. I would like to recall, in connection with §1.2.3, that my concern is ultimately not to create a theory of "meaning," but rather to explore what this discussion of "meaning" implies for my theory of comprehension.

It should also be noted that each author that will be in question here either: i) does not have an actual, rigorous definition of "meaning" in mind, or else ii) a definition of meaning is precisely their theoretical contribution, which another researcher will then attack. Quine (1951) and Fodor and Lepore (1992) are of the first sort. They will not tell you once and for all what "meaning" is, and what logical properties it has. But, they are often ready to tell you *some* properties that it has. Block (1986) comes nearer to being of the second sort. He proposes to equate "meaning" of a word with some sort of mental contents plus the extension of that word in the "world of non-symbols," and then Fodor and Lepore criticize its adequacy.

Ultimately, again, I think the reader should continue, in this section, to consider "meaning" a pretheoretical term, in the sense of §1.2.3.

With these caveats out of the way, let us now try to say what a *holistic* theory of "meaning" is. A holistic theory of "meaning" is one in which the "meaning" of each word in a language is a *function of* other words in that language, or the meanings of other words in that language. In other words, in a holistic theory, the "meaning" of one word cannot be given without mentioning other words in the same language.

The "language" in question can either be an idiolect (a single person's individual, perhaps idiosyncratic language, cf. Chomsky 1986), or a dialect (the language of a group of people), however one wants to construe these ideas.

There are, in theory, all sorts of ways to give a theory of meaning that could be considered holistic under this definition. For example, the work on semantic nets in the 1960's (cf., e.g. Quillian 1968, Brachman 1979) might be considered holistic if one wanted to view things in such terms. But, in practice, among modern theorists in the branch of philosophy under discussion, there is one particular theory, called *inferential role semantics*, that is, in the words of Pelletier, "[s]o influential that it is often taken to *be* semantic holism," (2009, p. 21, emphasis in original). I take the theory's modern proponent among philosophers to be Block (1986), though the approach can be traced back to Harman, Sellars and Field[1]. *Inferential role semantics will be the only kind of semantic theory we are interested in here.*

An inferential role model of language essentially assumes, as we have in §4, that a person's word-/world-knowledge is structured as a set of sentences in a logical language, such as (101):

$$\text{(101)} \quad \mathbf{K} = \left\{ \begin{array}{c} \forall x\, bachelor(x) \leftrightarrow (\neg married(x) \wedge male(x))\,, \\ \forall x\, married(x) \leftrightarrow has\text{-}spouse(x) \end{array} \right\}$$

As discussed on page 60, I will always use the symbol $\mathbf{K}$ to represent a person's word-/world-knowledge, and will often refer to this as the person's *knowledge-set*, or *belief-set*, interchangeably.

Now, in §4.1, we focused on sentences and sets of sentences. That is, a hearer was assumed to turn a natural language sentence, U, into a logical language sentence, $\mathfrak{l}_u$, from which it would compute

$$I_{\mathbf{K}}(\mathfrak{l}_u) = Cn(\mathbf{K} \cup \{\mathfrak{l}_u\})$$

However, when discussing inferential role semantics, it turns out to sometimes be easier to discuss things at the level of words, and ask what is the "inferential role" that a *word*, rather than a sentence, plays in a language (cf. ff. 2).

So, suppose we set out to characterize the role that each of the words (or symbols) like $bachelor$, $married$, etc., is playing in the inferences that follow from a knowledge-representing set like (101).

Let us speak intuitively for a moment about the "role in inference" that the various symbols of (101) are playing. It seems that $bachelor$ should be said to have an inferential "liaison" with $married$, because knowing that $bachelor(c)$ will allow one to conclude that $\neg married(c)$. But,

[1] Cf., Harman 1982, Sellars 1974, Field 1977.

knowing that $\neg married(c)$ allows one to conclude that $\neg has\text{-}spouse(c)$. So, knowing $bachelor(c)$ allows one to conclude that $\neg has\text{-}spouse(c)$. Thus, $bachelor$ should also have some sort of inferential liaison with the predicate $has\text{-}spouse$.

But, also note that we can prove, based on the **K** of (101), that if $has\text{-}spouse(c)$, then $married(c)$, and so $\neg bachelor(c)$. Thus, having information about whether $has\text{-}spouse$ applies to an individual allows one to conclude something about whether $bachelor$ applies to the individual. Thus, just as we have a path connecting $bachelor$ to $has\text{-}spouse$, in that direction, we also have a path going from $has\text{-}spouse$ to $bachelor$.

So, intuitively, we might think of the *inferential role* of a word, w, as the set of all liaisons that w has with other words. While the literature is a bit fuzzy about the definition of inferential role[2], it seems pretty clear how to give a rigorous one, given the framework of §4:

(102) With respect to the knowledge-set **K**, the *inferential role* of the symbol α is the set of statements in $Cn(\mathbf{K})$ which are either of the form (102a) or (102b):

a. $\forall x\ \alpha(x) \rightarrow \alpha^{Y}(x)$

b. $\forall x\ \neg\alpha(x) \rightarrow \alpha^{N}(x)$

(Recall that $Cn(\mathbf{K})$ is the set of all of the consequences that follow from **K** as discussed on page 60.)

So, given the **K** depicted in (101), the inferential role of of $married$ would include:

(103) $$\left\{ \begin{array}{c} \forall x\ married(x) \rightarrow has\text{-}spouse(x) \\ \forall x\ married(x) \rightarrow \neg bachelor(x) \\ \forall x\ \neg married(x) \rightarrow \big[male(x) \rightarrow bachelor(x)\big] \end{array} \right\}$$

Here, $\forall x\ married(x) \rightarrow has\text{-}spouse(x)$ is one of the statements in **K**, and thus it is one of the statements in $Cn(\mathbf{K})$. Also, it is of the form of (102a) and so is part of the inferential role of $married$.

Also, since all $bachelor$s are necessarily not $married$, by *modus tollens*, we can say that if anything *is* $married$, it cannot be a $bachelor$. Thus, we find $\forall x\ married(x) \rightarrow \neg bachelor(x)$, which is the form (102a), in the inferential role of $married$.

[2] Block (1993), who I take to be the theory's main proponent among philosophers, refers to it as "something of a fiction" (p. 2, ff. 2) that we know what it means to define the "inferential role" of a sentence. But, he says, if we assumed that we *could* define the inferential role of a sentence, then "the inferential role of a word as represented by the set of inferential roles of sentences in which it appears," (p. 2).

Lastly, consider $\forall x\ \neg married(x) \rightarrow \big[male(x) \rightarrow bachelor(x)\big]$. This example shows how detailed an inferential role can be. The thing to notice is that, if something is $\neg married$ *and* a $male$, it is a $bachelor$. So, once we know that something is $\neg married$, we only need to learn that it is a $male$ in order to conclude that it is a bachelor. In other words, one of the properties of $\neg married$ things is that they are "$bachelor$s if $male$."

So, basically, the inferential role of $married$ is the set of all inferences that one is prepared to draw about some individual c after either learning that $married(c)$ or $\neg married(c)$. So, what makes an inferential role theory holistic is that the inferential role of $married$ depends on $bachelor$ but, as should be obvious, the inferential role of $bachelor$ depends *in turn* on $married$.

This sort of bidirectionality, in which the "meaning" of a depends on b, while the "meaning" of b in turn depends on a, shows as a marked contrast in its "direction of explanation" when compared to a compositional theory.

In a compositional theory, the "meaning" of a whole, say c, is *only* a function of its parts, say d and e, and their mode of combination. So, the "meaning" of the larger part, c, is dependent on the "meanings" of d and e. But, d and e would each have a "meaning" which can—and in fact must—be stated independently of c's.

For example, in a classic compositional theory, we might have a constant like TOM and a predicate like $bachelor$. Now, TOM has as its extension, $[\![\text{TOM}]\!]$, some individual in a *universe*, and $bachelor$ has as its extension, $[\![bachelor]\!]$, a set of things in that same universe. Then the extension of $bachelor(\text{TOM})$, denoted by $[\![bachelor(\text{TOM})]\!]$, would be $true$ if and only if $[\![\text{TOM}]\!]$ is in the set $[\![bachelor]\!]$.

That is, the extension of $bachelor$ can be stated without reference to TOM, and vice versa, and the extensions of both are, crucially, stated without reference to that of $bachelor(\text{TOM})$. So, if one identifies "meaning" with "extension," then we get that the "meaning" of $bachelor$ can be stated without reference to TOM, and vice versa, and that the "meaning" of the whole depends on the parts, but *not* vice versa.

So, it feels as though we are dealing with two very different modes of approaching language. Indeed, as Pelletier (2009, forthcoming) points out in his survey of the area (in an essay in fact called, "Holism and Compositionality"), there is in practice often felt to be a clash—with fervent anti-holists (i.e., Jerry Fodor) typically being fervent compositionalists. However, given the slack allowed in the definitions of the terms, the two positions are not irreconcilable and the theory proposed below will blend the two kinds of "direction of explanation."

One of the most interesting facts about a holistic theory is that it can be shown, in a variety

of ways based on plausible assumptions, that if a theory is *at all* holistic, then it must be *strongly* holistic. By this I mean, if inferential role is defined as in (102), then the inferential role of each word is (typically) going to mention *all* other words in the language. This will be discussed in §5.1.2. But, first, we must stop to review some ideas of Quine which feature repeatedly in discussions of holism.

5.1.2 Some Legitimate Problems with Holism

I want to now explain why it is that some researchers are concerned about the viability of a holistic theory. The first thing to note in this regard is that it is widely agreed that any kind of semantic holism is going to be an "extreme" kind of holism.

To see what I mean by this, consider that, in general, in a holistic theory of meaning, the "meaning" of each word will "depend on" other words (where the notion of "depend on" can vary between theories). To say that any holistic theory will be an *extremely* holistic theory is to say that, if we allow that the meaning of each word will depend on *some* other words in a language, there is no way to guarantee that the meaning of that word will not depend on *all* other words in the language.

For example, say we fix a language with n words. Then, there is no way to predict, in advance just how many other words will be mentioned in some word's inferential role. That is, there is no way to prove that it will not be all n words. In fact, it probably will be.

To get an intuitive picture of how this happens, suppose that one believes that *tomato* and *stop sign* are both red. Then, one of the inferences about tomatoes is that they are the same color as stop signs. Now, if one of the facts about stop signs is that they are signs that control traffic, then one of the inferences about tomatoes is that anything which is a tomato is the same color as some kind of sign used to control traffic. So, when one learns a new fact about stop signs, such as that they control traffic, not only does the inferential role of stop sign change, but so will the inferential role of tomato.

Further, suppose that *traffic* is believed to be something which *transports* things. Then the inferential role of *tomato* involves the word *traffic*, because a tomato is the same color as something used to control traffic, which means that a tomato is the same color as something which involves the transport of things. Thus, when on learns a new fact about traffic, a concept which ostensibly has little to do with tomatoes, the inferential role of tomato will change.

Of course, there is nothing particular about tomatoes that makes their inferential roles so sensitive to change. Thus, it is generally agreed by proponents and opponents of holism alike that the inferential role for any given word will (or may) depend on all other words in the speaker's idiolect.

The result of this is that if two speakers have idiolects that differ at all, then the inferential role for each word that they know will differ. Similarly, if one person changes any of their beliefs, then the inferential role for each word that they know will also change.

Now, it is quite certain that no two people have the *exact* same beliefs. And, differences in beliefs anywhere, as we have just seen, will lead to differing inferential roles everywhere.

So, one problem that skeptics have with holism is that, if "meaning" is identified with inferential role, and we have some vague idea that when one person understands another sentence, they do it by understanding the "meaning" of that sentence, then if two people have different belief sets, they have different inferential roles for those sentences. So, then the "meaning" that the speaker intended to convey using some sentence is different than the "meaning" that the hearer gets for that sentence. So, says the skeptic, how could it be that communication succeeds?

The second problem is that if, again, "meaning" is identified with inferential role, and we have some, once again, vague idea that when some person accepts or rejects a statement, they accept or reject its "meaning," then when someone accepts some statement, S, at one time, then acquires some new belief which changes all inferential roles, including S's, the "meaning" of S has changed so that, apparently, what was accepted originally was not what is still accepted. Similarly, if one were to later reject S, what was rejected was not what was accepted. Here, the skeptic asks, how can we coherently form a notion of one's changing their mind when changing one's mind about some belief ϕ changes the meaning of ϕ?

5.1.3 Why the Theory of §4 is Holistic

Recall that, in §4.1, we said that, in order to understand a sentence, such as, "Socrates is a man," which translates to LL as $\Psi = man(\text{SOCRATES})$, the hearer would compute $I_{\mathbf{K}}(man(\text{SOCRATES}))$. Now, note that $I_{\mathbf{K}}(\Psi)$ is a function of all of the other words in speaker's (idiolectal) knowledge, $\mathbf{K}$, *as well as* the inferential relationships that hold between these words.

Now, if one loosely identifies $I_{\mathbf{K}}(\Psi)$ with the *meaning* of Ψ, then, clearly, our theory is holistic because the meaning of Ψ cannot be given without discussing other words in the language. Furthermore, even if one does not identify $I_{\mathbf{K}}(\Psi)$ with the meaning of Ψ, it is still the case that one cannot comprehend Ψ without reference to other words in their knowledge, $\mathbf{K}$, along with the inferential relationships between these words.

Ultimately, though, what is of greatest relevance in our discussion here, is the fact that the model of comprehension given in §4.1 is completely relative to the idiolect of the hearer. That is, the hearer

who has knowledge $\mathbf{K}_{\mathrm{H}}$, will understand the sentence Ψ by computing

$$I(\mathbf{K}_{\mathrm{H}}, \Psi) = Cn(\mathbf{K}_{\mathrm{H}} \cup \{\Psi\})$$

But if the speaker has knowledge $\mathbf{K}_{\mathrm{S}}$, and, quite certainly, $\mathbf{K}_{\mathrm{H}} \neq \mathbf{K}_{\mathrm{S}}$, then it would seem that the speaker could not know how it would be that the hearer would interpret their message. If the speaker were to hear Ψ, they would understand it by computing $I(\mathbf{K}_{\mathrm{S}}, \Psi)$ and so, quite certainly,

$$I(\mathbf{K}_{\mathrm{S}}, \Psi) \neq I(\mathbf{K}_{\mathrm{H}}, \Psi)$$

So, although critics of semantic holism have not had a chance to address the model of §4.1, it is pretty obvious that they would have the same concerns about my model of comprehension as they would have about a holistic theory of meaning. Moreover, I am concerned about this problem!

5.2 Quine: Confirmation Holism and the Analytic-Synthetic Distinction

We are going to consider now, in some detail, Quine's (1951) famous paper, "The Two Dogmas of Empiricism," in which he argues in favor of a notion of confirmation holism (the view that scientific theories are tested as whole), and against a notion of an analytic-synthetic distinction (a distinction between statements which hold in virtue of what words mean, versus those that hold contingently and empirically).

Beginning with the issue of confirmation holism, in the early part of the twentieth century, a group of philosophers who called themselves the Vienna Circle, and who were trying to negotiate a philosophy of science, propounded the *empiricist* doctrine, whose tenet of *reductionism* held that every meaningful empirical statement should be reducible to an observation (or a set of observations) that could be made of the world.

An example of a good reductionist theory would be a Skinnerian theory (cf., e.g., Skinner 1935) that specifies that if an animal is presented with a stimulus, S, and is rewarded with another, positive stimulus, S', for the response, R, to S, then the rate at which this animal will respond to S with R will increase. We can observe the animal, S, R, and S', and can specify their sequence. Thus, all statements in this theory can be reduced to observations of the world.

An example of a non-reductionist theory would be a Chomskyan (1957, 1965) theory of transformational grammar, with references to non-terminal nodes, base structures, or transformation rules,

none of which can be observed directly. Chomsky referred to levels of representation that could not be observed directly as "abstract" levels and, of course, it would now be unthinkable to do linguistics without these abstract levels. At the time, Chomsky argued that the existence of his abstract levels was to be evinced by the ability of his theory, *as a whole*, to represent the comprehension of language. Cf.,

> To understand a sentence, then, it is first necessary to reconstruct its analysis on each linguistic level; and *we can test the adequacy* of a given set of linguistic levels *by asking whether or not grammars formulated in terms of these levels enable us to provide a satisfactory notion of "understanding,"* (Chomsky 1957, p. 87, emphasis mine).

Well, this holistic view of how scientific theories should be tested came from the work of Quine under discussion[3] (as well as that by Hempel, and others). Quine argued against the idea that each individual statement in a theory should be able to be reduced to some observable statement. Arguing instead that it is scientific theories as *wholes* that make predictions. And, that when abstract principles in conjunction with observable statements make incorrect predictions, it is likely to be the abstract principles, rather than those that relate the abstract principles to observations, that will be dropped. In his typically cryptic language, Quine expressed this thesis thus:

> The totality of our so-called knowledge or beliefs . . . is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustments in the interior of the field, (1951, p. 31)

In other words, "our statements about the external world face the tribunal of sense experience not individually but only as a corporate body," (Quine, 1951, p. 38). This is the oft-quoted statement of the principle of *confirmation holism*—that theories as wholes, rather than individual statements, are what get tested in science.

One might wonder whether, from this principle of confirmation holism itself, a jump to some sort of "meaning" holism should be made directly. Fodor and Lepore (1992) argue that such a leap is not inevitable, and it is not my desire to argue for one. It is actually the other remarks in this paper—those on the analytic-synthetic distinction—that will be more important in our discussion

[3] In Newmeyer's (1986) opinion, the work by those like Quine and Hempel (1950) to discredit the empiricist philosophy was a precipitating factor in the loss of respect for behaviorism generally and empiricist linguistics in particular. Further, the influence on Chomsky himself is apparent in the acknowledgements to his revolutionary monograph: "In less obvious ways, perhaps, the course of this research has been influenced strongly by the work of Nelson Goodman and W. V. Quine," (Chomsky 1957, p. 6).

of semantic holism. But, this notion of confirmation holism is closely related to the possibility of an analytic-synthetic distinction[4], and is important for understanding what I will call the "spirit" of Quine's (1951) paper, which I will argue differs slightly from some of the comments as they might be interpreted literally.

The concept of the analytic-synthetic distinction can be difficult to characterize because it is stated in different and, I argue, incompatible ways. Quine gives, I count, three definitions. At first, Quine explains that he will militate against a distinction between "truths which are *analytic*, or grounded in meanings independently of matters of fact, and truths which are *synthetic*, or grounded in fact," (1951, p. 20). Second, he says that, "analytic statements are defined as statements whose denials are self-contradictory," (p. 20). Lastly, Quine draws a distinction, "between synthetic statements, which hold contingently on experience, and analytic statements which hold come what may," (p. 40).

Now, Quine, along with many who have discussed his ideas after him, freely switch, without comment between talking of a line between analytic and synthetic *truths*, on the one hand, and *statements*, on the other. To speak of a statement (as in a "sentence") is simple: we can define a well-formed statement by giving a syntactic description that specifies which of all possible strings constitute well-formed statements.

If we are going to speak about *truths*, however, in the sense that somehow these statements are connecting to the "world of non-symbols," does this mean we can only proceed by picking some theory of truth, such as the correspondence or coherence theory? Not necessarily. For, if we are going to speak of "statements whose denials are self-contradictory," then we are essentially speaking about tautologies, and a tautology is true regardless of the model for it. So, we do not need a way to figure out "which model" corresponds to the "real world," because a tautology would be true in each of them[5].

Even with this caveat in the open, we still have an important equivocation on Quine's part. We have still got multiple definitions of a distinction between analytic and synthetic statements. They are not necessarily all equivalent and, I will argue, they are, *in fact*, not all equivalent.

The first two definitions quoted are, I think, synonymous. That is, statements which are true for reasons "grounded in meanings independently of matters of fact," can be viewed as identical to the

[4]Cf., "The dogma of reductionism, even in its attenuated form, is intimately connected with the other dogma: that there is a cleavage between the analytic and the synthetic... The two dogmas are, indeed, at root identical," (Quine 1951, p. 37). Whether and how they might be completely identical is a debate that would lead us astray, however.

[5]Similarly, we can speak of analytic *falsehoods*, which are going to be false in all models.

"statements whose denials are self-contradictory." It is *completely* an intuitive judgment because Quine does not define "meaning," but I feel like equating these.

To me, both of these kinds of analycity are the "synonymy" kind of analycity. That is, "bachelor" is synonymous with (whatever that turns out to mean) "unmarried male." So, "unmarried male" is part of the "meaning" of "bachelor." Also, because the terms are synonymous, one would be contradicting oneself if one denied that bachelors were unmarried males.

But, *prima facie*, it seems to me to be quite a different thing altogether to speak of statements whose denials are self-contradictory, on the one hand, and those "which hold come what may," on the other. And, the reason is this. Let us continue to be sympathetic, for the sake of discussion, to the notion of analycity just discussed. That is, let us say that it is an analytic truth that "the number after 3 is 4," because 4 is *defined* to be, and so synonymous with, "the number after 3." And, so, the denial of this statement would be self-contradictory.

Now, suppose the body of the world's mathematicians were offered ten million dollars each to switch the names of the numbers 4 and 5 around. That is, they are asked to change the names of the numbers so that the (positive) natural numbers would instead be named, $1, 2, 3, 5, 4, 6, 7, \cdots$ (note that 4 and 5 are switched). And, the mathematicians accept the offer. Well, then "the number after 3 is 4" is no longer going to be true by definition. In fact, it will be false by definition, and thus is not a statement that we were willing to hold "come what may." But, it was, originally, a statement whose denial would have been self-contradictory.

What the above *reducio ad absurdum* example, supposing one was willing to follow along, was meant to show was that statements which had once been true by definition may later become false, when it becomes *profitable* to change definitions. But, this example might seem a bit trumped up. No one has the resources, nor the inclination to offer such a bribe to the world's mathematicians, one might say. Moreover, the men and women of mathematics are far too concerned for their craft to bend to the whims of such heretical donors.

But, examples can be found which are closer to the heart. Suppose we had found a number of facts which seemed to hold about "bachelors," where bachelors are "unmarried males": bachelors have messier apartments, they have less food in their fridges, they spend more time at the kind of night clubs that have single women, etc. I will refer to these supposed facts as *hypotheses*. Now, suppose this society comes to have many "common-law" partners, who are not legally married but who have lived together as though they were for many years.

In this case, we find that all of the striking regularities that we found for "bachelors" actually to do not apply to the males living in these common-law relationships for, though they are unmarried,

they share the behavior of married people. Really, the regularities only apply to the group of people who are "unmarried males, who are not in common-law relationships." Now, if our goal is, as for good scientists it should be, to account for the greatest range of data with the minimum theory, we have several options open.

First, we could change all of our hypotheses, such as "bachelors have messier apartments," to mention the new qualification. I.e., we can instead say, "bachelors who are not in common-law relationships have messier apartments," "bachelors who are not in common-law relationships have less food in their fridges," etc. But, clearly we would be adding the exact same verbiage to each hypothesis. Given that we have many hypotheses about bachelors, we will have many hypotheses that contain the redundant qualification "who are not in common-law relationships." From the point of view of parsimony, this is not an ideal solution.

So, we are then left with two options. On the one hand, we could make a new word, say, "crachelor," to stand for "unmarried males, who are not in common-law relationships." Or, we could simply amend the definition of bachelor to *now* mean, "unmarried males who are not in common-law relationships." Now, the first option is certainly plausible but the point is that all I need you to accept is that the second is as well. That is, we can find it fit to amend our theory, so that a statement that was once true by definition (e.g. "Someone who is an unmarried male is necessarily a bachelor") is no longer true.

The point of this discussion was, again, to show that statements whose denials are self-contradictory are not necessarily statements we would hold come what may. Thus, there are, for Quine, at least two definitions of analycity in play, which do not necessarily amount to the same thing. The reason I have been at such pains to stress this is that, I will argue, Quine was getting at a fundamental insight in arguing against the second sort of analycity, i.e. the "come what may" sort of analycity. One might even consider arguments against this sort as the "spirit" of his paper, even if the "letter" of his paper was an argument against a notion of analycity (which he was, in fact, quite explicit about).

Now, what is Quine's problem with analycity? Well, Quine is concerned that one cannot define analycity except with reference to the notion of synonymy, but that one also cannot define synonymy without reference to analycity. Since both words are, says Quine, mysterious, he is unhappy because we are not able to reduce the definition of a mysterious word to a mundane one.

That is, suppose we want to say that, "Bachelors are unmarried males," is analytic precisely because "bachelor" and "unmarried male" are synonymous. Well, how do we know that these two words are synonymous? Suppose we say that two words are synonymous if they can be interchanged *salva veritate*, i.e. preserving "truth". Well, how would we know that truth has been preserved? That

is, how do we know that, "Unmarried males are messy," can be changed to, "Bachelors are messy?" preserving truth? We might suppose that truth has been preserved precisely if "Unmarried males are messy if and only if bachelors are messy," is an analytic truth. But, "analytic truth" was the term we were looking to define!

This line of argumentation is rather enjoyable. However, Quine seems to give deliberately short shrift to the one obvious way to define synonymy, which is in terms of definition. That is, it seems as though it would be perfectly reasonable to say that "bachelor" is synonymous with "unmarried male" because it is *defined* to be. In other words, we can let the notion of *is defined as* be primitive, and let synonymy and analycity be defined in terms of that[6]. Let $\alpha =_{\text{Def}} \beta$ be read as "α is *defined as* β". Then:

(104) α and β are *synonymous* when $\alpha =_{\text{Def}} \beta$.

And, we implicitly include the following axiom in each person's knowledge:

(105) $\forall\alpha, \beta, \left[\left(\alpha =_{\text{Def}} \beta\right) \rightarrow \left(\forall x\ \alpha(x) \leftrightarrow \beta(x)\right)\right]$

Thus, if (106), then (107):

(106) $bachelor =_{\text{Def}} \left[\lambda x \left(\neg married(x) \wedge male(x)\right)\right]$[7]

(107) $\forall x\ \ bachelor(x) \leftrightarrow \left(\neg married(x) \wedge male(x)\right)$

Note that we might identify a contingent fact, such as that "All and only bachelors are males that go to night clubs":

(108) $\forall x\ bachelor(x) \leftrightarrow male\text{-}who\text{-}goes\text{-}to\text{-}night\text{-}clubs(x)$

But, despite the fact that the set of bachelors is the set of males who go to night clubs, "bachelor" is not synonymous with "male who goes to night clubs" as long as the theory does not contain the statement

$$bachelor =_{\text{Def}} male\text{-}who\text{-}goes\text{-}to\text{-}night\text{-}clubs$$

That is, we can tell the difference between predicates that happen to have the same extensions, such as is expressed in (108), versus those that are coextensive by definition, as in (106–107), because only those predicates that are coextensive by definition will satisfy the $=_{\text{Def}}$ relation.

---

[6] Or, we could just let *synonymous* be primitive. But "is defined as" is more common in mathematics and science, so I propose to use that.

[7] I had been using, as my logical language in this chapter, the first-order predicate calculus. Let us now assume that we are using, for the rest of the chapter, Church's (1940) *Simple Theory of Types*. Note that $=_{\text{Def}}$ is a relationship between relations.

So, here, *is defined as* is a primitive term, i.e. it is not defined in terms of any others. It is rather unmysterious and, from this term, we can derive the others, i.e. synonymy and analycity. The point is actually not lost on Quine, who basically points out the possibility, before moving on without comment:

> There does, however, remain still an extreme sort of definition which does not hark back to prior synonymies at all; viz., the explicitly conventional introduction of novel notations for purposes of sheer abbreviation. Here the definiendum becomes synonymous with the definiens simply because it has been created expressly for the purpose of being synonymous with the definiens. *Here we have a really transparent case of synonymy created by definition; would that all species of synonymy were as intelligible*, (1951, p. 26, emphasis mine).

So, what Quine has essentially conceded is that, yes, there is a perfectly consistent way to draw an analytic-synthetic distinction, but he is asking us to brush this off on the grounds that there might be other "species of synonymy" which still remain a mystery. But, it was only in the bargain to give one "species" of analytic-synthetic distinction. And, I think one can argue that this one fits our intuitive ideas about the topic quite squarely.

Note that the term $=_{\mathrm{Def}}$ here is a logically primitive term in the sense that it is not defined in terms of others. It is also not given a *definition in use*. A definition in use (Ayer 1936) would explain how to remove the term from any theory (i.e. set of sentences) that contains it. But, $=_{\mathrm{Def}}$ cannot be removed from a theory. That is, one cannot replace $\alpha =_{\mathrm{Def}} \beta$ with,

$$\forall x\ \alpha(x) \leftrightarrow \beta(x)\dagger$$

This is because, as we have been discussing, † could be a contingent fact and it is precisely the statement $\alpha =_{\mathrm{Def}} \beta$ that tells us it is not.

Also note that it would be inappropriate in the extreme to complain that we use the term *define* without defining it in terms of other words. It is a fact about *language* itself that, in general, if the language has a finite number of words, then either there must be words which are not defined in terms of others (i.e. primitives[8]) or one must have circular definitions. Rather than prove this fact, which involves some technical detail, I will merely quote from Hempel:

[8]This notion of "primitive" should not be confused with the nativistic kinds of primitives to be discussed in §5.3.1. Cf., ff. 10.

> Not every term in a scientific system, therefore, can be defined by means of other terms of the system: there will have to be a set of so-called primitive terms, which receive no definitions within the system, (Hempel 1966, p. 88).

So, I have endeavored just now to show two things. The first is that there is a difference between statements whose denial is self-contradictory, on the one hand, and statements which we would hold to "come what may." The former are not necessarily the latter (in fact, the latter category, I have argued, is empty). Further, I have endeavored to show that one *can* in fact come up with a perfectly consistent, rigorously defined notion of a "statement whose denial is self-contradictory," based on definition.

Now, Quine was quite explicit about the fact that he was arguing against a notion of analycity of the "statement whose denial is self-contradictory" sort. But, what I suggest is that there is a fundamental insight that Quine is getting at. This holds even if one does create a way to consistently view analycity.

This insight is: in a scientific theory, when something goes wrong, it might be any statement which might have to go. Even a statement of the form "X is defined as Y" might have to go, as we saw with our example about bachelors and common-law marriages. Thus, *even synonymy is, in some sense, contingent*. This, I propose, is Quine's fundamental insight, which is preserved even though we can come up with a sensible way to define synonymy.

Moreover, apart from definitional synonymy, there is no other sort of analytic statement. In particular, there are certain notions of analycity floating around, which Block (1993) refers to as "traditional ideas" about analycity (p. 3), in which sentences like, "All dogs are mammals," or, "All dogs are living," are considered analytic.

It seems to me that maybe such statements are thought to be analytic because they are so "obvious" or "common-place." But, the criterion for analycity that I have laid out has nothing to do with obviousness. If part of the definition of dog includes being a mammal, then, "All dogs are mammals," is analytic. If animalhood is not part of the definition of doghood, then this statement is not analytic. It is as simple as that.

So, to sum up, I think that, even though the letter of Quine's paper was a tirade against synonymy and analycity, we can in fact come up with a perfectly coherent notion of the two. But, one can argue that Quine's underlying maxims were actually that:

(109) Theories are tested as wholes.

(110) When something goes wrong with a scientific theory, no statement is safe.

These are fundamental tenets of the modern philosophy of science and were both absent from the "empiricism" that Quine was revolting against.

That said, it will also be prudent to make some comments about the domain of discourse. Quine's remarks are clearly in reference to the philosophy of science, in which the totality of science is seen as some kind of monolithic entity, almost existing independently of a thinker. That is, he refers to a notion of "total science," (p. 39). "Total science" is a theory which is being self-consciously tested.

Now, if we turn our attention to psychology, and to communication between humans via natural language, we may want to reach different conclusions[9]. That is, in §5.4, I am going to propose that communication requires a *kind of* analycity. That is, we will have need of a distinction between properties which are *necessarily* considered true, based upon the conventionalized word-meanings, on the one hand, and those which are *contingently* considered true, on the other. And, I will argue that this kind of analycity is really untouched by Quine's polemics. Thus, it has been important to consider Quine's actual arguments and concerns, rather than to simply swing around the slogan "there is no analytic-synthetic distinction," like a wooden mallet.

5.3 What Holism Buys

The questions just raised are legitimate and must be addressed if a holistic theory of meaning is to be taken seriously. But, before I answer them (i.e. in §5.4), we should first take a look at what a holistic model of the knowledge of language buys. That is, we should look at what it explains that other theories cannot—at why we should *want* to bother defending a holistic theory.

Now, we began our discussion of holism by noting that the theory of comprehension presented in §4, and especially §4.1, was a holistic one. We saw, in that section, an extremely simple but powerful model of comprehension, in which comprehension was modeled as the translation of a natural language sentence, U, into a logical language translation, $\mathfrak{l}_u$, from which inferences were computed in conjunction with word-/world-knowledge.

In that section, we saw that this sort of model can give meaning to the atomic parts—i.e. the

[9] I should note that, in Quine 1960, Quine attempted to use the letter, rather than the spirit, of his (1951) result in formulating conclusions about the interactions of speakers in speech communities. I will not discuss Quine (1960) in much detail (though cf. §5.4.3.2, for some discussion). I do not like his arguments there. All I will say is that Quine 1951 is both concise and nearly universally regarded as formidable in the philosophy of language, so I felt it worthwhile and somewhat necessary to rebut criticisms stemming from this in detail. Quine 1960 is not concise and does not have the same universal respect, and I furthermore do not like it much, so I propose to ignore it.

words—that compositional semantics can build the meanings of larger expressions out of. We also saw that this sort of theory has predictive ability, in that it can predict what sorts of questions people will be able to answer after being supplied with premises (cf. the discussion of §4.1).

So, one reason to defend holism is to defend our aforegiven model of the process of comprehension of a mature hearer. There are, in addition, other theoretical advantages to being able to adopt a holistic theory of meaning.

Consider that, in a holistic theory, each word is dependent for its meaning on the other words in the given idiolect. The primary conceivable alternative to such a theory is one in which there are some words which are not dependent for their meaning on any other words. Words that do not depend for their meaning on other words are referred to as *semantic primitives*, or *conceptual primitives*[10].

In §5.3.1, I will look at two theories that involve some kind of conceptual primitives, and conclude that there is great reason to be skeptical of the concept of semantic primitives. In §5.3.2, we will see how a holistic theory of word-meaning can avoid the use of primitives and their attendant problems.

5.3.1 The Short-Comings of a Primitives-Based Semantic Theory

Fodor 1975 puts forward several interesting claims about language. First, he overviews some arguments that there ought to be, in some sense, a "language of thought." That is, if thinking is computing, and computing requires the manipulation of representations, then there ought to be some mode of representation for these thoughts: "Computation presupposes a medium of computation: a representational system," (1975, p. 27).

This might seem to be an obvious truism, but Fodor claims some measure of controversiality because, "Wittgenstein is supposed to have proved that there can be so such thing as a private language," (p. 68, referring to Wittgenstein 1953).

---

[10] Unfortunately, there are two notions of *primitive* being used in this chapter. This problem runs deeper than unfortunate terminology, as the concepts named are as closely related as the names to describe them. A *primitive term* in a theory, whether scientific or mathematical, is a word that is not *defined* in terms of others. It cannot be eliminated from the theory. So, in §5.2, I said that *is defined as* was going to be a primitive (i.e. would not be defined in terms of anything). Then, I said that *synonymous* and *analytic* would be defined in terms of *is defined as*. Both *synonymous* and *analytic* could be removed from our theory without changing the conclusions of the theory, and so are not primitive.

A *semantic primitive* is one which does not depend for its *meaning* on any other words. This is a theory-laden notion, that depends on what notion of "meaning" one is working with. For example, in a theory that uses semantic primitives, *large* might be a primitive concept, which would mean that we can say what *large* "means" without having to refer to any other concepts such as, e.g., *small*. Typically, in a primitives-based theory, we are born already knowing what each of the primitives "means."

His most striking thesis comes when he gets down to the question as to just what this "language of thought" must be like. In particular, what sort of characteristics must this language have if it is going to enable us to learn an arbitrary natural language?

Well, "we have no notion at all of how a first language might be learned that does not come down to some version of learning by hypothesis formation and confirmation," (p. 58). Here, "learning a language involves at least learning the semantic properties of its predicates," (p. 59), and, "*S* learns the semantic properties of *P* only if *S* learns some generalization which determines the extension of *P*," (p. 59)[11]

Fodor continues:

> [O]ne cannot learn a language unless one has a language. In particular, one cannot learn a first language unless one already has a system capable of representing the predicates in that language and their extensions. And, *on pain of circularity, that system cannot be the language that is being learned*. (p. 64, emphasis rearranged)

His conclusion is the following: "one can learn [a language] L only if one already knows some language rich enough to express the extension of any predicate of L," (p. 80). Now, "language" can mean a great many things. But, what Fodor has in mind is that, whatever English, French, etc., are, the child *already* has one of these from birth. And, however it might be that English talks about things (i.e. by expressing extensions of predicates), the innate language can do the same things.

In other word, for each concept that we have, such as "cat," "dog," "Olympic games," "Communist Party," "non-deterministic Turing machine," etc., either this concept is a primitive concept, which we are innately capable of expressing, or else it is somehow a combination of these primitive concepts.

One might wonder how a complicated modern concept like "non-deterministic Turing machine" could be a primitive concept for an animal that has not evolved much in the past fifty thousand years. Fodor anticipates this sort of concern and so allows that some concepts can be expressed as a combination of other, primitive concepts. That is, maybe "airplane" is not a plausible primitive. But, maybe, he says, "airplane" can be decomposed into "flying machine," where "flying" and "machine" are primitives[12].

---

[11] Fodor goes on to complicate the picture for the benefit of those who prefer an intensionalist or a Putnam (e.g. 1975) stereotypic version of semantics, where a word is connected to its intension or stereotype, rather than its extension. These details are not really relevant for the point at hand.

[12] This example is actually Fodor's, rather than an adaptation. Cf. Fodor 1975, p. 96, for his discussion.

Of course, one still has to wonder whether "machine" is really such a reasonable primitive predicate, given the pace of evolution. If not, what primitives does it decompose as? It really should not be the job of the skeptic to ask questions like these. If Fodor, or anyone else who had read his work, had taken this theory seriously, the way to pursue it would be to list the semantic primitives that would be required to represent all of natural language and then show how a very large fragment of the words in some natural language, if not all of the words, can each be represented using combinations of these primitives.

The fact that Fodor has never done this is telling. The task seems hopeless to me and one can easily imagine that he realizes this as well. However, there is at least one group of researchers, led by Wierzbicka (1972), that has devoted considerable effort to the research of semantic primitives. As of 1994 (i.e. Goddard and Wierzbicka 1994), this group had found evidence for 36 semantic primitives. I will list them in full here:

(111) I, you, someone, something, people, think, say, know, feel, want, this, the same, other, one, two, many, all, do, happen to, no, if can, like, because, very, when, where, after, before, under, above, kind of, have parts, good, bad, big, small

One might doubt that were are going to get "non-deterministic Turing machine" out of this set. It is not even clear how one could get Fodor's "flying" or "machine" out of (111).

But, the reader might imagine, surely Wierzbicka and her colleagues have demonstrated that this list can, in fact, generate a substantial number of the words in use in natural language. Well, this would be an incorrect assumption. Wierzbicka and her colleagues have not endeavored to demonstrate that their primitives can generate a wide range of the words in natural language, nor do they seem to feel that this is important.

Wierzbicka feels the real problem facing a primitivist is that one must motivate each primitive added to the list rigorously, so that list of primitives does not contain any redundancy.

That is, the primary concern is that the list in (111) might contain *too many* primitives, and so the theoretical task that an article, such as one in Goddard and Wierzbicka 1994, is required to achieve is to demonstrate that the addition of a new primitive is actually necessary.

But, this methodology *assumes* that a primitive-based framework is correct. It does not give any evidence that the approach is in fact correct. The skeptic of primitives is concerned that a list like (111) is far, far, far too *small*. That is, we want to see evidence that the number of words generated by a list of primitives is large enough, not that the list of primitives itself is small enough.

One might speculate that the reason it does not occur to Wierzbicka to employ this methodology

is that the task is not possible. Suffice it to say that skepticism is fully justified until the explanatory power of this list has been demonstrated.

Of course, we can investigate Wierzbicka's proposal in some level of detail for she has actually endeavored, to some extent, to list what the primitives would be. Fodor has saved himself the embarrassment by not even trying. I think that one has to conclude from this discussion that there is very little in the way of positive evidence that a primitives-based program is workable.

Before moving on, it should be noted that Fodor did, in fact, express some concern that his talk of primitives was entering the realm of the ridiculous, suggesting that he, "should be inclined to view [his conclusions] as a *reductio ad absurdum* of the theory that learning a language is learning the semantic properties of its predicates, except that no serious alternative to that theory has ever been proposed," (1975, p. 82, emphasis mine). Well, in §5.3.2 we will see why holism provides us with a "serious alternative" for the explanation of the semantic aspect of language acquisition, sparing us from the apparent absurdity of the concept of primitives[13].

5.3.2 Holism Succeeds Where Primitives Fail

One of the major advantages of the sort of holistic model of knowledge that the model in §4.1 presents is that *there does not need to be a single innate or primitive semantic predicate*. There does not need to be a fixed set of features from which all concepts can be constructed. And, there does not need to be any language known in advance.

The maxim, is this: semantic knowledge is represented as a set of statements in a logical language. Period. All that the innate basis needs to provide is the mechanisms to store a knowledge set such as (101), repeated here, and to draw inferences from these.

$$\text{(112)} \quad \mathbf{K} = \left\{ \begin{array}{c} \forall x\ bachelor(x) \leftrightarrow (\neg married(x) \wedge male(x))\,, \\ \forall x\ married(x) \leftrightarrow has\text{-}spouse(x) \end{array} \right\}$$

That is, the innate basis provides only for the means of combination, by which symbols can be formed to create (some neural representation of) logical language statements, and to draw inferences from these. *With the means to combine and make use of arbitrary symbols, the child can learn to use any possible predicate, even if, at birth, they do not know how to use any.*

The reason for this is precisely the fact that there is no single "direction" in which meanings are built. That is, the meaning of $bachelor$ depends on $married$ and the meaning of $married$

[13] As we will see in §5.5, Fodor will later be the chief critic of the holistic theory that succeeds where his primitive-based theory fails.

depends on $bachelor$. We are not in a situation where we begin with one set of innate and primitive "meanings" and derive the rest. The various symbols conspire to define each other.

Let us consider how a child would be said to learn new words in a holistic model. Let us assume that the child knows enough syntax to know how to assign a semantic representation to sentences of the form, "Fs are G." Then, on hearing that, "Cats are furry," the child has learned something about both "cats" and "furriness." (That is, cats are things that are furry, and furriness is a property that cats have.) She stores this.

Then, upon hearing that, "Cats chase mice," she has learned more about "cats," as well as learning about "chasing" and "mice." The child simply keeps updating her knowledge with new facts/beliefs about each entity and relation as she hears them.

So, it seems to be precisely because of the fact that meanings of words can be learned together, on the basis of the meanings of one another, that we can get by without semantic primitives. If there were some meanings that did not depend on others, *those* would be the primitives, and one would have to ask where they had come from.

And, if the primacy of these primitives were crucial to the acquisition of the rest of the word-meanings, then we would be in precisely the problematic situation broached in the last section. First, we would have to ask where these primitives had come from. And, second, we would have the problem that, since the primitives are assumed to be crucial for the acquisition of the non-primitive symbols, the primitives would have to suffice to produce *all* concepts in use today.

Note that, in this holistic model of language acquisition, it is unproblematic to explain how the child learns to use predicates incrementally. That is, the child can learn that "Cats are furry" on one day and then that "Cats chase mice" on another. The child does not need to know that cats chase mice in order to learn that they are fuzzy.

Traditionally, extensionalists such as Lewis (1970), Fodor (1975), Lepore (1983) and Dowty et al. (1981) have held that, in learning to use words, the child must learn some way to determine the "extension" of each term in use. That is, when the child hears talk of "cats," their mind will allow some way to exploit an ability to recognize "the set of all cats."

But, intuitive as this sounds, this theory is riddled with problems, some of which were outlined in §2.2.2. In addition to those, there are criticisms of the following sort. If somebody has ever been lied to, then he clearly does not know how to recognize "the set of all liars." If somebody has ever lost money in the stock market, then he clearly does not know "the set of all good stock picks." So, obviously whatever the person knows about how to use a predicate, it is not *literally* how to fix that predicate's extension in the "world of non-symbols."

However, there is an additional problem pressing us in this regard when it comes to the acquisition of language, which is the question of just when the child is able to fix the extension of the words she is learning. Is the child prepared to fix the extension of each term after hearing the first fact about it? That is, is the child ready to fix the extension of "cat" after hearing that, "Cats are furry"? Many other things are furry. If the child is not immediately ready to exploit the connection between the word "cat" and the set of all cats, then what does she do until she is ready? How does she store this fact about cats until then? And, when is she going to be ready to fix the extension of all cats?

I am not aware that anyone has suggested how this might be done, despite the fact that it has been nearly forty years since Lewis blasted the internalist linguists for failing to exploit the word-world connection in their semantic theories.

In a holistic framework, the child's use of the word "cat" is based on her knowledge set at that time. It does not depend on the extension of "cat" but on its inferential role in her developing language. The child's concept of "cat" might be less full than it will be when she is mature. But, this poses no problem, it simply means that she will not draw as many inferences about cats as she will when she is older. Her "cat" concept can grow from nothing into a fully mature concept, at which point the child will be likely to label as "cats" the same stimuli that her other speech-community members label as cats.

Furthermore, we do not have the problem of explaining, as primitives-theorists like Fodor and Wierzbicka must eventually do (although they have not), how it is that the child ascertains which primitives some given word is made up of. That is, even if we assume a primitives based system, when the child encounters a sentence like (113), how does the child figure out which combination of primitives "airplane" consists of?

(113) We are going to take an airplane to Europe, Sally.

To my knowledge this has not been explained. This is despite the fact that the assumption that evolution had anticipated, in some crucial way, the full range of concepts that the modern human (as well as all future humans) are equipped with is so costly from a scientific perspective. One might think that with such a costly assumption as that a list of primitives to generate all words in any language can be found, the problem of how a child learns a word should be trivial. But, it is not. There is still no idea how to do it.

In contrast, a holistic theory has it that the child interpreting (113) would not need to break the term into primitives. If she knows that airplanes are vehicles, then she knows that the family

is taking a vehicle to Europe. If she knows that airplanes arrive at their destinations within two days after their departure, then she will know that the airplane will get to its destination within two days of its departure, etc. We have given an explanation of how both the child and the adult would understand a sentence in (4). Thus, the holistic theory is not only much less ontologically costly than the primitives based theory, it actually affords a much easier explanation of what comprehension consists of.

I want to stress that what I have given here are not just *arguments* in favor of structuring a model of knowledge or comprehension in a certain way; I have given the actual model *itself*. That is in §4.1, I have explained how comprehension should be modelled. In this section, I have explained, roughly, what innate basis is needed for the acquisition of adult semantic competence and, if only with broad brush strokes, how this acquisition takes place.

It would be a patent mistake, it seems to me, to think that those arguing against a holistic model, especially Jerry Fodor, have a model of language acquisition of this level or any level of detail that they are proposing instead. They do not. We are going to look at some of the attacks that have been made against the concept of holism and on holistic models in the next section and I will answer them. These are legitimate concerns, which are always of the form, "You still haven't explained how X happens in a holistic framework." (Some would call these research questions rather than fatal flaws.)

And, in addition to addressing the legitimate criticisms of holism in §5.4, I will also address, in §5.5, those of Fodor and Lepore (1992).

However, the point is that the holists are at least on the scoreboard here with a model of how we can have language with a minimal innate basis, how the adult competence can be acquired, and how comprehension can proceed once that competence is acquired. This is not a competition between rival theories (unless one counts as legitimate, e.g., Fodor's [1975] admittedly absurd primitives-based theory recounted above). There is one theory which can so far explain the aforementioned facts—holism—along with its detractors.

5.4 Answering the Legitimate Criticisms of Meaning Holism

Now, we are going to look at how the criticisms listed in §5.1.2 might be addressed. In §5.4.1, we will first look at some solutions that have been provided in the literature. I will argue that these answers are not satisfactory. Then, in §5.4.2, I will give my own solution.

5.4.1 Solutions from the Literature and their Short-Comings

5.4.1.1 Block's Two-Factor Theory

Block is a main proponent of inferential role semantics in the modern literature and, in Block 1986, 1993, he lays out a "two-factor" theory that is meant to address the sorts of criticisms discussed in §5.1.2.

To recapitulate, the criticism of meaning holism is that it seems that, if two people are drawing different inferences on the basis of a sentence, like, "Socrates is a man," then they have different meanings for this phrase. How can communication succeed when the speaker cannot know what the meaning of his statement will be to the hearer.

Block is going to propose that a word's inferential role is only one "factor" of a word's meaning. The second factor is its connection to its extension in the world. In other words, this second factor is really a full-blown extensional theory of semantics, of the sort discussed in §2.2.1.2. Thus, while the inferential role part of "cat" for Henry is unstable, and idiosyncratic to Henry himself, the connection of the word to the set of all cats in the world *is*, according to Block, stable. So, the instability of inferential roles is compensated for by the stability of an extensional theory.

Also, if I were to say that, "Toads cause warts," and then later deny it, saying, "Toads don't cause warts," then the inferential role of "toads" and "warts" has changed. Thus, if one equates meaning with inferential role, then the meaning of what I asserted was not meaning of what I denied.

To avoid this problem, Block proposes that, for analyzing disagreements, we do not use inferential role. We instead use the externalist factor—the "wide contents"—in which case the wide contents of "toads," and "warts," are the sets of toads, and warts, in the world, respectively. These do not change (or at least, are stable enough to satisfy externalist philosophers). This wide contents factor contrasts with the "narrow contents" factor of a word, which is its inferential role.

Essentially, we have a situation in which the externalists are criticising the inferential role semanticists for the perceived instability of inferential roles. Block evidently agrees that the externalist theory *does* exhibit meaning stability. So, what Block wants to do is to attempt to adopt both theories simultaneously.

Now, since Block has adopted *both* theories, one might ask what the point is of using the inferential role (i.e. narrow content) factor at all. Well, for one thing, it solves Frege's problem. That is, it gives "Morning Star" a different meaning than "Evening Star." Frege's (1892) distinction between *sense* and *reference* already did this, but in a way that has been criticized for naming a problem without analyzing it (cf., e.g., Davidson 1967). Montague's (1974) intensions also provided a solution

to Frege's problem, but some have cited qualms about psychological plausibility (cf., e.g., Partee 1980). So, one could argue that one merit of this gargantuan theory is a psychologically plausible solution to Frege's problem.

But, overall, inferential role plays a rather limited role in Block's paradigm:

> [N]arrow content [i.e. inferential role] has a role in *psychological explanation*. For purposes of certain kinds of psychological explanation, narrow content differences matter despite the fact that they dont make for differences in truth-conditional content.... *By contrast*, truth-conditional content attribution is useful for *communication* and other contexts where information is important, and where psychological differences dont matter, (1993, p. 17, emphasis mine).

In other words, one needs one kind of theory when doing psychology, and another kind of theory when dealing with communication. One might wonder what it is that communicates, if not brains.

Aside from this, the glaring problem here is that Block has taken two theories that are ostensibly, and in the minds of most practitioners of each, antagonistic and grafted them together. And, it is not as though he has done any kind of surgery, or theoretical work at all, to either position to make it more amenable to the other. He is simply proposing to take both theories, *in their entirety*, and use whichever theory happens to work better for the given phenomenon.

And, as Fodor and Lepore (1992) aptly note about these two apparently heterogeneous factors of meaning: "We now have to face the nasty question: *What keeps the two factors stuck together?*," (p. 170). Block does not say. Personally, I have no idea what would connect the mental and external factors that any externalist semantic theory presupposes. We discussed in §5.3.2 certain qualms about the disconnect between factors evinced by a person's inability to discern "the set of all good stock picks."

But, the criticism is an odd one coming from such avid externalists as Fodor and Lepore. I have to wonder why Block cannot appeal to whatever Fodor and Lepore have been appealing to to keep the mind's contents connected to the contents of the "world." But, in any case, it seems to be a fair criticism that if Block is going to simultaneously hold to the two theories, he ought to explain how they link up.

5.4.1.2 Meaning Similarity

Fodor and Lepore (1992) suggest (as a foil idea) that one method to explain interpersonal communication is to suppose that "meanings" need not be identical between conversation participants in

order for communication to succeed—they only need to be *similar*.

In other words, Annie's inferential role for red is quite certain to differ from Bob's. Again, since red then "means" something different for Annie than for Bob, critics are worried that communication about red will not go through. But, Annie's inferential role for red might be somehow *similar* to Bob's and maybe that would be enough to let communication go through.

Fodor and Lepore's criticism of this position is that a notion of similarity of inferential roles A and B presupposes a notion of identity for inferential roles. But, the only notion of identity that we have for inferential roles is complete identity.

That is, we can tell if A and B are *completely* identical but cannot make any further distinctions between them if they are not. So, we might try to split some symbol's inferential role into "components," and then to say that Annie and Bob have *similar* inferential roles exactly if many of their components, say $\mathrm{A}_1, \ldots, \mathrm{A}_\mathrm{n}$, and $\mathrm{B}_1, \ldots, \mathrm{B}_\mathrm{n}$, respectively, are equal. Well, none of their components are going to be equal because, as we assumed at the beginning, no inferential roles are going to be equal. So, this sort of notion of similarity cannot get started precisely because we have no identity in the system.

I think this is an apt criticism and so will conclude that meaning similarity is not the answer to our problem.

5.4.2 A New Solution

Stated once more, one problem raised in the holism literature is, if inferential role is identified with "meaning," and no two people will have identical inferential roles, then it would seem that people cannot communicate, since each word used by the speaker will have a different "meaning" for him than it will for the hearer.

Or, terms more similar to §4, the relevant questions are:

(114) How can people agree or disagree? If Annie and Bob infer different things from ϕ, then how can they agree on ϕ?

(115) How can someone change their mind? When Annie believes ϕ, the symbols in ϕ have one set of inferential roles. When she rejects ϕ, they have another set.

(116) How is it that *interpersonal communication* can occur when the **K** of each speaker is potentially different and when, corollarily, the inferential role of each word is different for each speaker? Is it not possible that the hearer will draw completely different inferences on the basis of an utterance than the speaker would have?

5.4.2.1 Agreement, Disagreement, and Changing One's Mind

The questions listed in (114) and (115) are the most easily handled. We have only to adopt (or, realize) the following maxim:

(117) Speakers agree and disagree about *sentences*, not their inferential roles.

I will illustrate (117) via the following example about Annie and Bob. Assume that $\mathbf{K}_A$ below is Annie's knowledge, and $\mathbf{K}_B$ is Bob's:

(118) a. $\mathbf{K}_A = \left\{ \begin{array}{c} \forall x\, bachelor(x) \leftrightarrow (\neg married(x) \wedge male(x))\,, \\ \forall x\, married(x) \leftrightarrow has\text{-}spouse(x) \end{array} \right\}$

b. $\mathbf{K}_B = \left\{ \begin{array}{c} \forall x\, bachelor(x) \leftrightarrow (married(x) \wedge male(x))\,, \\ \forall x\, married(x) \leftrightarrow \neg has\text{-}spouse(x) \end{array} \right\}$

Throughout this and the next section, assume that Annie and Bob each "think" in the same language, LL. That is, Annie's knowledge, and Bob's knowledge, which are $\mathbf{K}_A$, and $\mathbf{K}_B$, respectively, will be sets of LL sentences. We will also always assume that they both speak the same natural language, NL, which will be English.

Note that Annie defines "bachelor" in the traditional way, as "unmarried male," while Bob defines it unusually as "married male." Also, note that Annie defines "married" traditionally as "having a spouse," while Bob defines it unusually as "*not* having a spouse." Thus, $married$ has a different inferential role for Annie than it does for Bob, as does $bachelor$.

Furthermore, while Annie and Bob may disagree as to whether bachelors are married, they actually *agree* that bachelors do not have spouses. This is because Bob has two unusual definitions, which cancel each other out, essentially, in terms of the relationships between $bachelor$ and $has\text{-}spouse$.

So, if Annie were to say, "Peter does not have a spouse, so he is a bachelor," or, "All bachelors do not have spouses," Bob would agree, failing to realize the extreme level of disagreement between them about what bachelors and married things are.

However, if Annie were to say, "All bachelors are unmarried," by (117), Bob would disagree. That is, he holds to a different statement, regardless of the fact that the inferential roles differ. A contradiction has been elicited at the level of sentences.

A similar situation obtains with changing one's mind. Suppose we want to say that one changes their mind from believing that (119a) to (119b):

(119) a. Toads cause warts.

b. Toads do not cause warts.

Well, we can define *changing one's mind* as having some belief ϕ such that, at time t_0, $\phi \in \mathbf{K}$, where $\mathbf{K}$ is consistent (i.e. has no contradictions), and, then, at some later time $t' > t_0$, having $\neg\phi \in \mathbf{K}$.

Changing one's mind does not, as I have just defined it, refer to inferential role at all. So, when one trades the belief (119a) for (119b), the fact that the inferential roles of "toad" and "wart" have changed does not in any way preclude one's mind from being changed. Again, what a person accepts and rejects is the statement itself, not the inferential role.

So, (114) and (115) were easily handled by assuming (117). Before moving on, however, we should ask, what is the empirical nature of the statement (117)? Is it an assumption that gets our theory to work out? Is it an empirical statement that could be tested with an experiment? It would seem to be both.

First of all, with a simple change in outlook—i.e. to one in which it is *sentences* rather than meanings that are agreed and disagreed upon—problems (114) and (115) disappear. There does not seem to be any good reason why one should insist that sentences cannot be the locus of agreement and disagreement, and taking them to be immediately alleviates theoretical problems.

Furthermore, I have sort of stipulated Annie and Bob's behavior in this case. In other words, I have simply decreed that, if Annie and Bob's knowledges were as in (118a–118b), Annie's mention of bachelors being unmarried would prompt Bob's disagreement. Well, this seems to me like what would happen but, ultimately, one would be justified to ask to see some experimental evidence that this would indeed be the case.

5.4.2.2 Interpersonal Communication and Group Language

I will begin by repeating (116) here:

(120) How is it that *interpersonal communication* can occur when the $\mathbf{K}$ of each speaker is potentially different and when, corollarily, the inferential role of each word is different for each speaker? Is it not possible that the hearer will draw completely different inferences on the basis of an utterance than the speaker would have?

Note that we will have effectively addressed (116) if we can prove the following:

(121) There is a lower bound on the number of the same inferences that the hearer and the speaker will make on the basis of some sentence, ϕ.

(122) There is a guarantee that that there will not be any "intolerable" contradictions between what is inferred by the hearer and what is inferred by the speaker on the basis of some sentence, ϕ.

To elaborate on (121), suppose Annie knows/believes that "a bachelor is an unmarried male." Then, Bob says to Annie, "We need to find someone who is unmarried to attempt a dangerous mission." Then, Annie replies, "Tom is a bachelor." By this she intends that Bob will realize that Tom is an unmarried candidate for the mission.

In this case, Bob needs to infer from Tom's being a bachelor that Tom is not married. In other words, there will be a problem if we cannot put some lower bound on the number of inferences that Bob will draw that Annie will also draw.

5.4.2.3 Motivating a Notion of Social Analycity

In this section, I would like to address the notion of "intolerable" contradictions that was broached in (122), and which clearly seems somewhat fuzzy. I would also like to begin to address the question as to how a common set of inferences shared by linguistic community members might be negotiated.

Let us assume the following. Annie and Bob each have their own knowledge-set, which we will call $\mathbf{K}_A$ and $\mathbf{K}_B$, respectively. We will suppose that, between them, there is a *conversational record*, which is a pair of sets of sentences, and which will be denoted $CR = \langle CR_A, CR_B \rangle$. CR_A contains all of the sentences that Annie has spoken during the conversation. CR_B contains all of the sentences that Bob has spoken. So, we will assume that whenever either Annie, or Bob, utters some statement, that statement is added to CR_A, or CR_B, respectively.

Now, Annie does not have access to $\mathbf{K}_B$ directly, nor does Bob have access to $\mathbf{K}_A$. Annie will only notice that she and Bob have contradictory beliefs when $\mathbf{K}_A$ is not consistent with CR_B. (Correspondingly, Bob can only notice a contradiction with Annie once $\mathbf{K}_B$ and CR_A contains a contradiction.)

And, whenever Annie or Bob notices a contradiction between the premises in their own knowledge and those in the conversational record, we will assume that they will speak aloud each premise from their knowledge that was responsible for the contradiction, thus effectively putting a proof of the contradiction into the conversational record.

To see what I mean, suppose that Annie and Bob have knowledges $\mathbf{K}_A$ and $\mathbf{K}_B$ respectively, which are as follows:

(123) a. $\mathbf{K}_A = \left\{ \ \forall x \ bachelor(x) \leftrightarrow (\neg married(x) \wedge male(x)) \ \right\}$

b. $\mathbf{K}_B = \left\{ \begin{array}{c} \forall x\, bachelor(x) \leftrightarrow (married(x) \wedge male(x))\,, \\ married(\textsc{Mark}) \end{array} \right\}$

Now, suppose Bob says to Annie, "Mark is a bachelor. He's married." That is, we now have the conversational record looking like this:

(124) $CR = \langle CR_A, CR_B \rangle$
$CR_A = \varnothing$
$CR_B = \left\{ \begin{array}{l} bachelor(\textsc{Mark}) \\ married(\textsc{Mark}) \end{array} \right\}$

(That is, Annie has not said anything. Bob has said two things.)

Now, let us ask what Annie will infer on the basis of $\mathbf{K}_A$ and CR. Well, first of all, $married(\textsc{Mark})$ is in CR_B, so, on the basis of $\mathbf{K}_A$ and CR_B, one thing Annie would infer is $married(\textsc{Mark})$. But, Annie would also infer that Mark is *not* married. This is because $bachelor(\textsc{Mark})$ is in CR_B, and $\forall x\, bachelor(x) \rightarrow \neg married(x)$ is in $\mathbf{K}_A$[14]. So, she has derived both one statement and its negation, i.e. Annie has derived a contradiction.

Now, I said above that, upon discovering a contradiction between one's own knowledge and the conversational record, a speaker would effectively put the proof of that contradiction into the conversational record. What I meant by that was the following. Annie has derived a contradiction, i.e. by deriving both $married(\textsc{Mark})$ and $\neg married(\textsc{Mark})$.

Suppose that the list of statements, P, is Annie's proof of $married(\textsc{Mark})$ and P' is her proof that $\neg married(\textsc{Mark})$. Well, we are going to assume that she speaks aloud each of the premises on which P and P' are based that *are not in* CR_B.

For example, her proof of $married(\textsc{Mark})$ is:

$married(\textsc{Mark})$ (premise from CR_B)

(This is a simple proof because $married(\textsc{Mark})$ is one of the premises.

[14] Actually, it is in $Cn(\mathbf{K}_A)$.

Her proof of $\neg married(\text{MARK})$ is:

$bachelor(\text{MARK})$	(premise from CR_B)
$\forall x\, bachelor(x) \leftrightarrow \big(\neg married(x) \wedge male(x)\big)$	(premise from $\mathbf{K}_A$)
$\forall x\, bachelor(x) \rightarrow \neg married(x)$	(from the previous line and the basic rules of logic)
$\neg married(\text{MARK})$	(from the previous line and the first line)

(Here, I have marked next to each derived statement on what basis it was derived.)

These proofs relied on the following premises: $married(\text{MARK})$, from CR_B; $bachelor(\text{MARK})$, from CR_B; and †, below, which is in $\mathbf{K}_A$:

$$\forall x\, bachelor(x) \leftrightarrow \big(\neg married(x) \wedge male(x)\big)\dagger$$

So, Annie is going to speak aloud each of the premises that she had to supply (as opposed to those in the conversational record) to derive the contradiction. That is, she will speak aloud all of the premises in each proof that come from $\mathbf{K}_A$, which, in this case, is just †.

So, † is placed on the conversational record, which will now look like this:

(125) $CR = \langle CR_A, CR_B \rangle$

$$CR_A = \left\{ \forall x\, bachelor(x) \leftrightarrow \big(\neg married(x) \wedge male(x)\big) \right\}$$

$$CR_B = \left\{ \begin{array}{l} bachelor(\text{MARK}) \\ married(\text{MARK}) \end{array} \right\}$$

From here, Bob notices that Annie has put,

$$\forall x\, bachelor(x) \leftrightarrow \big(\neg married(x) \wedge male(x)\big)$$

into the conversational record. Meanwhile, Bob, as we have seen, holds that,

$$\forall x\, bachelor(x) \leftrightarrow \big(married(x) \wedge male(x)\big)$$

So, on the basis of a noticed contradiction by one of the parties (i.e. Annie, in this case), both of the parties are able to find out the root cause of the contradiction. That is, they are able to find out which underlying difference belief led to the noticed contradiction. I will take this to be a general result. That is, any time either Annie or Bob notices that the other has said something apparently contradictory, both will be able to discover which premises in $\mathbf{K}_A$ and $\mathbf{K}_B$ led to that contradiction.

Also, we will assume that Annie and Bob want to be part of a homogenous speech community (see §5.4.3.2 for discussion of this assumption). And, since it *feels* as though they are bound to have many mix-ups if this difference in word usage persists, we can say that at least one of them is going to have to change their views about $bachelor$s if they are actually going to be able to act as members of the same community.

So, are we going to assume that, each time Annie and Bob notice a contradiction between their beliefs, one or both of them will need to reconcile their beliefs to the other? Consider now a different case, in which both Annie and Bob define "bachelor" as "unmarried male" but differ as to whether they think Tom is married, say with Annie being quite sure that Tom is married, while Bob is quite sure that Tom is not. That is, suppose $\mathbf{K}_A$ (i.e. Annie's belief set) and $\mathbf{K}_B$ (i.e. Bob's belief set) are as follows:

(126) a. $\mathbf{K}_A = \left\{ \begin{array}{c} \forall x\, bachelor(x) \leftrightarrow (\neg married(x) \wedge male(x))\,, \\ married(\textsc{Tom}) \end{array} \right\}$

b. $\mathbf{K}_B = \left\{ \begin{array}{c} \forall x\, bachelor(x) \leftrightarrow (\neg married(x) \wedge male(x))\,, \\ \neg married(\textsc{Tom}) \end{array} \right\}$

Well, then, Annie might say, "Tom is married. He is not a bachelor." Then, Bob might say, "Tom is *not* married. He *is* a bachelor." (Remember that everything spoken aloud is added to the conversational record, CR.) So, now the conversational record contains two contradictions. That is, we have both $bachelor(\textsc{Tom})$ and $\neg bachelor(\textsc{Tom})$. Also, we have both $married(\textsc{Tom})$ and $\neg married(\textsc{Tom})$.

This second contradiction, however, has something of a different feeling than did the first. That is, in the first case, it seemed as though Annie and Bob agreed about the basic "fact" that Mark is married. The disagreement arose, it felt, from the fact they are using the word "bachelor" to "mean" different things. So, in that case, it seemed as though they could reconcile their communicatory problems by changing the "meanings" of words.

In this second case, it does not seem as though the problem is one of communication. It seems that there is a difference as to whether or not Mark is believed to be married. This is not an impediment to communication. The disagreement is precisely what Annie and Bob *are* communicating.

Obviously, this sort of talk has at least an inkling of the discussion of the analytic-synthetic distinction that was discussed at length in §5.2. That is, we are saying that a contradiction is tolerable, for the purposes of communication, when it arises from a difference in opinion about the "facts," rather than about word "meanings."

The fact that we would find ourselves in this situation is precisely the reason we found it necessary to discuss Quine's result at such length in §5.2. It certainly seems as though we are going to have to introduce some *kind* of notion of analycity. I say we will be after a "kind" of analytic-synthetic distinction because, as we have seen, there are different ways of casting such a distinction that are not equivalent.

I will propose we postulate what I will call a notion of *social analycity*. That is, a distinction between inferences that follow from the socially determined "meaning" of a word versus those that are considered to be "facts."

Interestingly, Asher and Lascarides (2003) also suggest, for different reasons, that some line should be drawn, for the purposes of analyzing communication, between aspects of a word that are *conventionalized* as part of its "meaning" versus those aspects that are not.

They suggest, for example, that it is a conventionalized fact about "eating" that "eating" takes an object. They contrast this to what they see as a contingent fact about eating that only animals with digestive tracts can eat. So, their distinction between analytic and synthetic information is somewhat different than what I am proposing. For them, conventionalized knowledge (which they equate with analytic knowledge, for better or worse) is used mainly for the purposes of parsing an utterance and a discourse. So, they are presenting a different concept of analycity (recall that I warned there were many concepts of analycity!), which I will not discuss further. Whatever one may think of this, I merely note here the idea that there is precedence among linguists for starting to suspect whether Quine's slogan should be stubbornly applied in all aspects of the theory of communication.

So, I suggest that Annie and Bob, in order to form a homogenous speech community, will need to agree that a "bachelor" is an "unmarried male" by virtue of the *conventionalized meaning* of that word, whatever that might mean. Suppose we identify the conventionalized meaning of the logical symbol α with the set of inferences about c that all speech community members agree to draw on the basis of learning $\alpha(c)$ or learning $\neg\alpha(c)$.

In such a case, conventionalized meanings would seem to say nothing about whether a predicate should apply to an *individual*. That is, nothing in the conventionalized meaning of $married$ would say whether or not the predicate should apply to TOM, although conventionalized meaning would say that if $married(c)$ then $\neg bachelor(c)$.

There might be all sorts of ways to implement conventionalized meaning, but in this and the next section, I will suggest that we will really only be interested in the specific case of definition.

That is, as in §5.2, let us introduce the notion of *definition* as a primitive term. That is, $=_{\mathrm{Def}}$ will be a symbol in LL, the language that Annie and Bob's (and everyone else in the speech community's)

thoughts are represented in. And, again, we tacitly understand that each **K** includes the (105), repeated here:

(127) $\forall\alpha,\beta,\big[\big(\alpha =_{\text{Def}} \beta\big) \rightarrow \big(\forall x\ \alpha(x) \leftrightarrow \beta(x)\big)\big]$ [15]

Then, what Annie and Bob (and other speech-community members) will need to agree on is to each believe a common set of statements of the form

$$\alpha =_{\text{Def}} \beta$$

Note that Annie and Bob can still disagree about "contingent" facts about "bachelors"—i.e. ones that do not follow from definitions—such as:

(128) $\forall x\ bachelor(x) \leftrightarrow male\text{-}who\text{-}goes\text{-}to\text{-}night\text{-}clubs(x)$

One fact which I have swept under the rug up to this point is that the cases for which we can strictly give necessary and sufficient conditions for class membership are limited. That is, it is rare that we can say that something is α if and only if it is β. "Bachelor" is famously one of the few examples, and even then many question whether there is nothing else to "bachelor" than "unmarried male," e.g., asking whether priests or infants should be called "bachelors."

To address this concern, I will briefly sketch how this discussion of definition can be expanded to accomodate more complex concepts. First, we can accomodate partial definitions by implicitly adding the following axiom to each **K**:

(129) $\forall\alpha,\beta\ \big(\alpha \rightarrow_{\text{Def}} \beta\big) \rightarrow \big(\forall x\ \alpha(x) \rightarrow \beta(x)\big)$

For example, we might not know exactly what constitutes an $animal$. But, we might be sure that an animal is something which can reproduce itself[17], and that this is an *inherent* part of animalhood. But, plants can reproduce as well, so we would not want to define "animals" as "reproducers of themselves." So, we use $\rightarrow_{\text{Def}}$:

---

[15] Recall that, what (127) says is that if,

$$bachelor =_{\text{Def}} \lambda x\big(\neg married(x) \wedge male(x)\big)$$

then,

$$\forall x\ bachelor(x) \leftrightarrow \big(\neg married(x) \wedge male(x)\big)$$ [16]

[17] Here, I mean that the animal can participate in reproduction, rather than be able to give birth itself. That is, males can "reproduce" themselves, in this sense.

(130) $animal \rightarrow_{\text{Def}} can\text{-}reproduce$

For further realism, one could expand on this notion of partial definitions with defeasible logic (e.g., Asher and Morreau 1995, generics (e.g., Carlson and Pelletier 1995) or prototype theory (e.g., Kamp and Partee 1995).

That is, one might point out that perhaps some animal will be unable to reproduce because it is injured, or something. In that case, we might want to say that a *normal* animal can reproduce, or else that an animal can *probably* reproduce. In this case, we can still make a distinction that properties that *normally* or *probably* follow by definition, versus those that normally or probably follow contingently.

5.4.2.4 Interpersonal Communication and Group Language (reprise)

Armed with this notion of analycity, as well as our justification of its coherence as a concept, let us go on to assume that, for a given community who speaks the dialect D, each speaker, i, of D shares a common set of beliefs. That is, there exists a set of statements, $\mathbf{K}_D$, such that for each i, i's knowledge, $\mathbf{K}_i$, includes $\mathbf{K}_D$ as a subset—i.e. $\forall i\ \mathbf{K}_D \subseteq \mathbf{K}_i$. Assuming that each $\mathbf{K}_i$ is consistent, a speaker of D can make any statement that is in or that follows from $\mathbf{K}_D$ and not expect to surprise or be surprised by any other speaker.

Clearly, $\mathbf{K}_D$ would place a lower bound on the number of shared inferences that follow from a statement. And, this lower bound on shared inferences will place an limit on the amount of contradictory inferences that will be drawn between hearer and speaker. In other words, $\mathbf{K}_D$ satisfies our desire to ensure (121) and (122).

Note that we are dealing now with a notion of what we might call *group language*. In other words, we have a notion that does not fit into the traditional Chomsksyan (cf. Chomsky 1986) distinction between *I-language*, which is the individual language internal to a single speaker, and *E-language*, which is a notion of language that exists apart from any language user.

D, and the corresponding set of sentences $\mathbf{K}_D$, are not internal to a single individual. One must look at the entire group of members in the community who speak D to ascertain what $\mathbf{K}_D$ consists of. But, at the same time, D is not a language that exists independently of language users, as it is tied to the community that speaks D.

So, assuming we want to pursue this notion of a common subset of all individual languages, i.e. $\mathbf{K}_D$, the following questions arise:

(131) What kind of statements are in $\mathbf{K}_D$?

(132) How is it determined which inferences will constitute $\mathbf{K}_D$?

(133) How is $\mathbf{K}_D$ communicated to new community members?

Well, though it would be an empirical question, let us just suppose that the answer to (131) is that:

(134) $\mathbf{K}_D$ can contain only definitions, i.e. statements of the form, $\alpha =_{\text{Def}} \beta$, or $\alpha \rightarrow_{\text{Def}} \beta$.

I want to just assume (134), and then give an answer to (132) which will, in turn, lend support to the credibility of (134). Let us assume that there is a set of predicates that get some significance independently of the logical system.

That is, it seems to me quite certain that there are predicates that are set via the visual and other sensory systems. For example, the predicate $green$ may have logical components, but, primarily, there is no logical means by which to discriminate $green$ from red—this is done by the visual system. A physicist could perhaps distinguish $green$ logically from red based on its wavelength. But, the ordinary person cannot do this and, I submit, the physicist can discuss the logical properties of $green$ and red only after his visual system has done the job of indicating to him which is which.

So, given some discourse referent, such as one for the wall, d_{wall}, whether or not $red(d_{wall})$ holds is something that is going to have to come from the outside the logical system.

As another example, let us consider what "lucky" might ultimately mean. Well, perhaps "lucky" means "one who is in a fortunate situation." Well, how does one identify a "fortunate situation"? Perhaps it is "a situation in which one is happy." And, how do we define "happy"? Well, primarily, "happy" is an identification of a feeling, and the ability to tell what feeling one is feeling is a sensory ability, like the ability to identify what color is in front of one, or whether one is hot or cold.

Now, someone of the perpetually doubtful sort might wonder how two people will come to agree that the word "happy" should be paired with the same feeling, when they cannot feel one another's feelings. Well, it only takes a moment's reflection about what goes on in the world to notice that happy experiences are often going to be shared and, in other cases, one can recognize by facial cues that another person is experiencing happiness. So, the mother might see the child displaying the symptoms of happiness and say, "You look happy." From this, the child will come to associate the word with the feeling.

What is the relevance of all this with regards to answering (132)? Well, suppose that Annie and Bob agree to the following:

(135) $lucky\text{-}to\text{-}be\text{-}unmarried =_{\text{Def}} \lambda x \left(\neg married(x) \wedge happy(x)\right)$

Then, suppose that, in Annie's opinion, all of the bachelors that she knows are not displaying the symptoms of happiness:

$$(136)\quad \mathbf{K}_A \supset \left\{ \begin{array}{c} \neg married(\textsc{Bill}) \\ male(\textsc{Bill}) \\ \neg happy(\textsc{Bill}) \\ \neg married(\textsc{Walter}) \\ male(\textsc{Walter}) \\ \neg happy(\textsc{Walter}) \end{array} \right\}$$

So:

$$(137)\quad Cn(\mathbf{K}_A) \supset \left\{ \begin{array}{c} \neg \textit{lucky-to-be-unmarried}(\textsc{Bill}) \\ \neg \textit{lucky-to-be-unmarried}(\textsc{Walter}) \end{array} \right\}$$

Also, suppose Bob has different opinion on the matter. That is, he feels that bachelors are, in general, the happy ones. And, moreover, he has noticed so many regularities about lucky bachelors that he is getting tired of saying "lucky bachelor" all the time. He would rather call them "lachelors," for short. So, Bob explains to Annie that a *lachelor* is an unmarried male who is lucky to be unmarried:

$$lachelor =_{\text{Def}} \lambda x \left(\begin{array}{c} \neg married(x) \wedge male(x) \wedge \\ lucky\text{-}to\text{-}be\text{-}unmarried(x) \end{array} \right)$$

Well, Annie is not going to be able to use this word, since, for her, it will not apply to anyone. That is, if she is obeying Grice's (1975) maxim, "Be relevant," then there is never going to be a time when Annie is going to say, e.g., "Look, there's a lachelor," or, "Hey, have you met Rob? He's a lachelor." So, "lachelor," as a word, is not going to get off the ground if there are many speakers like Annie.

The moral of this story is that, if a sentence like $\alpha =_{\text{Def}} \beta$, where α is a single predicate and β is a conjunction of other predicates, is going to enter into $\mathbf{K}_D$, then the cluster of predicates that β represents is going to have to be something that many people in the community will want to talk about. Thus, only properties that cluster together very tightly, as observed by everyone, are likely to be given their own symbols in $\mathbf{K}_D$.

Assuming the question as to how $\mathbf{K}_D$ can be negotiated has been settled, we now turn to (133), the question as to how newcomers to the linguistic-social community will learn $\mathbf{K}_D$.

The first thing to notice is that "is," which is already in several ways ambiguous, is now going to be ambiguous in more ways. That is, in addition to the following:

(138) a. Bill is Sammy's brother. (identity)

b. BILL = $brother\text{-}of(\text{SAMMY})$

(139) a. Bill is tall. (predication for an individual)

b. $tall(\text{BILL})$

(140) a. Bachelors are messy. (predication using quantification)[18]

b. $\forall x\ bachelor(x) \rightarrow messy(x)$

Now, "is" will be ambiguous in at least two other ways, having the following interpretations as well:

(141) a. Bachelors are unmarried males. (definitional equality)

b. $bachelor =_{\text{Def}} \lambda x \left(\neg married(x) \wedge male(x)\right)$

(142) a. Men are humans. (definitional implication)

b. $man \rightarrow_{\text{Def}} human$

Asking how "is" can be resolved in these different ways is essentially the question as to how explicature is resolved and, of course, we dealt with this question in great detail in §3–4. There we saw that, in general, the context-sensitive process by which explicature is resolved is a difficult one to model, but there is plenty of evidence that it is a general process, and that inroads can be made.

So, clearly the transmission of $\mathbf{K}_D$ to a newcomer will involve that newcomer explicating certain utterances as conveying a $=_{\text{Def}}$ or a $\rightarrow_{\text{Def}}$ relation. Now, there are basically two kinds of newcomers: toddlers (i.e. first-language learners) and foreign (i.e. second-) language learners.

In either case, the newcomer is typically motivated to assimilate their language usage to that of the natives, without much thought. The toddler is typically unaware of the writings of Chomsky or Quine. They are unaware of arguments about the philosophical legitimacy of their deviant individual-languages. They are rewarded for using words like those around them and discouraged from using words differently. They like rewards. They want to use words to get things from their parents, to gain favor with people and, later, to forge relationships. So, they internalize the statements they are being given and, in particular, those explicated as $=_{\text{Def}}$- and $\rightarrow_{\text{Def}}$-sentences.

The second-language learner will typically make it a point to internalize $\mathbf{K}_D$ as quickly as possible. One only has to look to see that it is a matter of pride for most of them just how quickly they

[18] Some, like Montague (1974), analyze the English sentences in (139) and (140) as both involving predication with quantification. Either way, "is" is at least in two ways ambiguous.

can do this. They will be responsive to criticisms that they have used words incorrectly and will seek out the correct usage.

So, in this way, once a $\mathbf{K}_D$ is in existence, it will remain so via a constant propagation to newcomers, and by continuing to allow profitable communication for those who have long ago internalized it.

5.4.2.5 Conclusion

In this section, we have seen how it can be that speakers in a community who speaks D can communicate, despite the fact that no two speakers, such as Annie and Bob, will have exactly the same knowledge set. I proposed that they can communicate because they internalize a set of common inferences, $\mathbf{K}_D$. I have conjectured that $\mathbf{K}_D$ might include only definitions, i.e., only relations of the form $=_{\text{Def}}$ and $\rightarrow_{\text{Def}}$. $\mathbf{K}_D$ gives each symbol therein a *conventionalized meaning*.

5.4.3 Scrutinizing Our Assumptions

In the previous section, I made some simplifying assumptions, which were allowed to pass without much comment, but which I think it would be wise to revisit.

The first, to be discussed in §5.4.3.1 is the assumption that we can essentially assume that, for example, Annie has a predicate $bachelor$ and Bob has a predicate by the same name, i.e. $bachelor$, and that, when Annie means to communicate a thought involving the predicate $bachelor$, Bob understands that Annie has communicated a thought involving the predicate $bachelor$. And, this $bachelor$ predicate is not ambiguous, the way that the natural language word "bachelor" might be[19]. I refer to this as the assumption that community members "speak in LL."

The second, to be discussed in §5.4.3.2, is that, when two people are talking and they notice some intolerable contradiction, they will try to sort the matter out, so that one of them will change their language usage in a way that removes that contradiction. This would be as opposed to, for example, Annie and Bob noticing that they use words in crucially different ways, and, rather than trying to harmonize their usages, trying to remember how one another uses words.

[19]The natural language word "bachelor" is ambiguous between meaning, for example, "unmarried male," on the one hand, and, "a young knight in the service of another during feudal times," on the other. I would presume that these two natural language uses of the word "bachelor" would translate to LL as different predicates, such as $bachelor_1$ and $bachelor_2$. So, it is a very strong assumption that if Annie means to communicate the logical predicate $bachelor$, this is the predicate that Bob understands.

5.4.3.1 Language Members are Effectively Speaking in LL

Taking Annie and Bob as our usual stand-ins for any given members of some given linguistic community, I have assumed in §5.4.2 that the predicates mentioned in Annie's knowledge, $\mathbf{K}_A$, and those mentioned in Bob's knowledge, $\mathbf{K}_B$, are the same. And, I further assumed that if Annie meant to communicate the thought $bachelor(\text{TOM})$, then Bob would understand Annie's communication as $bachelor(\text{TOM})$ exactly. Thus, I have assumed the following:

(143) **Speakers effectively talk in LL**: When Alice wants to communicate the LL statement α, and Bob understands this as the LL statement β, then $\alpha = \beta$. That is, the syntax of β is the same as α, and the symbols used are identical. This is referred to as "talking directly in LL."

It is arguably not clear why this should be a safe assumption. For example, if Annie were to try to communicate $bachelor(\text{TOM})$ and Bob were to parse her communication as $tall(\text{TOM})$, then there would not be a means to notice that a contradiction had been drawn.

This is a concern in this framework, while perhaps it has not been in others, because the system I have been describing *crucially* relies on negotiation about word-meaning in order to work. To see what could go wrong, suppose Annie were to try to communicate $bachelor(\text{TOM})$ to Bob. And Bob, who actually believed $\neg bachelor(\text{TOM})$, were to parse "Tom is a bachelor" as $tall(\text{TOM})$. Then, if Bob *did* believe that $tall(\text{TOM})$, then, because there are so many discrepancies between them, they would never notice their discrepancy about Tom's bachelorhood.

The reason that one cannot trivially assume that Annie's intention to communicate the logical idea $\alpha = bachelor(\text{TOM})$ will be matched by Bob's parsing her communication as β, where $\beta = \alpha$, is the following. Consider that when we write out a word in English, like "bachelor," we intend this to stand in for a sequence of phonemes, or perhaps phones or sounds.

That is, "bachelor" in this context is not an arbitrary name but it is also a *description* of a sound pattern. The "bachelor" that Annie says is something public. That is, we assume without qualm that the "bachelor" that Annie hears when she speaks the word is the same "bachelor" that Bob hears.

So, while the connection between the word "bachelor" and the idea of an unmarried male is arbitrary, as per Saussure, the connection between "bachelor" and the *sequence of phonemes* is not arbitrary because "bachelor" describes that sequence. (Well, as it happens, English spelling does not directly reflect the sounds that constitute a word. However, if we were willing to give up convenience, we could use phonemic spelling instead of English spelling.)

In contrast, consider that, our theory has a "level" of representation, called LL, in which we find forms like $bachelor(\text{TOM})$ and $\forall x\ bachelor(x) \rightarrow \neg married(x)$. These serve as the representation of "thoughts." But, we presume that, in the brain, this is all represented in neurons in a way that is not yet understood.

But, whatever configuration of neurons would represent the predicate I call $bachelor$, it is quite clear that $bachelor$ is not a structural description of that neural configuration, the same way that "bachelor" is a description of a sound pattern. In other words, $bachelor$, is an arbitrary name for a neural configuration representing a logical predicate, and instead of $bachelor$, we could just as well have used p_1, p_2 or $p_{aDK0123qX}$ instead.

The point is that it would be a mistake to think that that neural configuration in Annie's mind that represents $bachelor$ has any direct connection to a neural configuration in Bob's mind, simply because I have written the same symbol as a stand-in for both.

And, because of ambiguity, there is going to be more than one LL predicate that corresponds, in some cases, to a sequence of phonemes. That is, there are going to be two predicates, say $cool_1$ and $cool_2$ that the sequence "cool" must be mapped to in the following:

(144) a. Bob Marley is so cool.

b. The air is so cool today.

So, one cannot simply assume that each surface predicate is mapped to a single logical predicate.

And, there is a question that arises, in particular, in a holistic framework, which is this. If a child, like Little Annie, is trying to figure out which predicate to map "bachelor" to, maybe she will map "bachelor" to $bachelor$ and define $bachelor$ traditionally via,

$$\forall x\ bachelor(x) \leftrightarrow \big(\neg married(x) \wedge male(x)\big)$$

But, maybe she will translate "bachelor" as $chicken$ and then define $chicken$ as,

$$\forall x\ chicken(x) \leftrightarrow \big(\neg married(x) \wedge male(x)\big)$$

So, while this might (and, in fact, will) all turn out to be unproblematic, we at least need to do some work to convince ourselves that this sort of problem does not threaten the discussion of the previous section.

In order to justify the assumption (143), I am going to show that talking as though Annie and Bob are speaking directly in LL can be reduced to what would seem to be a much weaker and more

reasonable assumption, (146). (146) has a somewhat technical statement and I think it is more naturally introduced after some discussion.

After showing that the discussion of the last section effectively only need to assume (146), I will argue that, for the purposes of explaining how the *semantic* aspect of dialect comes to be negotiated, (146) is a reasonable assumption.

First, let us reduce (143) to (145) below. This reduction will be rather trivial, and is only an intermediate step in the proof, which allows for easier exposition. Let us suppose that two speakers, say Annie and Bob, are going to share their evaluation of the pragmatic context, whatever that may mean. That is, both will feel that the context is C. Also, Annie and Bob both speak the same natural language, NL, and each thinks in the same logical language, LL.

Now, relative to the context, each will have two functions. First, there is the speaking function, f_s, which is a function from LL into NL. Second, there is the parsing function, f_p, which is a function from NL into LL. These are inverses so that

$$f_p(f_s(l)) = l$$

And,

$$f_s(f_p(u)) = u$$

It should be understood that, in all cases, each f_s and f_p referred to is meant to be relative to (i.e. a function of) whatever the context is at the time even though I will not say so each time. This is how we will side-step the problems discussed at length in §3, in which I argued that there is not a context-insensitive mapping from NL to LL[20].

Now:

(145) **Identicality of Parsing**: Annie and Bob are each using the same f_s as their speaking function and f_p as their parsing function.

It is pretty easy to see that (145) is equivalent to our original (143).

That is, suppose Annie wants to communicate the idea $bachelor(\text{TOM})$. She is going to compute $f_s(bachelor(\text{TOM}))$, which will have the value "Tom is a bachelor." Then she will say this out loud, which Bob will hear. Bob will compute $f_p($ "Tom is a bachelor"$)$, which has the value $bachelor(\text{TOM})$.

[20] And, recall that I argued in that section that the mapping from NL to LL sentence is not foolproof. That is, the hearer may attribute a different explicature to the speaker than the speaker intended. We are ignoring this fact here, as an idealization.

So, given these assumptions, Annie and Bob might as well be speaking in LL itself. But, with (145), we are essentially still assuming that if Annie maps "bachelor" to $bachelor$ in sentence ϕ in context C, then Bob does as well, which is precisely the suspiciously strong assumption we were worried about.

But, we are not done here. The reduction of (143) to (145) is merely an intermediate step for expository reasons. We are now going to reduce (145) to an assumption which is, in fact, easily argued to be more reasonable.

So, let us assume that Annie thinks in the language LL^{A} and Bob thinks in a potentially different language LL^{B}, but that both speak the natural language NL. Annie has her own speaking and a parsing functions, say ${f_s}^A$ and ${f_p}^A$, and Bob has his own functions, say ${f_s}^B$ and ${f_p}^B$. That is, we will neither assume ${f_s}^A = {f_s}^B$ nor ${f_p}^A = {f_p}^B$. But, we will still assume that,

$${f_p}^A({f_s}^A(l)) = l_a$$

where $l_a \in \text{LL}^{\text{A}}$, and that,

$${f_p}^B({f_s}^B(l)) = l_b$$

where $l_b \in \text{LL}^{\text{B}}$. That is, we will still assume that if, e.g., Annie were to express $tall(\text{JIM})$ as "Jim is tall" in context C, then she would parse "Jim is tall" as $tall(\text{JIM})$ in context C. In general, if Annie were to express some LL form, $\ulcorner u$ as U in context C, then she would parse U as $\ulcorner u$ in context C, and we make the likewise assumption for Bob.

Now, suppose n_1 and n_2 are two statements in NL, such that ${f_p}^A(n_1) = {l_1}^A$, ${f_p}^A(n_2) = {l_2}^A$. What this says is that Annie will parse n_1 as ${l_1}^A$ and she will parse n_2 as ${l_2}^A$. Here, ${l_1}^A$ and ${l_2}^A$ are sentences in LL, the logical language. That is, ${l_1}^A$ and ${l_2}^A$ are *strings*, which are, in turn, sequences of symbols. Now, suppose that ${l_1}^A$ and ${l_2}^A$ both have a common symbol, say a.

Suppose that, in ${l_1}^A$, a is the i_1'th symbol and, in ${l_2}^A$, a is the i_2'th symbol. Furthermore, suppose that Bob (using ${f_p}^B$) parses n_1 and n_2 as ${l_1}^B$ and ${l_2}^B$ respectively. Then what we will assume instead of (145), is the following, which will be followed up quickly with an illustrative example:

(146) **Near Identicality of Parsing**: Suppose a is the common symbol between ${l_1}^A$ and ${l_2}^A$, and suppose that a is the i_1'th symbol in ${l_1}^A$, and a is the i_2'th symbol in ${l_2}^A$. Then, the i_1'th symbol in ${l_1}^B$ is identical to the i_2'th symbol in ${l_2}^B$.

Now, what (146) is saying is this. Suppose Annie parses "Tom is an American" as the LL form $american(\text{TOM})$ and that she parses "All Americans are humans" as the LL form $\forall x\, american(x) \rightarrow human(x)$. Now, Bob might parse these two sentences as any of the following:

(147) a. $banana(\text{TOM})$ and $\forall x\, banana(x) \rightarrow human(x)$

b. $saucer(\text{TOM})$ and $\forall x\, saucer(x) \rightarrow human(x)$

c. $pencil(\text{TOM})$ and $\forall x\, pencil(x) \rightarrow human(x)$

That is, because, in parsing these two sentences, Annie has used the symbol $american$ twice, (146) says that it must also be that Bob will use whatever symbol he will instead of $american$ in one place in *all of the same places* that Annie had used $american$.

In other words, (146) says that Bob *cannot* translate "Tom is an American" as $banana(\text{TOM})$ and then translate "All Americans are humans" as $\forall x\, saucer(x) \rightarrow human(x)$ because (146) says that if Bob uses $banana$ where Alice has used $american$ once, then he must do so all the time.

Now, suppose that it was in fact $banana$ that Bob's parser had used instead of American, as we see in (147a). Then, we could create a new $\mathbf{K}_B$ for Bob, $\mathbf{K}_B{}'$ in which wherever $\mathbf{K}_B$ had used the symbol $american$, $\mathbf{K}_B{}'$ would use the symbol $temporary$. And, wherever $\mathbf{K}_B$ had used the symbol $banana$, we would instead use the symbol $american$. We would then have to update Bob's speaking and parsing functions so that now $american$ (instead of $banana$) surfaces as "American," and so that $temporary$ surfaces as whatever $american$ used to surface as in ${f_s}^B$.

If we continued on like this, modifying Bob's knowledge and his speaking and parsing functions, then we would eventually arrive at a parsing function for Bob, ${f_p}^{B*}$, and a speaking function for Bob, ${f_s}^{B*}$, such that,

$${f_p}^{B*} = {f_p}^A$$

and,

$${f_s}^{B*} = {f_s}^A$$

In other words, we could massage Bob's knowledge and his speaking and parsing functions so that he would end up with identical speaking and parsing functions as Annie, which is the assumptions stated in (145).

Note that Bob's performance will not change throughout this sequence of changes to his knowledge and his speaking and parsing functions. That is, if Bob had begun with the LL belief $banana(\text{TOM})$,

which he would have verbalized as, "Tom is American," he would have finished with the LL belief $american(\text{TOM})$, which he would have *still* verbalized as, "Tom is American."

Furthermore, because he initially believed $banana(\text{TOM})$, if Annie had told him, "Tom is not an American," he would have translated this as $\neg banana(\text{TOM})$ and then complained about a contradiction. Well, once $american$ had been swapped in for $banana$ (and $temporary$ was swapped in for $american$), Bob would translate "Tom is not an American" as $\neg american(\text{TOM})$, which he *still* would have complained as contradicting his beliefs, because he would later, correspondingly, hold that $american(\text{TOM})$ (i.e. instead of holding that $banana(\text{TOM})$).

What all this has gone to show is that (145) is no stronger than (146). Since (143) is no stronger than (145), we have that (143) is no stronger than (146). Thus, if one is willing to assume (146), then our discussion of section §5.4 will go through.

But, what is (146) saying? Well, (146) effectively lets us assume the following. For each constant symbol (whether it denotes a predicate name or an individual), $\rho^A \in \text{LL}^\text{A}$, there is a corresponding predicate name, $\rho^B \in \text{LL}^\text{B}$. Now, suppose we have some sentence, $n \in \text{NL}$, and we know that Annie translates n as l_a. Then, we can get Bob's translation of n by simply taking Annie's translation, l_a, and replacing in l_a each occurrence of ρ^A by ρ^B.

For example, if Annie parses, "All bachelors are messy," as

$$\forall x \ \ bachelor^A(x) \rightarrow messy^A(x)$$

then Bob must parse it as

$$\forall x \ \ bachelor^B(x) \rightarrow messy^B(x)$$

In other words, Annie's parse and Bob's parse are effectively identical, except that each one's language of thought is allowed to contain different predicate names, so long as there is a "one-to-one correspondence" between the predicate names of each.

Now, (146) is much less theoretically offensive, I think, than (143). It does not require that Annie and Bob have the same logical predicate names, which was the crucial problem with assuming that speakers talk in LL that I highlighted.

However, it does assume that, aside from their choice of predicate names, Annie and Bob have identical parsers, and so, effectively identical grammars. This is still a strong idealization. First of all, different speakers might have subtle differences in their grammars. I would argue, though, that subtle differences in the grammars of mature speakers may slightly complicate the picture, but only in the minority of cases.

To see a second reason that assuming (146) is a strong assumption, suppose that Bob is a newcomer to the linguistic community that speaks NL, and Annie is a native speaker. Bob is not, initially, going to have the same grammar as Annie. But, I submit, there are many more words (many thousands) to be learned by a newcomer than grammatical constructions.

So, while there may be discrepancies between a model assuming (146), and reality, for each grammatical construction, g, being acquired by a newcomer, while they are still acquiring g, once g is learned, we can subsequently treat (146) as true, with respect to g itself, even if we cannot assume (146) for the whole language.

Thus, I feel that we are justified in having the discussion of §5.4.2 under the idealization that speaker's can communicate directly in a single LL.

5.4.3.2 The Principle of Charity?

Another concern that might arise could stem from the notion of a Wilson (1959)-Quine (1960) principle of charity. Quine made such a view famous by suggesting that, when a linguistic explorer goes into a new community and is trying to learn the language from the native speaker, the explorer should adjust his translations so that claims made by the native that sound silly are true. In other words, the explorer should assume that the native speaker is sensible, and that insensible apparently insensible statements evince poor translation[21]. That is, if one figures that the native speaker has said, "The sun comes out at night," which would be patently absurd, the translator should assume that something is wrong with his translation scheme.

Now, in §5.4.2, I assumed that if Alice and Bob realized that they had some "intolerable" contradiction in terms of how they used words, one of them would change their usage. That is, Alice and Bob would consider themselves part of a single speech community, so that if one defined "bachelor" as "unmarried male," and the other defined it as "married male," and the two figured that this was a difference which would impede conversation, then at least one of them would have to change their usage of "bachelor."

But, if say Annie were following Quine's advice, after hearing Bob say, "Bachelors are married males," Annie might take Bob's "married" to be her "unmarried," so that Bob's sentence would come out true. In this case, all contradictions would proliferate, and there would be no way to tell what anyone was saying.

[21]Cf., "assertions startlingly false on the face of them are likely to turn on hidden differences of language," (Quine 1960, p. 58).

Well, I submit that Annie and Bob, given some reason to consider themselves members of a single speech community would not follow Quine's advice for linguistic explorers and would, instead, reconcile any contradictions that they found between themselves.

This is ultimately an empirical question, and should, if we are seriously concerned, be tested via experiment. To test the matter via thought-experiment, suppose I were to go up to some rough-looking body builder with snake tattoos all over his arm who is clearly drunk at a shady bar and say to him:

(148) Your mother is of ill-repute.

The Quinean would suggest that this burly fellow, being quite certain that his mother is not of ill-repute, and being further quite certain that I am a sensible fellow who would recognize this, would interpret *my* (148) perhaps as his, "Your mother is *not* of ill-repute," or, perhaps, "Your mother is a fine lady."

I propose I would be punched. What this would show is that the tattooed man is interpreting my statement by *his own* rules, and treating me as a member of his own dialect, whose words should not be translated, but merely parsed according to his own rules.

There probably are times, such as when a person is in a new area where everyone seems to be speaking differently, that one might consider the possibility that they had understood something different than was intended. But, it would seem to me to be a fact that, in most cases, people treat other people who are ostensibly speaking the same language, as though they are *actually* speaking the same language and will react to such things that contradict their own beliefs, such when another reacts to, "Your mother is of ill-repute," with a punch.

Thus, I think we were justified in assuming in §5.4.2, that Annie and Bob would explore and try to resolve intolerable contradictions.

5.5 Answering Fodor and Lepore's Criticisms of Meaning Holism

Fodor and Lepore (1991, 1992), troubled what they saw as the overly hasty rise of holism, published a series of works cautioning against the holistic approach. Many of their arguments, according to Block (1993), are recycled. But, Fodor and Lepore claim to have an original argument against holism that everyone should know about. This argument is based on what they see as an inconsistent triad created by the principle of compositionality, the rejection of the analytic-synthetic distinction, and meaning holism, as they conceive of these terms.

Now, while I think that the criticisms listed in §5.1.2 were worthwhile criticisms whose resolution had led to new insights about language, Fodor and Lepore's argument is largely based on a series of confusions. But, I think it is still worthwhile to review it, first because it has been the topic of some debate, and second, because there may be some interesting notes brought up in the process. Furthermore, I would like to examine the reply by Block (1993) to Fodor and Lepore because Block's reply rests on the two-factor theory discussed in §5.4.1.1, which I argued is not tenable.

Fodor and Lepore's argument runs like this. Suppose we want to identify "meaning" with inferential role:

(149) "Meaning" is "inferential role"

And, suppose we want to adopt a compositionality thesis. This, say Fodor and Lepore, is the only account we have of how people can learn an infinite language on the basis of a finite description. So, we adopt the following assumption:

(150) The "meaning" of the whole is a function of the "meanings" of the parts and their mode of combination.

What they seem to have in mind in claiming (150) is that if something is a "brown cow," then that thing must be both "brown" and a "cow." And, whatever inferences are licensed by $brown(c)$ should also be licensed by $(brown(cow))(c)$. And, whatever inferences are licensed by $cow(c)$ should also be licensed by $(brown(cow))(c)$.

Now, the problem comes if someone believes that, "Brown cows are dangerous." In this case, $\forall x\ (brown(cow)) \rightarrow dangerous(x)$ does not follow from the inferential roles of $brown$ and cow. Thus, the "meaning" of the whole is not a function of only the "meanings" of the parts, as Fodor and Lepore conceive of things.

Now, say Fodor and Lepore, the person who wants to save something like (149) and also maintain (150) will have to adopt the following course. Identify "meaning" with the "analytic" aspect of inferential role:

(151) "Meaning" is "the analytic aspect of inferential role"

And, the analytic aspect of a compound's inferential role would have to be those inferences that follow from compositional rules.

Then $\forall x\ (brown(cow))(x) \rightarrow brown(x)$ is true by the rules of the compositional theory presumed and those inferences warranted by the compositional rules are analytic. Thus, the analytic

part of inference *is* compositional. This is circular but that is alright because the system hangs together and does what was asked of it. (It is not as though we proved some assumption on the basis of itself, which would be an inadmissible circularity.)

But, then, of course, say Fodor and Lepore, (151) runs up against (152). That is Fodor and Lepore note that most feel that Quine (1951) argued decisively that there can be no analytic-synthetic distinction.

(152) There is no principled analytic-synthetic distinction.

Of course, this matter was discussed at great length in §5.2, where it was argued that one *could*, in fact, draw a kind of analytic-synthetic distinction based on definition. But, actually, my rebuttal of Fodor and Lepore will not refer to that argument, in order to better illustrate in how many ways their argument is faulty.

There are two problems with Fodor and Lepore's argument that I will focus on. The first is that they are in several ways confused about the way that compositionality works. The second is that they are carrying the word "analytic" around from context to context and are insensitive to the fact that the definition of the term is changing. They derive a conclusion using one definition of analytic, and then evaluate it using another, finding absurdity in tautology, as we will see.

Beginning with compositionality, recall that I discussed the matter in detail in §3.2. In that section, we saw that if "meaning" was identified with model-theoretic truth conditions, then the compositionality thesis was clearly false if we are talking about the "meaning" of a natural language utterance. Recall we have considered such examples as:

(153) I was there.

(154) Sam is not tall enough [to play on this basketball team].

We also saw that if the compositionality thesis was meant to apply to LL, the logical language which natural language utterances are translated into for processing, it was basically true by definition. That is, LL is a context-free language, of essentially the sort pioneered by Frege, chosen to be amenable to the giving of explicit rules of inference and a Tarski-style truth predicate. And, we saw that essentially the definition of a Tarski-style truth definition *is* a compositional one.

To recap, then, the compositionality thesis clearly does not apply to the surface level of language[22], but does, basically by definition, apply to the logical language our theory says that utterances are translated into for inferential processing. The first problem that Fodor and Lepore make is failing to distinguish between the two.

Now, constraining attention to the domain in which the compositionality thesis does apply, i.e. LL, what do we find? What Fodor and Lepore argue is that, because compositionality is required to explain how we can make infinite use of a finite basis for language, there cannot be inferences which do not arise from the "meanings" (i.e. inferential roles) of the parts. This is incorrect.

What we require to show that infinite use can be made of finite means in language is that *some* of the inferences for complex expressions follow from the meaning of their parts. To see what I mean, consider that one might know that brown things are colored in such a way that they resemble dark yellow things. And, one might know that cows give milk that people can drink. That is:

(155) $\forall x \;\; brown(x) \rightarrow resembles\text{-}yellow(x)$

(156) $\forall x \;\; cow(x) \rightarrow gives\text{-}milk(x)$

Then, if we assume that,

$$\forall x \;\; (brown(cow))(x) \rightarrow \big(brown(x) \wedge cow(x)\big)$$

it will follow that:

(157) $$\begin{aligned} &\forall x \;\; (brown(cow))(x) \rightarrow \\ &\quad \big(resembles\text{-}yellow(x) \wedge gives\text{-}milk(x)\big) \end{aligned}$$

In other words, the facts that brown cows are things that resemble yellow in color, and things that give milk, follow from the compositional organization of this system.

Now, suppose we were to also add to **K** the statement:

(158) $\forall x \;\; (brown(cow)) \rightarrow dangerous(x)$

Well, (158) does not follow by any compositional rules. But, it also does not block the other inferences, such as (157), that *do* follow from such rules.

That is, Fodor and Lepore are proposing that a person's knowledge should not contain a statement like (158). But, if one adds (158) to some **K**, to yield some **K**', then that speaker is only going to be able to draw *more* inferences on the basis of **K**' than they could from **K** alone. So, if they

[22]This is why we were required in §5.4.3.1 to specify that each parsing and speaking function was relative to the context.

were able to understand an infinite range of sentences on the basis of **K**, they will still be able to understand an infinite range of sentences on the basis of **K**'.

Furthermore, suppose one knows that "funny" things make people laugh (159), and that $funny$ is an intersective adjective (160):

(159) $\forall x \;\; funny(x) \rightarrow \textit{make-people-laugh}(x)$

(160) $\forall x \;\; (funny(N))(x) \rightarrow \big(funny(x) \wedge N(x)\big)$

Then suppose one considers a "funny brown cow." They would then conclude that,

$$\forall x \;\; funny((brown(cow)))(x) \rightarrow \big(funny(x) \wedge (brown(cow))(x)\big)$$

and, so,

$$\forall x \left[\begin{array}{l} funny((brown(cow)))(x) \rightarrow \\ \qquad \left(\begin{array}{c} \textit{make-people-laugh}(x) \wedge \textit{resembles-yellow}(x) \wedge \textit{gives-milk}(x) \\ dangerous(x) \end{array} \right) \end{array} \right]$$

Note that, now, the idiosyncratic fact about brown cows—i.e. that they are dangerous—which did not follow from compositional rules is now *participating* in compositional rules. That is, by virtue of the fact that brown cows are dangerous, and $funny$ is intersective, we know that "funny brown cows" are "brown cows" and so dangerous.

The moral of the story is that the ability to draw inferences about complex predicates, on the basis of knowledge about the parts can work *in addition* to the ability to learn information about complex predicates that does not follow compositionally. One does not preclude the other.

The removal of (150)—in the way that Fodor and Lepore construe it—from their list of assumptions immediately destroys their argument. We could stop here. But, let us go on.

The second way that Fodor and Lepore get themselves confused is, as I said, that they use the word "analytic" in several different ways without realizing it. That is, they define analycity as follows: "for an inference to be analytic is just for it to be warranted by the meanings of its constituent expressions," (Fodor and Lepore 1991, p. 336). But, recall that we are identifying "meaning" with inferential role.

So, what this amounts to is that if one of the inferences that a person believes is,

$$\forall x \; (brown(cow))(x) \rightarrow dangerous(x)$$

then, "Brown cows are dangerous," is analytic, because one can infer it on the basis of the inferential roles of one's beliefs.

But, on this interpretation of analycity, *every statement* in some person's **K** is going to be analytic. This is because every inference that one can make on the basis of $\alpha(c)$ is going to be an inference licensed by the inferential role of α in **K**. So, if one's definition of an analytic inference about α is one that is licensed by inferential role, then all inferences are analytic. This is a tautology. The only reason it would have any shock value is that one is using a *different* (traditional) definition of analycity to evaluate the conclusion than we used to arrive at it.

And, indeed, shock value is all that Fodor and Lepore are ultimately trading on. Block (1993) puts it perfectly as he recounts Fodor and Lepore's "argument" against the idea that all statements would be analytic:

> The sum total of [Fodor and Lepore's] argument in (1992) is that this idea [i.e. that all statements are analytic] is 'preposterous on the face of it' (p. 164), 'patently preposterous' (p. 174), that it is not 'possible to take seriously' (p. 174), that surely this is preposterous' (p. 182), that is 'perfectly mad' (p. 182), and 'incapable of being taken seriously' (p. 183), and that it is . . . an option which Quine, quite sensibly, didn't even bother to consider. . . (p. 183), (Block 1993, p. 8)

Again, the only reason that this *tautology* sounds "preposterous" is that Fodor and Lepore are mixed up as to how they are defining their terms.

Before moving on from this point, we might briefly consider Block's reply to the charge that all inferences are analytic is preposterous, if only because I have quoted his reply to the charge for its aptness. For reasons internal to his two-factor theory (cf. §5.4.1.1), Block (1993) feels he cannot let it stand that all statements are analytic.

His response is to argue, via Putnam-style Twin Earth thought experiments, that his inferential roles—which he calls "narrow contents," as opposed to "wide contents," which involve the "world of non-symbols," cf. §5.4.1.1—can neither be true nor false and so cannot be analytic. Here, he is defining analycity as follows: "[a]nalytic truths are true in virtue of meaning," (1993, p.18). First of all, it seems Block is himself caught in the snare of multiple definitions of analycity.

Second, it is fortunate that we have rebutted Fodor and Lepore without recourse to this sort of two-factor theory because, as I argued in §5.4.1.1, the two-factor theory is of dubious merit, effectively grafting an externalist theory onto an internalist in a way that is unspecified and which predicts nothing.

In conclusion, though a minor fuss has been made about Fodor and Lepore's criticisms of meaning holism, there is ultimately not much to them, besides confusion.

5.6 Conclusion

In §5.1, I explained what a holistic model of meaning is. And, I recounted some of the apparent difficulties for this kind of theory. The chief difficulty is to explain how interpersonal communication can succeed if the inferential role for each word in a person's language is idiosyncratic to the person themself. I then explained why the model of comprehension described in §4 is one that would be called "holistic."

In §5.2, I reviewed Quine's (1951) famous arguments against an analytic-synthetic distinction, with an eye towards appealing to some notion of analycity later in the chapter.

In §5.3, we saw that a holistic theory affords a straightforward and elegant explanation as to how it is that language can be represented as well as how it can be learned incrementally. It does so in a way that does not require conceptual primitives, which we saw are problematic.

In §5.4, I reviewed and criticized some proposed solutions to the problems of meaning holism that can be found in the literature. I then showed how a notion of conventionalized word-meanings could be used to enforce sufficient agreement about the inferential roles of words in the idiolects within a speech community, so that communication could be guaranteed to succeed. In formulating this notion of conventionalized word-meaning, we appealed to the evaluation of Quine's remarks reviewed in §5.2.

In §5.5, I rebutted some particular arguments against holism due to Fodor and Lepore, which have gained particular attention as of late.

Ultimately, I think one has to conclude that a holistic model of the knowledge of the predicates in a person's language is viable. And, this is fortunate, for the holistic model of knowledge presented in §4.1 was seen, therein, to be a simple but powerful one.

Chapter 6

Conclusion

6.1 Summary

We began this inquiry with the goal of determining what the overall structure of a theory of comprehension should look like. This was an open-ended question. We had to figure out both what the theory should do and how it should do it.

§2 was spent scrutinizing compositional and model-theoretic semantics. In particular, I decided that giving model-theoretic truth-conditions for a natural language is basically just translating that natural language sentence into a logical language. Calling the target language for this tranlsation project the "meta-language" or the "truth-conditions" of the sentence does not lift it above the realm of tranlsation. It does not illuminate any connection to the world of non-symbols.

So, I decided that our theory should try to use the manipulation of representation in a way that could be said to constitute something we would call a process of "comprehension."

We also saw that translation does not, as Lewis was ironically always pointing out, constitute, in and of itself, a meaningful semantic theory. One reason for this, as Thomason acknowledged, is that compositional semantics does not give an acount of the "meanings" of the individual words that get combined in compositional semantics.

§3 highlighted the fact that the speaker's intended "full propositional form," or explicature, cannot be read right off of the acoustic signal by rules insensitive to the speaker's word-/world-knowledge. The creation of the explicature would involve a guess as to what the speaker had intended to communicate, both "literally" and via "implicature."

In sum, we decided that we would want a model to do the following:

- Explain "comprehension" in terms of the manipulation of representation.
- Explain how different words of the same grammatical category make different contributions to a sentence.
- Explain how the full propositional form communicated on an underspecified acoustic signal is recovered.

In §4 we saw a model that did, I argued, all of that. Understanding a sentence was equated with drawing the inferences that followed from it, in conjunction with word-/world-knowledge. But, in order to draw inferences, the logical form from which to infer would first have to be guessed at.

It might have seemed overly ambitious to attempt a model of comprehension in a work of this length, but I think we have seen that it was, to a large extent, possible.

In §5, I explained that certain problems awaited this model of comprehension. That is, because a hearer's comprehension of a sentence is relative to his own idiosyncratic word-/world-knowledge, we would need some way to ensure that there is some uniformity in the way that sentences are understood by different speakers.

This was accomplished by positing that certain inferences that follow on the basis of a given word would have to become conventionalized. In retrospect, it would seem that Quine's (1951) result—i.e. that there could be no meaningful distinction between things that would be "true" of an **N**, by virtue of what **N** "meant," and those that would be "true" contingently—was so shocking is that conventionalized word-meaning is, indeed, a part of the way we communicate.

This is not to say that Quine was wrong, per se, but only that he, and the commentators after him, have failed to distinguish between whether they were discussing the philosophy of science or language as it is used by speech communities. That is, it is possible that "x goes around the Earth" might have at one point been a conventionalized inference that followed from "x is a star." That does not make it "true" in the scientific sense, as Quine was correct to point out.

In conclusion, I think we can say that we have seen here the beginnings of a fairly powerful model of comprehension.

6.2 Directions for Future Research

This inquiry essentially suggests a new paradigm for analyzing the comprehension of language and, as such, I think the range of possible directions for future research that it opens up are vast.

I would like to begin with the fact that I repeatedly stressed in §1 that my goal was to create an empirical model, whose success ought to, eventually if not immediately, be assessed in terms of its ability to make predictions about observables. I think that we have succeeded in making certain non-trivial predictions, especially in §4. For example, our simple model of comprehension was able to predict that Annie would be able to resolve Bob's implicated answer about the fact that Socrates would be a citizen of Greece in (78) on p. 62.

And, our more complex model of comprehension was able to make the *highly* non-trivial prediction that Bob would know to resolve "he" to Jones in interpreting Annie's answer about whether or not Jones had a girlfriend in (87) on p. 69.

Admittedly, I have not tested the validity of either of these results in the lab, but this is only because I propose that we can trust our intuitions in these regards so well that we can forego actual empirical verification. Thus, these may be, I think, considered to be "empirical" predictions about observables.

There is a problem, however, which arises from the fact that, because our theory is somewhat simplistic, it is going to make a wide range of mis-predictions, and a question arises about what to make of these. One major source of mis-predictions will be the fact that I have modeled the hearer, in §4.1, as a machine which is able to compute *all* inferences that follow from some new statement *instantaneously*. Obviously nothing could be less realistic.

For example, our theory would predict that any person, upon being given the axioms necessary to solve Fermat's Last Theorem, would do so instantly. Obviously this will be a mis-prediction. Two questions that then arise are the following. First, should we lose all hope in this model for this reason? Second, is there maybe a better starting point, for the purposes of beginning the process of theory-creation in a new paradigm, than the highly unrealistic one that I have proposed? The answer to both questions, I think, is no.

First of all, the only thing that we need to have faith in to believe that this model can eventually be used to create highly precise predictions is that, at some point, some cognitive science will produce a theory of human inferential ability—i.e. a theory that models the kinds of inferences people can compute.

And, since people have differing levels of intelligence, a model of what some person can compute will often have to be relative to the individual. Blanket statements about what people can do will also lead to false predictions.

Consider, in this regard, some unpublished findings of Pelletier and Coppola, in which they

demonstrate that people have an easier time computing the answer to (161) than they do to (162)[1]:

(161) Consider the following sorts of designs: White Square, Black Square, White Circle, Black circle. Each of these designs will be called a "Thog" if and only if it has:

a. The *same* shape and the *same* color as the White Square, or

b. A *different* shape and a *different* color than the White Square.

(162) Consider the following sorts of designs: White Square, Black Square, White Circle, Black circle. Each of these designs will be called a "Thog" if and only if it has:

a. The *same* shape but a *different* color as the White Square, or

b. A *different* shape than but the *same* color as the White Square.

(Note that the answer in the first case is that only White Square and Black Circle are Thogs. In the second case, the answer is that only the Black Square and White Circle are Thogs.)

In other words, as people find the first problem easier, and they find the kind of disjunctive reasoning in the second problem to be more difficult. Now, it would seem as though a highly advanced cognitive scientific theory of human inferential ability would be able to predict this result from more general principles.

When such a theory *is* available, it can be combined with my theory, that assumes infinite computational ability, to yield a highly realistic picture of language use. Admittedly, I am saying that the ability of my theory to make highly precise predictions depends upon a field that is not yet extant. This would seem to be a slightly suspicious practice. But, at the same time, it seems to me quite plausible that a theory of what kinds of inferences are easier than others is not a highly controversial one to assume.

So, I have argued that we should not be worried that, at this time, our theory of language use makes somewhat imprecise predictions, if we are willing to assume that, at some point, cognitive science will be at a point at which it is able to predict what sorts of inferences people are capable of. In the meantime, should those studying the interface between language and thought pick a different, more realistic starting point than I have given here?

I would argue that an infinite model of computation is the perfect starting point. This sort of model totally abstracts away from all of the messy details of a theory of inferential ability. It would seem that the alternative is to consider, in addition to our model, which is complex enough as it is, some, at present, poorly developed theory of inferential ability. Such a model, being poorly

[1]These findings are based on variations on Wason's (1968) "Thog" experiment.

developed, will lead to its own incorrect predictions, and one would have to wonder what had been gained for adding so much complexity to papers of this sort.

So, in conclusion of the discussion of mis-predictions of our model, I would say that it will be important for me, eventually, to be able to combine my infinitely capable model of comprehension with a more realistic model of human inferential ability. But, I do not think that it would necessarily be prudent to incorporate the messy details of such models in every single work on the interface between language and thought.

Aside from this issue, there is one matter that I feel deserves special attention. Recall example (78), repeated here:

(163) a. Annie: Is Socrates a citizen of Greece?

b. Bob: Socrates is a Greek.

I said at the time that our inferential model of comprehension could explain how Annie and any bystanders could realize that Bob had answered the question via implicature. Recall that, in that section, we saw that Annie and Bob both believed that all Greeks were both citizens of Greece, as well as toga-wearers. But, out of the two inferences that Annie would draw, clearly one inference—i.e. that Socrates is a citizen of Greece—is special, in this case. At present, my model is not sophisticated enough to differentiate between all of the other inferences that follow from Socrates being a Greek, and the particular inference that Bob especially wants Annie to draw, which is that Socrates is a Greek.

In other words, our model does not have Annie thinking, "Well, there are a range of things that Socrates' Greekhood implies. But, I realize that Bob has the specific intention that I realize that Socrates' Greekhood implies that he is a citizen of Greece." Somehow, though, it seems to me that the model should somehow accomodate this.

Bibliography

Alchourròn, C., Gärdenfors, P., and Makinson, D. (1985). On the logic of theory change: Partial meet contraction functions and their associated revision functions. *Journal of Symbolic Logic*, 510–530.

Alshawi, H., and Crouch, R. (1992). Montonic semantic interpretation. In *Proceedings of the 20th annual meeting on Association for Computational Linguistics*.

Andrews, P. B. (1986). *An Introduction to Mathematical Logic: To Truth through Proof*. Orlando: Academic Press.

Asher, N., and Lascarides, A. (2003). *Logics of Conversation*. Cambridge: Cambridge University Press.

Asher, N., and Morreau, M. (1995). What some generic sentences mean. In G. Carlson and F. J. Pelletier (Eds.), *The Generic Book*, (pp. 300–338). Chicago: University of Chicago Press.

Ayer, A. J. (1936). *Language, Truth and Logic*. New York: Dover Publications.

Bach, K. (1994). Conversational implicature. *Mind and Language*, 9, 124–162.

Bach, K. (2005). Context ex machina. In Z. Szabó (Ed.), *Semantics vs. Pragmatics*, (pp. 15–44). Oxford: Oxford University Press.

Bach, K. (2006). Impliciture vs. explicature: What's the difference? In *Granada workshop on Explicit Communication*. Can be retrieved at http://userwww.sfsu.edu/ kbach/Bach.ImplExpl.pdf.

Bar-Hillel, Y. (1954). Indexical expressions. *Mind*, 63, 359–379.

Block, N. (1986). Advertisement for a semantics for psychology. In P. A. French (Ed.), *Midwest Studies in Philosophy, Volume X*, (pp. 615–678).

Block, N. (1993). Holism, hyper-analycity, and hyper-compositionality. *Mind and Language*, 8, 1–27.

Bos, J. (1996). Predicate logic unplugged. In *In Proceedings of the 10th Amsterdam Colloquium*, 133–143.

Brachman, R. J. (1979). On the epistemological status of computer networks. In *Associative Networks: Representation and Use of Knowledge by Computer*, (pp. 3–50). New York: Academic Press.

Carlson, G., and Pelletier, F. J., (Eds.). (1995). *The Generic Book*. Chicago: University of Chicago Press.

Carnap, R. (1956). *Meaning and Necessity*. Chicago: University of Chicago Press. Originally published in 1947.

Carston, R. (1999). The semantics/pragmatics distinction: a view from relevance theory. In K. Turner (Ed.), *The Semantics/Pragmatics Interface from Different Points of View*. Oxford: Elsevier.

Carston, R. (2000). Explicature and semantics. *UCL Working Papers in Linguistics*, 12, 1–44.

Carston, R. (2004). Truth-conditional content and conversational implicature. In C. Bianchi (Ed.), *The Semantics/Pragmatics Distinction*. California: Stanford University.

Chierchia, G., and McConnell-Ginet, S. (2000). *Meaning and Grammar*. Cambridge, MA: MIT Press.

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory*, II-2, 113–124.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger Publishers.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge: MIT Press.

Church, A. (1940). A formulation of the simple theory of types. *Journal of Symbolic Logic*, 10, 56–68.

Davidson, D. (1967). Truth and meaning. *Synthèse*, 17, 304–323. Page numbers from reprint in, S. Davis and B. Gillon (Eds.), Semantics: A reader, Second Edition, Oxford University Press, Oxford, 2004, 222–233.

Davidson, D. (1986). A nice derangement of epitaphs. In E. Lepore (Ed.), *Truth and Interpretation*, (pp. 433–446). Oxford: Blackwell.

Davis, S., and Gillon, B. S., (Eds.). (2004). *Semantics: A Reader*. Oxford: Oxford University Press.

Dowty, D., Wall, R. E., and Peters, S. (1981). *Introduction to Montague Semantics*. Dordrecht: Reidel.

Elio, R., and Pelletier, F. J. (1994). Belief revision as propositional update. *Cognitive Science*, 17, 165–187.

Field, H. (1977). Logic, meaning, and conceptual role. *The Journal of Philosophy*, 74, 379–409.

Fodor, J., and Lepore, E. (1992). *Holism: A shopper's guide*. Oxford: Blackwell.

Fodor, J. (1975). *The Language of Thought*. New York: Thomas Cromwell.

Fodor, J. (2001). Language, thought, and compositionality. *Mind and Language*, 16, 1–15.

Fodor, J., and Lepore, E. (1991). Why meaning (probably) isn't conceptual role. *Mind and Language*, 6, 328–43.

Frege, G. (1879). *Begriffsschrift, eine der arithmetischen nachgebildete Formelesprache des reinen Denkens*. Halle: L. Nebert.

Frege, G. (1892). Sense and reference. *Zeitscheift für Philosophie and philosophische Kritik*, 25–50. Reprinted in English in, P. Geach and M. Black (Eds.), Translations from the Philosophical Writings of Gottlob Frege, Blackwell, Oxford, 1952, 56–78.

Gazdar, G. (1979). *Pragmatics: Implicature, Presupposition, and Logical Form*. New York: Academic University Press.

Gillmore, P. (2005). *Logicism Renewed*. Wellesley, Massachusetts: A. K. Peters, Ltd.

Goddard, C., and Wierzbicka, A., (Eds.). (1994). *Semantic and Lexical Universals: Theory and empirical findings*. Amsterdam: John Benjamins.

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. Morgan (Eds.), *Syntax and Semantics Volume 3: Speech Acts*, (pp. 41–58). New York: Academic Press. Page numbers from reprint in, H. P. Grice, Studies in the Way of Words, Cambridge, Harvard University Press, 1989, 22–40.

Gundel, J. K., Hedberg, N., and Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69, 274–307.

Harman, G. (1982). Conceptual role semantics. *Notre Dame Journal of Formal Logic*, 23, 242–256.

Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases*. Doctoral Dissertation, University of Massachusetts Amherst.

Heim, I., and Kratzer, A. (1998). *Semantics in Generative Grammar*. Oxford: Blackwell.

Hempel, C. (1950). Problems and changes in the empiricist criterion of meaning. *Revue Internationale de Philosophie*, IV, 41–63.

Hempel, C. (1966). *Philosophy of Natural Science*. Toronto: Prentice-Hall.

Joshi, A. K., Levy, L., and Takahashi, M. (1975). Tree adjunct grammars. *Journal of the Computer and System Sciences*, 10, 136–163.

Kamp, H. (1981). A theory of truth and semantic representation. *Formal Methods in the Study of Language*, 1, 277–322.

Kamp, H., and Partee, B. H. (1995). Prototype theory and compositionality. *Cognition*, 57, 129–191.

Kamp, H., and Reyle, U. (1993). *From Discourse to Logic: Introduction to Model Theoretic Semantics of Natural Language, Formal Logic, and Discourse Representation Theory*. Studies in Linguistics and Philosophy, 42. Dordrecht: Kluwer.

Kaplan, D. (1977). *Demonstratives*. Unpublished manuscript. Page numbers from reprint in, S. Davis and B. Gillon (Eds.), Semantics: A Reader, Second Edition, Oxford University Press, Oxford, 2004.

Karttunen, L. (1968). *What do referential indicies refer to?*. RAND corporation report P 3854, unpublished.

Karttunen, L. (1976). Discourse referents. In J. McCawley (Ed.), *Syntax and Semantics 7: Notes from the Linguistic Underground*. New York: Academic Press.

Katz, J. J., and Postal, P. (1964). *An Integrated Theory of Linguistic Descriptions*. Cambridge, MA: MIT Press.

Lenat, D. (1997). From *2001* to 2001: Common sense and the mind of HAL. In D. G. Stork (Ed.), *HAL's Legacy: 2001's Computer as Dream and Reality*. Cambridge, MA: MIT Press.

Lepore, E. (1983). What model theoretic semantics cannot do? *Synthèse*, 54, 167–187.

Levinson, S. C. (2000). *Presumptive Meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Lewis, D. (1970). General semantics. *Synthèse*, 22, 18–67.

Mann, W. C., and Thompson, S. A. (1987). Rhetorical structure theory: A framework for the analysis of text. *IPRA Papers in Pragmatics*, 1, 1–21.

Montague, R. (1968). Pragmatics. In R. Klibanski (Ed.), *Contemporary Philosophy*, (pp. 102–121). Florence: La Nuova Italia Editrice. Reprinted in, R. Thomason (Ed.), Formal Philosophy: Selected papers of Richard Montague, Yale University Press, New Haven, 1974, 95–118.

Montague, R. (1970a). English as a formal language. In *Linguaggi nella società e nella tecnica*. Milan: Edizioni di Communità. Reprinted in, R. Thomason (Ed.), Formal Philosophy: Selected papers of Richard Montague, Yale University Press, New Haven, 1974, 188–221.

Montague, R. (1970b). Pragmatics and intensional logic. *Synthèse*, 22, 68–94. Reprinted in, R. Thomason (Ed.), Formal Philosophy: Selected papers of Richard Montague, Yale University Press, New Haven, 1974, 119–147.

Montague, R. (1970c). Universal grammar. *Theoria*, 36, 373–378. Reprinted in, R. Thomason (Ed.), Formal Philosophy: Selected papers of Richard Montague, Yale University Press, New Haven, 1974, 373–398.

Montague, R. (1974). The proper treatment of quantification in ordinary English. In R. Thomason (Ed.), *Formal Philosophy: Selected papers of Richard Montague*, (pp. 247–270). New Haven: Yale University Press.

Newmeyer, F. (1986). *Linguistic Theory in America: Second Edition*. Orlando: Academic University Press.

Partee, B. H. (1980). Semantics—mathematics or psychology? In R. Bäuerle, U. Egli, and A. von Stechow (Eds.), *Semantics from Different Points of View*, (pp. 1–14). Berlin: Springer-Verlag.

Partee, B. H. (1996). The development of formal semantics in linguistic theory. In S. Lappin (Ed.), *The Handbook of Contemporary Semantic Theory*. Oxford: Blackwell.

Pelletier, F. J. (1994). The principle of semantic compositionality. *Topoi*, 13, 11–24.

Pelletier, F. J. (2009, forthcoming). Compositionality and holism. In *Oxford Handbook of Compositionality*. Oxford: Oxford University.

Pollard, C., and Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Prince, A., and Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar*. Rutgers University Center for Cognitive Science Technical Report 2. Reprinted as, A. Prince and P. Smolensky, Optimality Theory: Constraint interaction in generative grammar, Blackwell, Oxford, 2004.

Quillian, M. R. (1968). Semantic memory. In *Semantic Information Processing*. Cambridge, MA: MIT Press.

Quine, W. v. O. (1951). Two dogmas of empiricism. *The Philosophical Review*, 60, 20–43.

Quine, W. v. O. (1960). *Word and Object*. Cambridge: Cambridge University Press.

Reyle, U. (1993). Dealing with ambiguities by underspecification: Construction, interpretation and deduction. *Linguistics and Philosophy*, 10, 123–179.

Sellars, W. (1974). Meaning as functional classification. *Synthèse*, 61, 3–16.

Skinner, B. F. (1935). Two types of conditioned reflex and a pseudo type. *Journal of General Psychology*, 12, 66–77.

Sperber, D., and Wilson, D. (1986). *Relevance*. Oxford: Blackwell.

Sperber, D., and Wilson, D. (1995). *Relevance: Communication and cognition*. Oxford: Blackwell.

Sperber, D., and Wilson, D. (1998). The mapping between the public and the private lexicon. In P. Carruthers and J. Boucher (Eds.), *Language and Thought: Interdisciplinary Themes*, (pp. 184–200). Cambridge: Cambridge University Press.

Stanley, J. (2000). Context and logical form. *Linguistics and Philosophy*, 23, 391–434.

Tarski, A. (1935). Der wahrheitsbegriff in den formalisierten sprachen. *Studia Philosophica*, I, 265–405. Page numbers from reprint in English as "The concept of truth in formalized languages," in, J. H. Woodger (Trans.), Logic, Semantics, Metamathematics, Oxford University Press, Oxford, 1956, 152–278.

Thomason, R. (1974). Introduction. In R. Thomason (Ed.), *Formal Philosophy: Selected Papers of Richard Montague*. New Haven: Yale University Press.

Thomason, R., and Stalnaker, R. (1973). A semantic theory of adverbs. *Linguistic Inquiry*, 4, 195–220.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20, 273–281.

Wierzbicka, A. (1972). *Semantic Primitives*. Frankfurt: Athenaum.

Wiggins, D. (1997). Meaning and truth conditions: from Frege's grand design to Davidson's. In B. Hale and C. Wright (Eds.), *A Companion to the Philosophy of Language*, (pp. 3–28). Oxford: Blackwell.

Wilson, N. L. (1959). Substance without substrata. *Review of Metaphysics*, 12, 521–539.

Wittgenstein, L. (1921). *Tractatus Logico-Philosophicus*. New York: Routledge.

Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.