ON THE DISTINCTION BETWEEN
LITERAL AND NON-LITERAL LANGUAGE

by

David Sturdee

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Philosophy
University of Toronto

© Copyright by David Sturdee 1999
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

0-612-41319-5
On The Distinction Between Literal and Non-Literal Language.
David Sturdee, Doctor of Philosophy, 1999
Department of Philosophy, University of Toronto

Abstract

My topic is the nature and scope of literal language. Broadly stated, my objective is to examine how philosophers in the Western tradition have made use of the concept of literal meaning, in particular, how they have treated the distinction between literal and non-literal language. My approach is historical. I trace how various philosophers, from Plato and Aristotle to Frege to Grice and Davidson, treat this distinction, arguing that it is not until the advent of analytic philosophy in general, and natural language semantics in particular, that it acquires genuine philosophical import. I argue further that although natural language semantics depends on the availability of a principled distinction, its attempt at an identification of literal meaning with context-independent meaning is problematic. Instead, drawing on Davidson's notion of first meaning, I argue that meaning qua meaning is fundamentally intentional and originates with individual acts of communication whereas literal meaning depends for its content on the standards of the linguistic community to which the speaker belongs.
Capturing the ever-changing flux in words is like attempting to map a cloud.

—Attributed to Cratylus
Acknowledgments

This thesis benefitted greatly from the moral and philosophical involvement of many of my colleagues. Indeed, without their support it would not have been possible.

In particular, I would like to thank Andrew Latus and Joshua Mozersky for their valuable suggestions for improving drafts of chapters one and two. Also, I am grateful to Cheryl Misak for the care and acuity with which she read early drafts of the thesis. In addition, I owe a special gratitude to Ian Hacking for the wisdom and grounding he brought to the thesis. Most of all, I am indebted to Bernard Katz. He made many constructive suggestions for improving the structure and content of the thesis, virtually all of which were incorporated into the final draft. More important, his enthusiasm and support for the project never wavered and, indeed, motivated me during the most difficult stages of its writing. Finally, I would like to thank Cath and my family for their love, their support and their never tiring faith in me.

I would also like to acknowledge the financial support I received from the Social Sciences and Humanities Research Council of Canada during the time in which this thesis was written.
# Contents

Abstract ii

Acknowledgments iv

Introduction 1

‘Communicated’ and ‘Strict’ Literal Meaning 4

Outline 5

I. The Role of the Literal/Non-literal Distinction in the History of Western Philosophy 14

   Early Greek Philosophy and Rhetoric 14
   Medieval Theologians and Scriptural Hermeneutics 19
   Modern Philosophy 25
   Summing Up 32

II. Frege and the Emergence of Natural Language Semantics 38

   The Grandfather of Analytic Philosophy: Gottlob Frege 41
   Frege’s Objective: Securing the Foundations of Mathematics 42
   Frege’s Logic 43
   Logic and Language: Truth 46
   Logic and Language: Sense 49
   Frege’s Anti-Psychologism Revisited 53
   Frege’s Literal Language 57
   Demarcating the Bounds of Sense 62

III. The Orthodox Position 69

   Frege’s Influence 69
   Compositionality and Truth Conditions 72
   The Orthodox Position 74

IV. On the Notion of Context-Independent Meaning 76

   What Does ‘Context-Independent’ Mean? 77
   Traditionally Cited Barriers to Context-Independent Meaning 79
   Searle’s Challenge to Context-Independent Meaning 82
   Searle’s Background Assumptions 87
   Responses to Searle 89
   Berg’s Response to Searle 89
   Katz’s Response to Searle 93
   Searle and the Orthodox Position 102
   Searle on Meaning 104
   Getting Back to Linguistic Basics 108
V. Grice and the Foundations of Linguistic Communication

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Foundation: Grice</td>
<td>111</td>
</tr>
<tr>
<td>Branch One: Saying and Implicating</td>
<td>112</td>
</tr>
<tr>
<td>Saying vs. Implicating and the Literal/Non-Literal Distinction</td>
<td>116</td>
</tr>
<tr>
<td>Literal/Non-Literal vs. Literal/Figurative</td>
<td>117</td>
</tr>
<tr>
<td>Conventional Implicature</td>
<td>118</td>
</tr>
<tr>
<td>Implicature and 'Strict' and 'Communicated' Literalness</td>
<td>119</td>
</tr>
<tr>
<td>Branch Two: Meaning</td>
<td>122</td>
</tr>
<tr>
<td>Meaning as Intending</td>
<td>123</td>
</tr>
<tr>
<td>Non-Natural Meaning, Speaker Meaning and Sentence Meaning</td>
<td>128</td>
</tr>
<tr>
<td>Connecting the Branches: 'Say' and Utterance-Type Meaning</td>
<td>131</td>
</tr>
<tr>
<td>Problem: Indeterminacy of Utterer's Occasion Meaning and Figurative Language</td>
<td>133</td>
</tr>
<tr>
<td>The Deeper Problem: Interpreting Figurative Language on the Gricean Model</td>
<td>135</td>
</tr>
<tr>
<td>Summing Up</td>
<td>140</td>
</tr>
</tbody>
</table>

VI. Davidson's First Meaning and the Literal/Non-Literal Distinction

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>On the Apparent Regularity of Language Use</td>
<td>143</td>
</tr>
<tr>
<td>Mrs. Malaprop and Conventional Meaning</td>
<td>144</td>
</tr>
<tr>
<td>First Meaning</td>
<td>146</td>
</tr>
<tr>
<td>A Conventionalist Response</td>
<td>149</td>
</tr>
<tr>
<td>Radical Interpretation</td>
<td>152</td>
</tr>
<tr>
<td>First Meaning and Radical Interpretation</td>
<td>157</td>
</tr>
<tr>
<td>Non-Literal Language and Radical Interpretation</td>
<td>165</td>
</tr>
<tr>
<td>Davidson on the Distinction Between Literal and Non-literal</td>
<td>166</td>
</tr>
<tr>
<td>Davidson and Non-Verbal Communication</td>
<td>167</td>
</tr>
<tr>
<td>First Meaning and Literal Meaning</td>
<td>171</td>
</tr>
<tr>
<td>Reconciliation</td>
<td>177</td>
</tr>
</tbody>
</table>

Conclusion

Appendix: Dictionary Meaning

References
Introduction

My topic is the nature and scope of literal language. Broadly stated, my objective is to examine how philosophers in the Western tradition have made use of the concept of literal meaning, in particular, how they have treated the distinction between literal and non-literal language. Roughly, a literal utterance is an utterance in which the principal intention of the speaker is to communicate what she says—for example, I say ‘Tomorrow in Toronto it will rain’ intending to communicate to you that tomorrow in Toronto it will rain. A non-literal utterance, in contrast, is an utterance in which the principal intention of the speaker is to communicate something other than what she says—for example, Alfred Lord Tennyson writes ‘Sleep is the brother of death’ intending not that we attribute to sleep and death a common parentage, but rather that we notice that there are salient similarities (and differences) between the physical states of sleep and death, just as there are important similarities (and differences) between male siblings. Of course, to set this characterisation of the distinction on a firm philosophical footing, we must provide an account of ‘says’ that does not make use of the notion of literal meaning.

The motivation for this investigation is a desire to reconcile two seemingly conflicting observations about language. The first is that there exists no sufficiently determinate set of sufficiently subtle criteria to support a principled distinction between

---

1 I say non-literal language as opposed to figurative language because I don’t want to prejudice the investigation—there may be non-figurative non-literal uses of language.

2 This applies equally to stage and other fictional utterances, though expectations about the truth of such claims will be different.
literal and non-literal language. Indeed, although it is a distinction that many philosophers and linguists take for granted, arguments for it rarely proceed beyond the swapping of examples tailored to fit particular theses. When we redirect our attention from these tailored examples to examples that are more representative of ordinary discourse, we find many cases for which we are uncertain whether to categorise them as literal or non-literal. This uncertainty is aggravated by the fact that the origin of many literal expressions is in metaphor—‘a hollow idea’, ‘buried in work’, ‘an ugly personality’, ‘moral fibre’, ‘immersed in thought’, ‘a budding theory’, ‘a wealth of ideas’, ‘a bottleneck in traffic’. Following H.W. Fowler, we might characterise such expressions as dead metaphors—that is, as once metaphorical expressions that have been used so frequently that “speaker and hearer have ceased to be aware that the words used are not literal” (A Dictionary of Modern English Usage 359). In contrast, live metaphors are those that “are offered and accepted with a consciousness of their nature as substitutes for their literal equivalents” (A Dictionary of Modern English Usage 359). But the transition from live metaphor to dead metaphor—i.e., from metaphorical status to literal status—is neither smooth nor easily discernable. Expressions which we might call dying metaphors—that is, expressions that are neither clearly literal nor clearly metaphorical—pervade ordinary language. We can imagine, for example, a time at which the above expressions were vying for an established position in the lexicon. Expressions

---

3 cf. Lakoff and Johnson, Metaphors We Live By.
today such as 'surf the web' might fit into this category.\textsuperscript{4}

It would be wrong, however, to conclude from the above that the literal/non-literal distinction is artificial and of no genuine philosophical importance.\textsuperscript{5} It would be wrong because, although we may be unable to provide a principled distinction between literal and non-literal language, nevertheless it seems that we need something like it to explain (in part) how so-called figurative devices work. Indeed, this is the second observation about language that I want to highlight. Metaphor, irony, and the like are parasitic on literal meaning. That is, one cannot understand a figurative utterance without first understanding its literal meaning—for example, one cannot begin to understand Tennyson’s metaphor without knowing what the words ‘sleep’, ‘brother’ and ‘death’ standardly mean in our linguistic community. It would seem, then, that to explain adequately how we interpret figurative utterances, we need to hold literal meaning constant. Thus the concept of literal meaning seems to do genuine philosophical work.

My task, then, is to reconcile the observation, on the one hand, that there exists no principled distinction between literal and non-literal language with the observation, on

\textsuperscript{4} Richard Rorty suggests that expressions acquire literal status when and only when they acquire a place in a pattern of justification of belief, or, more precisely, “when these unfamiliar noises acquire familiarity and lose vitality through being not just mentioned... but used: used in arguments, cited to justify beliefs, treated as counters within a social practice, employed correctly or incorrectly” (“Hesse and Davidson on Metaphor” 295). But, again, locating precisely when this occurs is impracticable.

\textsuperscript{5} Mary Hesse, for example, draws this conclusion. In her dialogue on metaphor with Rorty and Susan Haack, she claims that the literal/non-literal distinction “is one of those dichotomies which might have gone the way of synthetic/analytic, discovery/justification, explanation/description, or reasons/causes” (Hesse, “Tropical Talk: The Myth of the Literal” 298; Rorty, “Hesse and Davidson on Metaphor”; Haack, “Surprising Noises: Rorty and Hesse on Metaphor”).
the other hand, that we need something like the literal/non-literal distinction to explain (in part) how figurative utterances work.

‘Communicated’ and ‘Strict’ Literal Meaning

Before we begin, a further observation is worth making. Philosophers typically conflate what are at least two ways in which we employ the term literal. Consider the sentences ‘Ian’s boss gave him the boot’, ‘Kal always uses chopsticks’, ‘Nobody knows the trouble I’ve seen’ and ‘The room was silent’. Each sentence may be used to communicate a message to a member of our linguistic community. The first may be used to communicate the message that Ian’s boss fired him. The second may be used to communicate the message that whenever Kal eats, he uses chopsticks. The third may be used to communicate the message that nobody other than the speaker knows the trouble that the speaker has encountered. Similarly, the fourth may be used to communicate the message that the people/things in the room were very quiet. Indeed, in each case, we might say that this is what the utterance means literally. But each sentence also has a strict reading which we might say gives its literal meaning. That is, one might argue that the first sentence means literally that Ian’s boss gave him the boot—i.e., his boss gave to him a certain item of footwear. Likewise, one might argue that the second sentence means literally that Kal always uses chopsticks (i.e., whether he is eating or not), that the third means literally that nobody knows the trouble that the speaker has encountered (i.e., not even the speaker) and that the fourth means that the people/things in the room were absolutely silent. Thus there are two ways in which we employ the term literal meaning.
I will call the first sense *communicated literal meaning*—roughly, the meaning that is communicated to a typical speaker of the language by the utterance. I will call the second sense *strict literal meaning*—roughly, how a logician or an unforgiving lawyer might interpret the utterance. In the chapters ahead, this distinction proves important in making sense of the commitments of various competing accounts of literal meaning.\(^6\)

**Outline**

The thesis divides roughly into two parts. The first three chapters assess the historical significance of the distinction between literal and non-literal language, beginning with pre-twentieth century philosophy in the Western tradition and ending with Gottlob Frege and the emergence of what I will call the *orthodox position*. In the first chapter I argue that, with the exception of proponents of scriptural hermeneutics, philosophers in the Western tradition, from Plato and Aristotle to Hobbes, Locke and Leibniz, treat the distinction as unimportant philosophically. Interestingly, as we will discover, with respect to language, the chief concern of these thinkers is not to demarcate between language that is literal and all else, but rather between language that is clear and standard and language that is obscure and uncommon. Indeed, I argue that it is not until the advent of natural language semantics at the stewardship of Frege that the distinction

\(^6\) Note also that sometimes we use the term ‘literal’ (i) to issue a preemptive protest that we are not exaggerating—for example, expecting disbelief, we might say ‘The winning fish was *literally* twice as big as the second place fish’, and (ii) to *assist* us with our exaggeration—for example, ‘He was *literally* ten times stronger than anyone I’ve met!’. As we will find, in these cases we would say that the term ‘literal’ (and its cognates) is being used *non*-literally.
between literal and non-literal language acquires genuine philosophical import.

The objective of chapter two is to tell a story of how Frege’s underlying interest in the foundations of mathematics, in particular his interest in logical inference, greatly influences his claims about natural language. Frege argues that in order to explain certain features of language—for example, our ability to mean the same thing by the same sentence—we must posit meanings as mind-independent Platonic entities. Understanding someone’s utterance, according to Frege, consists in attaching to the sentence uttered the same meaning that the speaker attaches. I argue that implicit in Frege’s description is that these meanings are literal meanings. But whereas Frege touts the importance of demarcating clearly between what a sentence means and all else (for example, the psychological effects it has on a hearer), I contend that he says too little about how this is done.

Frege proves to be an important figure in our investigation because various of the ideas found in his writings on language are taken up by subsequent analytic philosophers including Bertrand Russell, Ludwig Wittgenstein and Rudolf Carnap. It is the objective of chapter three to argue that various important theses found in Frege’s writings—for example, his distinction between the meaning of a sentence and the use to which that sentence is put, and his appeal to truth conditional semantics—provide the framework within which further philosophising about language is done. I outline the main tenets of this framework and dub it the orthodox position. Because among these tenets is the claim that in order to construct a semantics for natural language we must distinguish the literal meaning of a sentence (or expression) from its psychological effects or the purposes to
which it is put, Frege and certain subsequent analytic philosophers raise the stakes on the importance of making a principled distinction between literal meaning and all else. Indeed, I argue, these analytic philosophers introduce the problem of demarcating clearly between literal and non-literal language.

Having provided an understanding of the main suppositions and, indeed, the scope and influence of the orthodox position, in the second part of the thesis I turn my attention to some of its problems. Specifically, in chapter four I consider in some detail what is perhaps the orthodox position's most controversial claim: the claim that literal meaning is identified with context-independent meaning. Here I enter into the debate between Jerrold Katz, a proponent of the orthodox position, and John Searle, one of its most forceful critics.

Katz proposes an anonymous letter scenario as a heuristic device for making sense of the notion of context-independent meaning and, so, of literal meaning. We are invited to imagine that a speaker of a language receives an anonymous letter consisting only of a single sentence of the language without any knowledge of the motive, circumstances of transmission, or any other fact relevant to understanding the sentence on the basis of its context of utterance. Katz argues that context-independent meaning consists in the meaning that a speaker of the language would attribute to the sentence in such a situation. Searle challenges Katz, arguing that in general the notion of the literal meaning of a sentence only has application against a set of background assumptions, assumptions which are not all and could not all be realised semantically. I argue that Searle is right, and that to save the anonymous letter scenario as an explanatory device, Katz must invoke
both the notion of *minimal semantic content* and the notion of an *ideal interpreter*. I conclude that although this move shields him from Searle's criticisms, it nevertheless renders the anonymous letter scenario explanatorily inert.

Dissatisfied with the *orthodox position*'s identification of literal meaning with context-independent meaning, at the end of chapter four I propose that we change two of the basic assumptions with which we are operating. First, rather than assume, as the *orthodox position* does, that meaning consists in something abstract and external (Fregean *senses* or the like), I propose instead that meaning results from individual acts of communication (or, better, individual attempts at communication). Secondly, rather than suppose, again as the *orthodox position* does, that every sentence has a context-independent meaning which is to be identified with its literal meaning, I propose that we supplant this with the supposition that every sentence, in its context, has a literal meaning. I argue that these changes enable us to work toward a new approach to understanding the distinction between literal and non-literal language, one that incorporates the following ideas: (i) that meaning is not a precursor to understanding, but rather the result of understanding; (ii) that context plays a critical role in all cases of linguistic communication; and (iii) that linguistic communication is but one element of the broader phenomenon of interpersonal communication.

Equipped with these changes, in chapter five I begin to explore an alternative approach to understanding literal meaning, one based on the works of H.P. Grice. To

---

7 A notion which corresponds closely to my notion of *strict literal meaning*. 
begin, I argue that Grice's distinction between what someone *says* and what he *implicates* by his utterance (i.e., what he suggests, implies, indicates, and the like) provides us with a framework for understanding the distinction between literal and non-literal meaning. The idea is this: the literal meaning of an utterance is identified with what the person *says*; all non-literal effects are *implicatures*. Accordingly, I argue that metaphor, irony, hyperbole, and the like are kinds of implicatures. Although this is a promising line for understanding the distinction between literal and non-literal meaning, its success depends on the availability of an account of Grice's notion of *says*. I turn, as a result, to Grice's attempt to provide such an account.

Grice proposes an approach to meaning that is grounded on a general understanding of how people can and do mean things by their actions. Employing a broad notion of *utterance*, one that incorporates both verbal and non-verbal expressions, Grice argues that for a speaker to mean *p* by issuing an utterance *x*, she must intend her audience to believe that she believes that *p*, and the audience must arrive at this belief in part by their awareness that this is what the speaker intends by her utterance. I call this the *Gricean analysis of understanding*. It is of interest to our investigation because it appears to provide us with the foundation for a non-circular account of his notion of *says*, and so of literal meaning. It has the further advantage of situating literal meaning in the broader context of the general phenomenon of interpersonal communication.

But the *Gricean analysis of understanding* is an analysis of *speaker meaning* (or what Grice calls *utterer's occasion meaning*), not of *sentence meaning* (or what Grice calls *utterance-type meaning*). In order to generate an account of sentence meaning,
Grice appeals to the notion of a linguistic convention. Roughly, an utterance-type \( X \) means \( p \) if it is a convention (a Lewisian convention) of the members of a linguistic community (or the habit of an individual speaker in the case of idiomatic utterances) to utter \( X \) with the intention of inducing the belief that \( p \) (or at least the intention of inducing the belief that the speaker believes that \( p \)), based upon the *Gricean analysis of understanding*. Because, according to Grice, sentence meaning is derivative of speaker meaning, he must assume that utterances of the same type are made, as a general rule, with the same basic intentions. But any sentence, from the most simple to the most fantastic might be employed for any number of purposes: to make an assertion, to tell a joke, to issue a promise, to make a metaphor, and so on. Indeed, I argue, this exposes the inability of Grice's account to explain how figurative utterances are interpreted. In the case of a figurative utterance, the intention with which the utterer issues her utterance cannot be determined unless the audience *already knows* the meaning of the utterance—figurative interpretation is parasitic on literal interpretation. Thus I argue that Grice's attempt to construct an account of sentence meaning based on an account of speaker meaning must fail.

Nevertheless, despite its shortcomings, the *Gricean analysis of understanding* provides us with a persuasive basic description of what actually occurs during the interpretive process. Further, it meets our supposition that meaning arises out of understanding, it provides a critical role for context and it treats linguistic meaning as one element of the broader context of interpersonal communication. Moreover, his concepts of *saying* and *implicating* provide us with a promising template for characterising the
distinction between literal and non-literal language. Thus the objective of chapter six is to construct an account of sentence meaning which makes use of something like the Gricean analysis of understanding. I argue that we find the substance of such an account in the works of Donald Davidson.

Davidson uses the example of malaprops to argue that in order to preserve the distinction between sentence meaning and speaker meaning, we must pry apart what is literal in language from what is conventional. In place of the notion of literal meaning, which Davidson thinks carries with it too much philosophical baggage, he introduces the notion of first meaning. Distinct from conventional meaning, first meaning applies to sentences uttered by particular speakers on particular occasions. I argue that Davidson’s explanation of first meaning builds on the Gricean analysis of understanding, with one important difference. The first meaning of an utterance is specified by what Davidson calls the speaker’s first intention, i.e., the intention to utter words that will be interpreted in a certain way, as having certain truth conditions. Thus, in contrast to Grice’s account, Davidson’s appeal to first intentions provides him with an means of accounting for sentence meaning, not speaker meaning. Mrs. Malaprop’s utterance ‘This is a nice derangement of epitaphs’ means that the demonstrated item is a nice arrangement of epithets because this is how she intends it—i.e., her utterance—to be understood.

Having introduced Davidson’s notion of first meaning, I consider two important challenges to it. The first of these is the claim that although his malaprop argument shows that knowledge of conventional meaning is not sufficient to interpret Mrs. Malaprop’s utterance, it does not show that this knowledge is not necessary to interpret
her utterance. I agree with this criticism, noting that to show that shared conventions are not an essential element of the interpretive process, Davidson needs to show that one could understand Mrs. Malaprop's utterance without knowing what it means conventionally. Nevertheless, I argue that Davidson has an argument for this, though this argument is found elsewhere in his writings, specifically, in his work on radical interpretation.

The second challenge that I consider is Catherine J.L. Talmage's claim that Davidson's methodology of radical interpretation does not always yield theories of meaning that give first meanings. Talmage argues that in cases of intended misinterpretation the meaning generated by the speaker's first intention and the meaning arrived at through the process of radical interpretation are in fact distinct. I argue that by refining the notion of first meaning—in particular, by adding the condition that the speaker equip the audience with enough clues so that she might reasonably expect that a radical interpreter would interpret her utterance as she intends—it is possible to keep together first meaning and meaning generated by the methodology of radical interpretation.

I then argue that, equipped with his notion of first meaning, Davidson is well positioned to make a principled distinction between literal and non-literal language: there is only one kind of meaning, first meaning, and this is equated with literal meaning. All non-literal devices, including metaphor, hyperbole, irony, and the like, are kinds of uses to which we might put a sentence that has a particular literal meaning. Making a metaphor, then, is like telling a joke.
I argue, however, that Davidson's equation of first meaning and literal meaning is problematic. Though we might agree that Mrs. Malaprop's utterance has as its first meaning that the demonstrated item is a nice arrangement of epithets, we would deny that it means this literally. Contra Davidson, I argue that we should distinguish between first meaning and literal meaning, allowing the standards of a linguistic community to set the bounds of the latter. This distinction provides us with a framework for understanding the distinction between communicated and strict literal meaning. For any utterance these classifications might apply: the utterance may have a first meaning as specified by the speaker's first intention; further, the utterance may have a communicated literal meaning as specified by the standards of his linguistic community; finally, the utterance may have a strict literal meaning, again, roughly, how it might be interpreted by a logician or an unforgiving lawyer. More importantly, treating literal meaning this way allows us to reconcile the apparently contradictory observations that first motivated our investigation. On the one hand, because literal meaning is determined by social constraints, and because these constraints are dynamic, determining the bounds of literal meaning for all sentences of a language is impracticable. On the other hand, Davidson's notion of first meaning allows us to make a principled distinction between what a speaker says and what she implicates by her utterance (or uses her utterance to do). Ultimately, then, whether an utterance is literal or non-literal depends on which notion of meaning is taken as the foundation.
Chapter One
The Role of the Literal/Non-Literal Distinction in the History of Western Philosophy

Tracing the roots of the distinction between literal and non-literal language—that is, determining in what contexts it is first employed, either implicitly or explicitly—is nigh impossible. No matter, our interest lies not with the origins of the distinction, but rather with its role in various philosophical projects. Indeed, our present objective is to provide a sense of how and why the distinction between literal and non-literal language matters to philosophers.¹

Early Greek Philosophy and Rhetoric

A distinction between literal and non-literal language is at least implicit as early as Plato (427-347 B.C.)² However, although himself a master of metaphor, Plato never discusses explicitly the topic of metaphor, nor the more general topic of non-literal language. He alludes to the existence of non-literal language in so far as he warns against the emotive power of poetry. Poetry, he writes, “irrigates and tends to [the passions] when they should be left to whither, and it makes them our rulers when they should be our subjects” (Republic 606d, insert mine). In other words, poetry is dangerous because it detracts

¹ Because our final interests are in contemporary Western analytic philosophy, the roots traced here will be those most influential on this school. I will have nothing to say about how this distinction figures in other philosophical traditions.

² William Bedell Stanford argues that the first occurrence of the Greek word for metaphor, μεταφορά, occurs in Isocrates’s (436-338 B.C.) Evagoras (190d) (Greek Metaphor 3). Stanford adds that before Isocrates, the historians Herodotus (c.485-425 B.C.) and Thucydides (c.456-404 B.C.) used the word μετονόμασειν which meant ‘to call by a new name’ or ‘to change something’s name’, but that in none of its contexts did it carry the same import as μεταφορά.
from reason, inviting us to indulge in the passions. If we are not already of impeccable moral inner constitution, he warns, poetry will lead to treating appearances as though they represent the underlying truth. In contrast to legitimate experts, “all [the poet] understands is how to represent things in a way which makes other superficial people, who base their conclusions on the words they can hear, think that he’s written a really good poem about shoemaking or military command or whatever else” (Republic 601a, insert mine). As an example of the rogue nature of poetry, he cites the ancients who “used poetical forms which concealed from the majority of men their real meaning” (Theaetetus 180d). Poets are not, however, the only targets of Plato’s criticism. In addition, he observes disapprovingly that Sophists, “by the power of their language, make small things appear great and great things small” (Phaedrus 267a).

Interestingly, in none of these passages does Plato specifically mention metaphor, or any other form of non-literal language, as the offending device. Rather, the distinction he draws is between useful, clear language, the kind of language used by men who “plainly demonstrate their meaning so that even shoe-makers may hear and assimilate their wisdom” (Theaetetus 180d) and others who, like “a painter [who] creates an illusory shoemaker, when not only does he not understand anything about shoemaking, but his audience doesn’t either... uses words and phrases to block in some of the colours of each area of expertise” (Republic 601a, insert mine). It would appear, then, that particular cases of non-literal language (as well as of literal language) may fall within either of the above two categories—that is, whether a particular utterance is clear appears for Plato to be irrespective of its status as literal or non-literal.
In contrast to Plato, Aristotle (384-322 B.C.) offers a somewhat systematic treatment of language. In the *Poetics* he tells us that "every word is either standard or foreign or metaphorical or ornamental or newly-coined or lengthened or shortened or altered" (*Poetics* 1457b). He adds, "by 'standard' I mean a name which a group of men ordinarily use" (*Poetics* 1457b). Metaphor, conversely, "consists in giving the thing a name that belongs to something else; the transference being either from genus to species, or from species to genus, or from species to species, or on grounds of analogy" (*Poetics* 1457b). A standard name, then, is a name whose use within a linguistic community is common and well understood. A metaphor, in contrast, constitutes a deviance from ordinary use since it necessarily involves a transfer of a name to an object for which it is not the name.

At first it may appear that here Aristotle means to introduce the distinction between literal and non-literal language. Indeed, in his seminal piece on the history of metaphor in the philosophical tradition, Mark Johnson cites Aristotle's definition of metaphor and announces, "the fatal separation—figurative vs. literal—has been made" ("Metaphor in the Philosophical Tradition" 6). But Johnson's conclusion is misleading. Granted, by defining metaphor Aristotle distinguishes between standard words and metaphorical words, but nonetheless this does not itself constitute the distinction between literal and figurative language. Aristotle's definition is of metaphor only (which, to be fair, for Aristotle includes simile (*Rhetoric*, 1406b)); it does not, however, apply to irony, hyperbole and other figurative uses of language which, typically, employ 'standard' language.
Further enquiry reveals that the distinction between literal and non-literal language is not Aristotle’s primary concern, just as it is not Plato’s. Consider the remainder of Aristotle’s list of types of words. Foreign words are those that are standard according to a different group of speakers (so the same word may be “both foreign and standard, but not to the same group” (Poetics 1457b)). A newly-coined word is one “not used at all by a group of men but introduced by the poet” (Poetics 1457b). The last three categories, words that are lengthened or shortened or altered, are the result of syntactic variations. (Of words that are “ornamental” Aristotle says nothing further, and so their exact nature remains a mystery.) As it turns out, it is not a distinction between literal and non-literal language that interests Aristotle—in fact, he never discusses such a distinction—rather, his interest lies with a different distinction: that between language that is ‘standard’ and language that is ‘strange’ (Poetics, 1458a). As already indicated, standard use is ordinary use (perhaps established by convention). “By ‘strange’,” Aristotle writes, “I mean names which are foreign or metaphorical or lengthened or any of those which are not standard” (Poetics, 1458a). Again, this list of ‘strange’ names includes metaphor and simile—clear cases of non-literal language—but it also includes variations on syntactic structure, foreign words, ornamental language and newly coined terms, each of whose classification as non-literal is, at the very least, not assured.

The impetus for Aristotle’s distinction between names that are standard and those that are strange becomes clear later in his discussion when he employs the distinction in an effort to school us on the proper use of language:

One virtue of language is to be lucid without lacking dignity. Now the most lucid kind of language is that which uses standard names, but such
language lacks dignity... Language which uses strange names has dignity and avoids what is commonplace. *(Poetics 1458a)*

Poetry, then, "should use a sort of mixture of the above [two kinds, standard and strange names]; for the use of foreign names and metaphors and ornaments and the other forms will prevent language from being commonplace and undignified" *(Poetics 1458a, translators' insert)*. When we are not engaged in poetic musing, however, Aristotle urges us to make every effort to use language that is clear, and he notes that "clearness is secured by using the words (nouns and verbs alike) that are current and ordinary" *(Rhetoric 1404b)*.

Aristotle's treatment of the distinction between 'standard' and 'strange' language sets the tone for many of the subsequent discussions of language found in the writings of classical rhetoricians, such as in Cicero's (106-43 B.C.) *De Oratore*. As with Aristotle, in *De Oratore* we find that no explicit distinction is made between literal and non-literal language. Rather, Cicero's discussion centres on the distinction between language that is clear and common, and language that is oblique and unusual, of which metaphor is but one case. Cicero, lifting a page from Aristotle, characterizes metaphor as "a species of borrowing... a brief similitude contracted into a single word; which word being put in the place of another as if it were in its own place" *(De Oratore III, xxxviii-xxxix)*. Save for passing remarks on simile (III, xli) and irony (II, lxv), this definition of metaphor represents the extent of Cicero's concern with explicating the distinction between literal and figurative language.

A refreshing exception to this trend is found in Quintilian's (c.35-c.100) *Institutio Oratoria*. In contrast to his predecessors, Quintilian centres his discussion not on
metaphors in particular, but on tropes in general: "By a trope is meant the artistic alteration of a word or phrase from its proper meaning to another" (*Institutio Oratoria* VIII, vi). In stark contrast to the earlier preoccupation with the distinction between language that is standard and that which is not, Quintilian acknowledges that "some [tropes] arise from words used *properly* and others from words used *metaphorically*" (*Institutio Oratoria* VIII, vi, insert mine). He adds, "I regard those writers as mistaken who have held that tropes necessarily involved the substitution of word for word" (*Institutio Oratoria* VIII, vi). In his discussion he notes that there are many types of tropes which employ the proper senses of words, including irony and hyperbole.

Nonetheless, Quintilian does little to direct us to clear differentiations between the various kinds of tropes he discusses, nor does he explain what determines proper meaning from which tropes are "artistic alterations." Rather, he assumes that the reader understands these distinctions, and, following Aristotle, centres his discussion on the appropriateness of various uses of language to proper oratory. So, even though in Quintilian we find an emphasis on the distinction between the standard use of language and various figurative uses, how exactly the distinction is made is left largely to our imagination.

*Medieval Theologians and Scriptural Hermeneutics*

In contrast to the relatively minor role that it plays in many early discussions of poetry and rhetoric, the distinction between literal and figurative language gains philosophical purchase with the advent of Scriptural hermeneutics. A central concern of Jewish and
Christian scholars is that the literal interpretation of Holy texts is not always feasible, nor even desirable. There are well known, problematic passages which, interpreted literally, undermine the dignity of the word of God because they appear nonsensical or trivial or, worse, inconsistent. The most difficult passages fall into one or both of two categories: first, those that contain apparent contradictions or tell of physical impossibilities; and, second, those that involve divine predication—that is, those that assign human attributes to God. For example, though God is omnipresent, we read that Moses entered into the thick darkness where God resides (Exodus 20:21), and though God does not possess human form, we are told that “the sound was heard of the Lord God walking” (Genesis 3:8).

This problem is as old as the Scriptures themselves. The initial reaction to such passages was to interpret them literally, taking each story at face value. This, however, proved unacceptable to many as it involved abandoning certain pre-theoretic guiding principles, in particular, consistency—How can God both be omnipresent and possess human form? Some even suggested that if a passage was particularly problematic we should question its authenticity. But the stock of such cases was huge and pervasive. Eventually, the widespread dissatisfaction with the literalist approach served as the impetus for the development of alternative methods of interpretation.

By the time Christianity emerged, Jewish hermeneutics had evolved to a point where scholars found themselves divided roughly into one of three schools: Midrashic, Pesher and Allegorical (Jeanrond, Theological Hermeneutics 16-17). Of particular interest to our investigation is the third. Werner Jeanrond, in his historical account of
theological hermeneutics, writes,

*Allegorical Interpretation* of the Scriptures attempt[s] to understand the spiritual meaning of a text with the help of perspectives not derived from the text and usually not even informed by it. Rather the text is understood symbolically: it points beyond itself to a deeper reality. (*Theological Hermeneutics* 17, insert mine)³

The first theologian to campaign actively and effectively for an allegorical approach to Scriptural interpretation was Philo of Alexandria (c. 25 B.C.-50 A.D.).⁴

In his writings, Philo claims that nothing in the Scriptures is insignificant or misplaced. Rather, he argues, almost everything in the Scriptures is intended to be interpreted allegorically. He regards as overly conservative those exegetes who see only what is on the surface of the texts, adding that they are inflicted by "partial blindness of the soul's eye" (*The Worse is wont to ambush the Better* 22, qtd. in Williamson, *Jews in the Hellenistic World: Philo* 152). Philo builds his case for allegorical interpretation partly on the claim that the authors of the Holy texts wrote while possessed by the divine Spirit. To such a person "nothing is unknown" because he possesses "a spiritual sun and unclouded rays to give him a full clear apprehension of things unseen by sense but

---

³ In contrast to the Allegorical school, both the Midrashic and the Pesher schools stressed the need for "spiritual" interpretation. The Midrashic school was founded on seven rules, developed by Rabbi Hillel (1st century B.C.), which stressed the role of context in any effort to interpret a passage whose true meaning transcends its immediate obvious meaning. Members of the Pesher school simply claimed special insight into divine mysteries. See Jeanrond, *Theological Hermeneutics* (16-17).

⁴ Although Philo is often credited as the first to apply allegorical interpretation to Scriptural texts, Ronald Williamson argues that this is not justified (*Jews in the Hellenistic Tradition: Philo* 147). Williamson cites passages in Philo's writings where Philo alludes to the use of allegorical interpretation by his predecessors, partly as a justification for his own application of this method. Nonetheless, Williamson agrees that Philo was unique both in the scope and the quality of his allegorical exegesis.

Philo’s writings proved influential. His allegorical approach to Scriptural interpretation was met with agreement by many important Jewish and Christian theologians. St. Augustine (354-430), for one, agrees with the spirit of Philo’s approach. Augustine tells us that he was inspired by Ambrose, Bishop of Milan, to apply figurative interpretation to Scripture. He writes of a speech given by Ambrose:

> Above all, I heard first one, then another, then many difficult passages in the Old Testament scriptures figuratively interpreted, where I, by taking them literally, had found them to kill. So after several passages in the Old Testament had been expounded spiritually, I now found fault with that despair of mine, caused by my belief that the law and the prophets could not be defended at all against the mockery of hostile critics. (*Confessions* V, xiv)

Though they share important similarities, there is a notable difference between Philo’s allegorical approach and Augustine’s use of figurative interpretation. Specifically, whereas Philo argues that “all or most of the law-book is an allegory” (*On Joseph* 28, qtd. in Williamson, *Jews in the Hellenistic World: Philo* 158), Augustine stops short of endorsing a wholesale figurative interpretation of Scriptural texts. Rather, Augustine’s approach centres on an effort to synthesize the two methods, literalist and figurative, allowing for the figurative interpretation of certain passages, but not all (*On Christian Doctrine* II & III). He agrees with Philo that we must not take passages intended figuratively as literal, but warns “we must also pay heed to that which tells us not to take a literal form of speech as if it were figurative” (*On Christian Doctrine* III, x, 14).

This leads naturally to a question of central importance: how does one determine
whether a passage is intended literally or figuratively? Augustine tells us:

Whatever there is in the word of God that cannot, when taken literally, be referred either to purity of life or soundness of doctrine, you may set down as figurative. Purity of life has reference to the love of God and one’s neighbour, soundness of doctrine to the knowledge of God and one’s neighbour. (On Christian Doctrine III, x, 14)

Put succinctly, according to Augustine, first we should consider the literal interpretation of a given passage; if it fails to meet his criteria of being consistent with “purity of life and soundness of doctrine,” a figurative interpretation should be sought out. Passages of the kinds described at the beginning—that is, those that contain apparent contradictions or tell of physical impossibilities and those that involve divine predication—would, by Augustine’s criteria, be interpreted figuratively.

But why would the authors of the Holy texts choose to convey their message figuratively? Augustine answers: “it is pleasanter in some cases to have knowledge communicated through figures [because] what is attended with difficulty in the seeking gives greater pleasure in the finding” (On Christian Doctrine II, vi, 8, insert mine). St. Thomas Aquinas (c.1224-1274) agrees that some Scriptural passages must be interpreted figuratively. But, in contrast to Augustine, he argues that this is not just a matter of style; rather, the use of metaphors in Scripture is “necessary” (Summa Theologicae I, 1, 9ad.1). The use of metaphors is necessary because, for example, God is utterly unlike anything on earth and so it is impossible to apply (literal) earthly attributes to Him. Metaphors are indispensable in our quest to understand God because they allow us to employ characteristics that are familiar—that is, they help us make sense of Him in terms we understand. Aquinas writes,
So it is that all names applied metaphorically to God are applied to creatures primarily rather than to God, because when said of God they mean only similitudes to such creatures. For as *smiling* applied to a field means only that the field in the beauty of its flowering is like to the beauty of the human smile by proportionate likeness, so the name *lion* applied to God means only that God manifests strength in His works, as a lion in his. (*Summa Theologiae* I, 13, 6)

The employment of figurative interpretation in Scriptural exegesis represents a significant turning point in our investigation. Though Plato, Aristotle and early rhetoricians employ the distinction between literal and non-literal language in their effort to separate language that is clear from language that is obscure, this is a project that is of philosophical concern only indirectly. In contrast, the interpretive projects undertaken by theologians of the above sort depend importantly on being able to make sense of such a distinction. Put succinctly, a coherent understanding of the concept of God requires that we are able to distinguish between characteristics attributed to Him literally and characteristics attributed to Him only figuratively. One’s approach to Scriptural interpretation, then, will be informed by answers to several important questions. How is a literal interpretation established? What distinguishes a literal interpretation from a non-

---

5 Note that the figurative approach to Scriptural exegesis finds long standing support during subsequent centuries. For example, Leibniz makes the following case for the figurative interpretation of certain passages: “[S]uppose that on the one hand we have the literal sense of a text from Holy Scripture, and that on the other we have a strong appearance of a logical impossibility or at least a recognized physical impossibility; then is it more reasonable to give up the literal sense or to give up the philosophical principle?” (*New Essays on Human Understanding* IV, xviii, 9). Answering his question, Leibniz writes, “There are certainly passages where there is no objection to abandoning the literal sense—for instance, where Scripture gives God hands, or attributes to him anger, repentance and other human affects.” Others who acknowledge explicitly the useful role of figurative interpretation in understanding Holy texts include Descartes (“Comments on a Certain Broadsheet” 309), Locke (*An Essay Concerning Human Understanding* III, ix, 9), Berkeley (“Alchiphron” VI, 7), Spinoza (*Letters: July 1664-September 1665* 21) and Mill (“Coleridge” 144).
literal one? What (if any) are the rules for the application of figurative interpretation?

Of the above theologians, Augustine comes closest to providing answers to these questions. To begin, he distinguishes between language that is proper and language that is figurative: “[Signs] are called proper when they are used to point out the objects they were designed to point out... [and] figurative when the things themselves which we indicate by the proper names are used to signify something else” (On Christian Doctrine II, x, 15, inserts mine). Unfortunately, this characterization of the distinction raises more questions than it answers. In particular, Augustine leaves unanswered the question of how it is determined what a sign is “designed” to point out. Is it by convention, by individual decree, by some recognized authority, or by some other procedure? Each of these answers will result in very different sets of literal and non-literal language. Just as important, Augustine never tells us how to treat apparently literal terms that “point” to nothing—for example, adjectives, numbers, logical connectives, and so forth.

So, though with Scriptural hermeneutics we encounter our first philosophical project for which it is of seemingly fundamental importance, the exact nature of the distinction between literal and non-literal language remains largely a mystery. Moreover, almost as soon as non-literal language acquires a degree of philosophical respectability, we find it again under attack for its perceived lack of clarity and apparent capacity to deceive.

Modern Philosophy

In modern philosophy the most sustained attacks on non-literal language are found in the
writings of Thomas Hobbes (1588-1679), John Locke (1632-1704) and Gottfried Wilhelm von Leibniz (1646-1716). To understand the motivation for their disapproval of non-literal language, we need first to understand how they envision the role of language.

As much as they might disagree about fundamental issues such as the origins of human knowledge, there is a common thread that runs through the various dominant philosophies of the modern era: language is a medium through which we communicate ideas. Empiricists and Rationalists agree, on the one hand, that ideas constitute knowledge, and on the other, that language is a social device that we use to share and explicate our ideas. Hobbes explains that the role speech plays is that of the medium “whereby men register their Thoughts; recall them when they are past; and also declare them to one another for mutual utility and conversation” (Leviathan I, iv). In the same spirit, Locke says that the three chief purposes of language are, “first, to make known one man’s thoughts or ideas to another; secondly, to do it with as much ease and quickness as possible; and, thirdly, thereby to convey the knowledge of things” (An Essay Concerning Human Understanding III, x, 23). Leibniz, borrowing explicitly from Locke, writes, “words serve (1) to make our thoughts understood, (2) to do this with ease, and (3) to provide a way into the knowledge of things” (New Essays on Human Understanding III, x, 23).

Given that the role of language is the communication of ideas, it is not surprising that central to the discussions on language entered into by all three philosophers are accounts of the various misuses of language. Hobbes discusses at length assorted types of “absurd assertions” that constitute the core of improper uses of language (Leviathan I,
v). Likewise, both Locke and Leibniz write about the various "abuses of words," each of which marks a failure in one of the above three stated purposes of language (An Essay Concerning Human Understanding III, x, 34; New Essays on Human Understanding III, x). Perhaps not surprisingly, all three philosophers locate non-literal uses of language within the scope of these misuses. Hobbes writes that absurd assertions include "the use of metaphors, tropes, and other rhetorical figures, instead of words proper" (Leviathan I, v). Similarly, Locke includes non-literal language among his list of the abuses of words: "[All the artificial and figurative application of words eloquence hath invented, are for nothing else but to insinuate wrong ideas, move the passions, and thereby mislead the judgment... they are certainly, in all discourses that pretend to inform or instruct, wholly to be avoided" (An Essay Concerning Human Understanding III, x, 34). Leibniz, quoting directly from Locke, writes, "all the art of rhetoric, besides order and clearness, all these artificial and figurative applications of words, serve only 'to insinuate wrong ideas, move the passions, and... mislead the judgement'" (New Essays on Human Understanding III, x, 34).

As with Aristotle's discussion of the distinction between standard and strange words, however, the topic of non-literal language plays but a limited role in Hobbes's, Locke's and Leibniz's accounts of the proper and improper uses of language. In all three accounts the central motivation is to ferret out the various ways in which language can detract from the straightforward communication of clear ideas. Hobbes writes,

---

6 Notice the similarity between this and Plato's claim that poetry "irrigates and tends to [the passions] when they should be left to whither, and... makes them our rulers when they should be our subjects" (quoted on p.14 above).
The Light of human mind is Perspicuous Words, but by exact definitions first snuffed, and purged from ambiguity; Reason is the pace; Increase of Science, the way; and the Benefit of man-kind, the end. And on the contrary, Metaphors, and senseless and ambiguous words, are like ignes fatui; and reasoning upon them, is wandering amongst innumerable absurdities; and their end, contention, and sedition, or contempt. (Leviathan I, v)

In his discussion, Hobbes lists seven cases in which language misleads us. Interestingly, only one of these cases concerns non-literal language exclusively: "[I]n reckoning, and seeking of truth, such speeches [i.e., metaphors, tropes, and other rhetorical figures] are not to be admitted" (Leviathan I, v, insert mine). In each of the other cases of misleading language, literal language is a possible, even likely, vehicle—for example, Hobbes warns against the use of terms which lack proper definition, and against the use of terms which signify nothing (Leviathan I, v.). Locke and Leibniz also each list seven cases of the abuse of words. Again, almost all of these constitute misuses of literal language. For example, Locke complains that people sometimes apply names “very unsteadily, making them stand now for one and by and by for another idea” and other times use names before learning what ideas correspond to them (An Essay Concerning Human Understanding III, x, 23). Leibniz shares Locke’s complaints, and adds examples such as when “men, having by long-standing usage attached certain ideas to certain words, imagine that the connection is an obvious one and that everyone accepts it” (New Essays on Human Understanding III, x, 22). As with Hobbes, only once in each of Locke’s and Leibniz’s lists of abuses of language does any mention of non-literal language appear.

To our disappointment, none of the three philosophers takes pains to explicate
what exactly the distinction between literal and non-literal language is. Thus, we are in
the frustrating position of being left without a satisfactory account of what constitutes the
literal (and, so, non-literal) use of a term. Instead, as with Aristotle and early
rhetoricians, we are led into a discussion of the distinction between the proper use of
language—roughly that which expedites the accurate communication of ideas—and
improper use of language—that is, all else. Though non-literal language falls entirely
on the side of improper uses of language, it shares this space with numerous improper
literal applications of language. So, even with their rather detailed discussions of proper
and improper uses of language, Hobbes, Locke and Leibniz afford us few if any clues to
their conception of the distinction between literal and non-literal language.

This shortage of insight is compounded by the fact that no other philosopher of the
time offers the sort of explicit treatment of language found in Hobbes, Locke and
Leibniz. To be sure, other philosophers allude to the distinction between literal and
figurative language. For example, René Descartes (1596-1650), in the context of a
discussion about the possibility of something causing its own existence, writes, “There
are some who attend only to the literal and strict meaning of the phrase ‘efficient cause’”
(“Objections and Replies” 79). Elsewhere he criticizes another author for “his frequently

---

7 This deserves qualification. Like Aristotle, all three philosophers condemn the use of non-
literal language in the context of reasoning or informative discourse. This doesn’t address
the question of the applicability of figurative language in other, un-philosophical pursuits. Indeed,
both Locke and Leibniz acknowledge a useful role for non-literal language. Locke writes, “I
confess, in discourses where we seek rather pleasure and delight, than information and
improvement, such ornaments as are borrowed from [figurative speeches and allusion] can scarce
pass for faults” (An Essay Concerning Human Understanding III, x, 32, insert mine). Similarly,
Leibniz writes that figurative language “is acceptable in discourses where the aim is merely to
give pleasure” (New Essays on Human Understanding III, x, 34).
resorting to irony” (“Comments on a Certain Broadsheet” 307). George Berkeley (1685-1753), in reference to the notion that matter supports extension, writes, “It is evident support here cannot be taken in its usual or literal sense—as when we say that pillars support a building” (“The Principles of Human Knowledge” I, 16). Elsewhere he warns, “a philosopher should abstain from metaphor” (Of Motion 31). Jeremy Bentham (1748-1832) employs the distinction in an effort to explain the difference between the “literal or unfigurative sense” and the “figurative sense” of “motive” (An Introduction to the Principles of Morals and Legislation X, iv). Nonetheless, even a simple explicit statement of what constitutes the distinction between literal and non-literal language is absent from the writings of these philosophers, as it is from the writings of many other central figures of the era, including those of David Hume (1711-1776) and Benedict de Spinoza (1632-1677).

The badly bruised reputation of non-literal language finds some relief in the writings of Immanuel Kant (1724-1804) and John Stuart Mill (1806-1873). Although Kant does not discuss non-literal language explicitly, he pays homage to the role of metaphor in his discussion of the imagination. With metaphor, he writes, the poet endeavours “to go beyond the limits of experience and to present them to sense with a completeness of which there is no example in nature” (Critique of Judgement §49). He adds,

The imagination (as a product of faculty of cognition) is very powerful in creating another nature, as it were, out of the material that actual nature gives it. We entertain ourselves with it when experience becomes too commonplace, and by it we remould experience, always indeed in accordance with analogical laws, but yet also in accordance with principles which occupy a higher place in reason (laws, too, which are just as natural
to us as those by which understanding comprehends empirical nature).  
(Critique of Judgement §49)

Whereas Hobbes, Locke and Leibniz argue that anything which needs to be said can be said using proper language—i.e., some appropriate subset of literal language—Kant acknowledges that sometimes the artist reaches after something “which lies beyond the bounds of experience” and, so, beyond the bounds of literal language (Critique of Judgement §49). Indeed, according to Kant the mark of genius consists in a special relation between the faculties of understanding and imagination whereby the imagination, though submitting to the constraints placed on it by the faculty of understanding (namely, our stock of concepts and techniques for relating these concepts), finds hitherto unnoticed relations between concepts, and finds them without the aid of a set procedure. One way to achieve this is through a metaphor whereby, using the very language whose bounds render it incommunicable literally, one conveys a original idea by comparing it to that which is familiar. Despite this obvious respect for metaphors, Kant does not attribute cognitive import to them. Metaphors, he would argue, strictly speaking do not contribute to the stock of knowledge because they depend neither on rules nor on definite concepts, but rather on noticing relations between existing concepts. Nonetheless, they, as well as the other tools of the artist, are not merely ornamental as they afford us refreshing insight into the workings of the world around us.

Like Kant, Mill recognises a positive role for metaphor. Though his remarks are brief, they represent an important stage in the defence of metaphor as philosophically significant. Broadly, Mill envisions a potential for metaphors as legitimate heuristic devices. Apt metaphors, he writes, “assume the proposition which they are brought to
prove: their use is, to aid the apprehension of it" (*A System of Logic* V, v, 7). Thus, though a metaphor cannot itself constitute the proof of a proposition, it can play an important role in directing us to such a proof. Mill cites the metaphor "if the soil is left uncultivated, weeds will spring up" as an argument for education: "For the reason why weeds grow in an uncultivated soil is that the seeds of worthless products exist everywhere, and can germinate and grow in almost all circumstances, while the reverse is the case with those which are valuable" (*A System of Logic* V, v, 7). He adds that there is significant value in this approach to reasoning: "This mode of conveying an argument, independently of its rhetorical advantages, has a logical value, since it not only suggests the grounds of the conclusion, but points to another case in which those grounds have been found, or at least deemed to be, sufficient." An apt metaphor, then, though not an argument, can serve as an indication that an argument exists.

**Summing Up**

What have we learned thus far from this brief historical investigation? Perhaps the most important lesson is that among the various competing claims about the appropriateness of non-literal language to tasks ranging from poetic discourse to scriptural exegesis to philosophical argument, one common feature stands out: save for Aristotle’s brief definition of metaphor (and its echo in Cicero), Quintilian’s shopping list of tropes, and Augustine’s flawed definitions of literal and figurative terms, nowhere is the distinction between literal and non-literal language discussed explicitly. In each of the above projects, including Aristotle’s, Quintilian’s and Augustine’s, the nature of the distinction
between literal and non-literal language remains unspecified. What are we to make of this? Surely these projects depend on this distinction. How are we to heed the warnings against the misuse of non-literal language issued by Plato, Aristotle, Cicero, Quintilian, Hobbes, Locke, Leibniz and others if we cannot identify exactly what is to count as a breach of literal use? Can we follow the advice of Philo, Augustine and Aquinas and interpret difficult Scriptural passages metaphorically if we have no criteria of identity for metaphor? Similarly, what use are Kant’s and Mill’s praises for non-literal language if we are unable to say what constitutes an instance of figurative language?

Before we reward ourselves for having uncovered an apparent oversight shared by many of the great minds of the last two and a half millennia, we need to re-examine the force of the claims underwritten by the distinction. Plato, Aristotle, Cicero, Quintilian and modern philosophy as represented by Hobbes, Locke and Leibniz, strive to rid serious, informative discourse of any rogue elements such as nonsense, ambiguity, vagueness and, yes, figurative language. But here the distinction between literal and figurative language may, without risk to the general point, rest on rough pre-theoretic grounds. That is, for these projects to succeed we need not agree in advance on all possible candidates and their appropriate classification as literal or non-literal. Put differently, nothing in these discussions rests fundamentally on our ability to make a rigorous distinction between literal and non-literal language. Rather, the common underlying message is simple and practical: use language that is clear and concise, and know whereof you speak. To be sure, there may be disagreement about particular cases. Hobbes, for example, cites “the way goeth” as an example of a metaphor (Leviathan I,
v). Locke and Leibniz may disagree with him on the suitability of this example, but nevertheless, such cases can be decided individually without harm to the more general aim of perspicuity.

What of Scriptural exegesis, Kant's employment of figurative language in his discussion of the imagination and Mill's appropriation of metaphor in philosophical argument? It was noted already that of all the projects discussed above these are special because they assign philosophical relevance to figurative language. They depend fundamentally on the nature of the distinction between literal and non-literal language, do they not? Granted, there are certain points about the nature of the distinction between literal and non-literal language that are important to these projects. Specifically, it is critical that such a distinction exists at some level. But that the distinction be made precise—that is, that we find deciding criteria for literal and non-literal language—is not at issue. Consider the case of Scriptural exegesis. For Philo, allegorical interpretation is applied throughout, with little or no regard for the feasibility of literal interpretations. For Augustine and Aquinas, the application of figurative interpretation to Scriptural texts is reactionary—that is, it is employed in response to passages that for one reason or another cannot be interpreted literally. These points are important. They mean that the Scriptural exegete need not carry with her a handbook of rules for identifying figurative language, stopping periodically to check whether particular passages qualify. Rather, in the case of Philo, an allegorical interpretation is applied automatically; in the case of Augustine and Aquinas, the decision to interpret a particular passage as an allegory or as containing metaphor depends largely on broader considerations such as its relevancy and
coherence—that is, generally, if a literal interpretation is deemed untenable, then a figurative one is applied. But just as I might treat a newspaper article as a joke without being prepared to offer rigorous criteria for jokes, I can treat a Scriptural passage as allegorical or as containing metaphor based on similarly rough understandings of these notions. Thus, though the notion of figurative language is important to Scriptural exegesis, the outcome of the investigation into the nature of the distinction between literal and non-literal language bears no more weight here than it does for Aristotle and others who want simply to separate language that is clear from language that is obscure.

A similar point can be made with respect to Kant's use of metaphor in his discussion of the faculty of the imagination, and with respect to Mill's argument in favour of metaphors as heuristic devices. Again, these accounts get by on rough understandings of the distinction between literal and figurative. For Kant, one might even settle for a rough teleological criterion: if the poet succeeds in presenting reality in an utterly novel way, then he necessarily uses figurative language; for to transcend the bounds of our current mode of conceptualization, one must transcend the limits of literal language. Mill's point is simpler: by means of an apt comparison, possibly through metaphor, one can bolster one's argument by pointing to an analogous case for which an argument is already in place (or, at least, is apparent). For neither philosopher is it critical that they demarcate clearly between literal and figurative language. In fact, none of the projects undertaken by the above philosophers, from Aristotle to Kant, depends fundamentally on being able to make a principled distinction between literal and non-literal language. To be sure, these projects depend on there being a distinction between literal and figurative
language, and for this we can fault their inattention to explicating this distinction, but they survive even if the distinction rests on crude pre-theoretic grounds.

Indeed, there may be another explanation as to why none of the above philosophers, with the exception of Augustine, attempts to make explicit the distinction between literal and non-literal language. Perhaps it is not just because the distinction is, at best, of secondary interest to them, but also because they consider it a task for rhetoricians, not for philosophers. Even the extensive manuals of rhetoric published in the mid nineteenth century, however, contain few clues as to how the distinction between literal and non-literal language is properly construed. Noted rhetorician George Campbell writes that figures of speech "are but different modes of exhibiting a comparison... [whereby] without anything like an explicit comparison, and commonly without warning or apology, the name of one thing is obtruded upon us, for the name of another quite different, though resembling in some quality" (The Philosophy of Rhetoric 285, insert mine). But this vague characterization is nowhere supported by anything like a theory of literal meaning or literal comparison. Rather, like his predecessors, Campbell assumes the distinction and focuses on the grander task of counselling us on the proper employment of language. Similarly, Campbell’s contemporary, Richard Whately, makes short work of the distinction between literal and figurative language, noting that a trope is "any word turned from its primary signification" (Elements of Rhetoric 195). Again, we are left wondering, among other things, what notion of primary signification underwrites Whately’s definition.

So, again, what have we learned? At the very least, whereas we might have been
tempted previously to think either that the distinction between literal and non-literal language is historically well documented or that it is of central interest in the Western philosophical tradition, we can now conclude, for better or for worse, that this is not the case. Moreover, though many important projects depend on the existence of the distinction, and though it appears that no-one adequately explicates the nature of the distinction, the ground on which these projects rest is nonetheless stable (in this respect anyway). But before we answer the question, ‘Why concern ourselves with the distinction between literal and non-literal language?’ with a resigned ‘We needn’t’, we must extend our historical investigation to particular trends in late nineteenth and early twentieth century philosophy. Afterwards we will find ourselves issuing a remarkably different answer.
Whereas the projects of the philosophers discussed in chapter one depend importantly, but not essentially, on being able to sustain a distinction between literal and non-literal language, certain projects in twentieth century analytic philosophy rest *fundamentally* on the assumption that the argument for a principled distinction has been made or is nearby. The arrival of analytic philosophy in general, and of the semantic study of natural language in particular, signals a new era of importance for the distinction between literal and non-literal language.

In contrast to their predecessors, analytic philosophers locate human conceptualisation at the level of language. According to this approach, it follows that understanding our conceptual framework is best achieved through the analysis of our language. Summing up the impetus behind linguistic analysis, Michael Dummett writes, "[U]ntil we have first achieved a satisfactory analysis of the meanings of the relevant expressions, we cannot so much as raise the questions of justification and truth, since we remain unclear about what we are attempting to justify or what it is about whose truth we are enquiring" (Dummett, *Frege* 667). Thus analytic philosophers make it their task to answer certain fundamental questions about language: *What is the nature of linguistic meaning? What is the relation between words and sentences and the things they signify, or the facts they describe?*

To be sure, analytic philosophers are not the first to cite the importance of the analysis of language in the pursuit of philosophical understanding. In Plato’s *Cratylus*
(c.390 B.C.) we find Socrates arguing that only an adequate language can serve as a vehicle for framing and communicating knowledge. In modern philosophy, Locke, for example, appeals to our use of language in his discussion of the relation of matter to body: “But we see that though it be proper to say, ‘There is one matter for all bodies’, one cannot say, ‘There is one body for all matters’: we familiarly say, ‘One body is bigger than another’; but it sounds harsh (and I think is never used) to say, ‘One matter is bigger than another’” (An Essay Concerning Human Understanding III, x, 15).

Likewise, David Hume (1711-1776) anticipates twentieth century analytic philosophy: “To convince us of this proposition, that where there is no property, there can be no justice, it is only necessary to define the terms, and explain injustice to be a violation of property” (An Enquiry Concerning Human Understanding XII, iii). Nonetheless, the examples offered by Locke and Hume concern metaphysical and jurisprudential matters—that is, the answers to their questions necessarily lie beyond the scope of empirical investigation. Neither philosopher counsels an appeal to linguistic analysis in the case of empirical claims. In contrast, according to a certain strand in the analytic tradition, it is not just metaphysical disputes that are settled through the analysis of language, but certain kinds of empirical ones as well—for example, some analytic philosophers would argue that to understand what a volcano is, or what the colour red is, we must direct our attention to what the terms ‘volcano’ and ‘red’ mean.¹

According to analytic philosophy, language is not simply the medium through

¹ This strand of analytic philosophy (to be discussed in more detail in chapter three) includes works by Frege, Russell, the early Wittgenstein and the Positivists.
which we *communicate* our understanding of the world, it is the medium that makes understanding possible in the first place. In contrast to the empiricism of Hobbes, Locke and others, analytic philosophers deny the claim that ideas are primary and language secondary—that is, that first we have ideas, then we find ways to put them into language—rather, their claim is that there is no conceptualisation without language; enquiry into extra-linguistic reality is beyond both practical and theoretical bounds. So if we want to understand how we conceptualise our environment we need a proper analysis of our language. Naturally, then, the focus of analytic philosophy is largely on what and how bits of language *mean*.

Hence, despite their differences, logical positivists, Wittgenstein (throughout his career), ordinary language philosophers, Quineans, Davidsonians and others in the analytic tradition all agree that philosophical enquiry necessarily involves an analysis of the ways we use language. In his book, *Origins of Analytical Philosophy*, Dummett writes,

> What distinguishes analytical philosophy, in its diverse manifestations, from other schools is the belief, first, that a philosophical account of thought can be attained through a philosophical account of language, and, secondly, that a comprehensive account can only be so attained. (4)

Fundamental to this new approach to philosophy, then, is the idea that the meaning or structure of a language can be examined and explored in a manner in which extra-linguistic reality *per se* cannot. It is no wonder, then, that Russell, Wittgenstein, Carnap, Quine, Davidson and other philosophers who follow in this tradition, make it among their primary tasks to explore the nature of linguistic meaning.
The Grandfather of Analytic Philosophy: Gottlob Frege

Many would agree that the grandfather of analytic philosophy is Gottlob Frege. Indeed, some locate the seed of analytic philosophy in a specific passage in Frege’s *The Foundations of Arithmetic.* At a critical juncture in his investigation—specifically, during his quest for a criterion for numerical identity—Frege raises the question, ‘How are numbers given to us, granted that we have no idea or intuition of them?’ His answer is astounding: “Since it is only within the context of a proposition that words have any meaning, our problem becomes this: To define the sense of a proposition in which a number word occurs” (Frege, *The Foundations of Arithmetic* §62). According to Frege, to get at the structure and content of our thoughts we must examine the structure and content of our sentences: “The sentence can be regarded as an image of the thought in that the relation between the part and the whole within the thought there by and large corresponds [to] the same relation between the part of the sentence and the sentence” (*The Foundations of Arithmetic* §62). Thus Frege, perhaps unwittingly, shifts the focus of his philosophical enquiry from thoughts or ideas to language. Commenting on this move, Dummett writes,

[T]he investigation therefore takes the form of asking how we can fix the senses of sentences containing terms for numbers. An epistemological enquiry (behind which lies an ontological one) is to be answered by a linguistic investigation. (*Origins of Analytical Philosophy* 5).

Interestingly, Frege offers no explicit acknowledgement of this methodological shift, let alone a justification for it.

---

2 See, for example, Dummett, *Origins of Analytic Philosophy* (5).
In any case, whether Frege is or is not to be credited with the origins of analytic philosophy is not of importance to our investigation. What is of importance is that his method of logical analysis, outlined below, serves as the foundation upon which analytic philosophy is built. Accordingly, we would do well to delve both into its nature and into the motivations that underlie it.

Frege’s Objective: Securing The Foundations of Mathematics

Though he is known among philosophers mainly for his work in formal logic and in the philosophy of language, Frege’s primary interest was in the foundations of mathematics. Indeed, he began his career at a time when mathematics was undergoing radical and extensive changes. These changes prompted philosophically minded mathematicians to subject existing mathematical concepts and proofs to unprecedented levels of critical scrutiny. Indeed, some opined that the entire field was in need of philosophical reassessment. G.P. Baker and P.M.S. Hacker reflect, “the general philosophical question about the true sources of mathematical knowledge became urgent and unavoidable.” (Frege: Logical Excavations 33).

---

3 In his introduction to Frege’s philosophy, Anthony Kenny explains: Euclidean geometry, which had been regarded as a system of necessary truths for over two millennia, lost its unique status early in the nineteenth century. Euclid had derived the theorems of his system from five axioms: it was now shown that one of these axioms, far from being a necessary truth, could be denied without inconsistency, and non-Euclidean geometries were developed on the basis of alternative axioms. (Kenny, Frege 2)

In addition to the emergence of new geometries, there were important developments in number theory. Imaginary numbers, for example, though considered during the eighteenth century to be an eccentric curiosity, were now shown to serve a purpose in the representation of motion in a plane.
Frege was among those troubled by the state of mathematics. Though impressed by the emerging theories, he cautioned that they rested on ill understood foundations. Reflecting on his work late in his life, he wrote, "the most pressing need, it seemed to me, was to provide arithmetic with a better foundation" ("Notes for Ludwig Darmstaedter" 253). In his opinion, mathematicians had yet to explain satisfactorily some of the most basic concepts of their discipline—for example, the philosophical foundation of mathematical induction and the nature of natural numbers. Frege believed that only by making explicit the nature of these basic concepts could we assess the legitimacy of more complex theories. As a result, he devoted his career to setting out the logical and philosophical foundations of mathematics.⁴

*Frege's Logic*

Frege's entire treatment of the problem of the foundations of mathematics is motivated by his claim that mathematical proofs depend on the same valid forms of reasoning that govern all modes of thinking. He argues that mathematical inferences are based not on laws particular to mathematics, but on the general laws of logic. Moreover, he argues that the foundations of mathematics can be formalized without appealing to any non-logical concepts or axioms. That is, he contends not only that concepts such as

---

⁴ His results are contained in three major works. The first of these, discussed below, is *Begriffsschrift* (1879), in which he develops his system of logic. This was followed by *The Foundations of Arithmetic* (1884) in which he attempts to demonstrate that the concepts of a number and of numbers are definable in purely logical terms. The culmination of his thinking, however, appears in two volumes of *The Basic Laws of Arithmetic* (1893, 1903) in which he attempts (ultimately unsuccessfully) to give a precise demonstration of his thesis that mathematics is reducible to logic.
mathematical induction can be derived entirely from the general laws of logic, independent of any considerations peculiar to mathematics, but also that mathematical concepts are not in any way empirical. In particular, he denies that mathematical concepts are generalizations about our encounters with the world.

Further, in contrast to many of his predecessors, Frege denies that valid forms of reasoning are merely reflections of normal mental processes. Indeed, he earned a reputation as an ardent critic of psychologism, the doctrine that the laws of logic are determined by the nature and structure of the human mind. He tells us, rather, that logic is founded on the laws of truth, not of assertion or thinking ("The Thought" 18). He writes, "It is not the holding of something to be true that concerns us but the laws of truth" ("Logic" 145). Whereas assertion and thinking take place in accordance with psychological laws, Frege writes, "If being true is... independent of being acknowledged by somebody or other, then the laws of truth are not psychological laws" (The Basic Laws of Arithmetic 1 xvi). The laws of truth, like natural laws, exist independently of humans and are to be discovered, not invented.

Believing that "the firmest method of proof is obviously the logical one" because it is topic-neutral and "based solely upon the laws on which all knowledge rests," Frege set as his first task the development of a new symbolism designed to make perspicuous the logical relationships that lie concealed in natural language (Begriffsschrift, Preface). It was these laws, he believed, that would ultimately inform our understanding of the foundations of mathematics. The result was a new formal system called the concept-
script, elucidated in his first major publication, *Begriffsschrift* (1879).5

Frege’s concept-script was revolutionary. As a systematic treatment of logical relations, it was far more sophisticated than Aristotle’s syllogistic logic, considered by philosophers as late as Kant to be “a closed and completed body of doctrine” (*Critique of Pure Reason* B, viii). In contrast to Aristotle’s syllogisms, Frege’s concept-script is an axiomatic system in which the laws of logic are derived formally, by specified rules of inference, from a set of primitive principles. Indeed, Frege’s concept-script is the first systematic formulation of the modern propositional calculus. Moreover, it includes quantification theory, the basis of our current systems of first-order logic.

Why is Frege’s concept-script of concern to our investigation? Because, as we will discover, his work in logic profoundly influenced his philosophy of language. Indeed, in a short work written in 1906 entitled “What may I regard as the result of my work,” he wrote, “It is almost all tied up with the concept-script” (185). Though the foundations of mathematics remained his primary interest, in a series of papers published in 1891 and 1892, Frege explored the implications of his work in logic to our understanding of natural language.6 Though perhaps considered ancillary by their author, these papers set out a highly original approach to the study of language, one that proved extraordinarily influential on the work of subsequent analytic philosophers.

5 Though the foundations of mathematics were his primary interest, Frege intended his notation to apply to every field which makes use of rigorous proofs, including all branches of science.

6 These are “Function and Concept” (1891), “On Sense and Reference” (1892) and “On Concept and Object” (1892). Other papers in which Frege discusses natural language include “The Thought: A Logical Inquiry” (1918) and the posthumously published “Logic” (believed to be written in 1897).
Logic and Language: Truth

Frege’s analysis of natural language is deeply influenced by his work in logic. He uses his concept-script as a standard against which to elucidate the properties and quirks of natural language. This is not to say that Frege intends his concept-script to replace natural language—indeed, though he writes, “work in logic just is, to a large extent, a struggle with the logical defects of language,” he adds, “[natural] language remains for us an indispensable tool” (“My basic logical Insights” 252, insert mine). The concept-script is designed by Frege to bring into sharp focus all and only those elements essential to the validity of proof. For this particular purpose, Frege deems natural language, with all of its irregularities, too cumbersome and too often misleading:

In place of ordinary language with its proliferating, indefinite logical forms, I put a small number of forms. This seems to me essential for making the movements of thought precise. (qtd. in Sluga, Gottlob Frege 79)

Indeed, Frege blames many of our errors in reasoning on the imprecise and often awkward logical forms of natural language.

Frege’s analysis of natural language focuses almost entirely on what he calls conceptual content. Conceptual content is meant to capture all and only those elements of natural language relevant to logical inference; this is why he calls his logical system

---

7 Whether Frege intended his concept-script to show up natural language as inferior is a matter open to debate. Baker and Hacker, for example, claim that “to a degree, [Frege] regarded as ‘logical defects’ in a language those features of it which fail to correspond with the articulations of his Begriffsschrift” (Frege: Logical Excavations 37, insert mine) whereas others, including Kenny, claim rather that “Frege was not under the illusion that his concept script was a new and perfect language which would show up natural languages as imperfect” (Frege 13).

8 See “Logic” (143).
the concept-script. He writes, "I have designated by conceptual content that which is of sole importance for me" (Begriffsschrift Preface). Indeed, he forswears expressing in the concept-script anything "which is without importance in the chain of inference" (Begriffsschrift Preface). Accordingly, when he turns his attention to natural language, his objective is to elucidate what exactly is relevant to inferential relations—that is, what exactly constitutes conceptual content.

To begin, Frege argues, our analysis of language must focus only on that which is capable of being true or false; all else is incapable of bearing on the validity of an argument. His analysis of natural language, then, begins with the question: what in natural language is capable of being true or false? Frege claims first that sentences, not words, are candidates for truth or falsity. This seems uncontroversial; words by themselves are incapable of being true or false. Nonetheless, though all sentences of the concept-script are, in virtue of being well-formed, candidates for truth or falsity, Frege observes that not all sentences of natural language are similarly candidates for truth or falsity.

According to Frege, we must differentiate between the sentences of natural language that are capable of being true or false and those that are not. In "The Thought" Frege contends that when considering matters of truth, "only those sentences in which we communicate or state something come into question" ("The Thought" 21). Thus, in

---

9 This is a corollary to Frege's famous context principle, first outlined in the introduction to Foundations of Arithmetic. According to the principle, and contrary to the received opinion, Frege argues that sentences, not words, are the basic unit of meaning: "Only in the context of a sentence has a word a meaning."

10 The exceptions, if they exist, would be one-word indicatives.
virtue of their structure, non-indicatives are not capable of being true or false. For example, the interrogative, 'Who is the tallest Marx brother?' does not put forth a particular claim about who is the tallest Marx brother, rather it asks who is the tallest Marx brother. Similarly, the imperative, 'Sing it loud!' is not a claim that you are singing loudly (or that you are singing softly), rather it is an order to sing loudly. Likewise, the optative 'May our armies conquer' is an expression of hope rather than a report a particular state of affairs. Indeed, Frege denies that imperatives, interrogatives and optatives are capable of being true or false ("The Thought" 21). Prima facie, then, it appears that Frege wants to identify the sentences of natural language that are capable of being true or false with indicatives.

But Frege denies that all indicatives of natural language are capable of being true or false. In particular, any sentence containing an expression that fails to refer is neither true nor false according to Frege. Though in a formally respectable language such as the concept-script every term possesses a reference, in natural language, Frege observes, "to be sure, exceptions to this regular behaviour occur" ("On Sense and Reference" 58). Among the referring expressions of natural language that may fail to refer are the terms of fiction—for example, 'Santa Claus', 'The Millennium Falcon' and 'Humpty Dumpty'. In addition, Frege offers examples of the sort, 'the celestial body most distant from the earth' and 'the least rapidly converging series' ("On Sense and Reference" 58). As might be expected, Frege regards it as a weakness of natural language that many referring expressions fail to refer ("On Sense and Reference" 70).

We are now in a position to answer the question of which sentences of natural
language are, according to Frege, capable of being true or false and, so, of figuring in inference. On Frege’s analysis, a sentence of natural language is true or false if and only if it is an indicative containing no referring expression that fails to refer.\(^\text{11}\) I will call this subset of natural language indicatives, indicatives,.

*Logic and Language: Sense*

According to Frege, because only indicatives, are capable of being true or false, only indicatives, are relevant to logical inference. But the matter of conceptual content is not yet settled; that is, truth and falsity are not all that is relevant to inference and, so, not all that is relevant to conceptual content. To understand why, we need to return again to the concept-script.

In the concept-script, Frege identifies a concept as a function whose value for an argument is a truth-value. This characterisation of a concept is, in Baker and Hacker’s words, “the innovation of the 1890s enabling Frege to complete the logicist programme” (*Frege: Logical Excavations* 278). That is, this characterisation of a concept provides the model for Frege’s analysis of natural language indicatives,.

The predicate expression ‘is red’, for example, stands for a function that takes an argument, say a swatch of cloth, to the True if the argument is red and to the False otherwise. As we have discovered, the truth-value of an indicative, is understood by Frege to be an integral element of its

\(^\text{11}\) Though he does not mention them specifically, Frege may also want to exclude value judgments from the realm of indicatives that are true or false. It is impossible to say what exactly his stance would be. If, on the one hand, he were to analyze value judgments from the perspective of their grammatical structure, then he would likely include them among sentences that are true or false. If, on the other hand, he were to analyze value judgments from the perspective of their contribution to rigorous enquiry, he would likely exclude them.
conceptual content. Frege realises, however, that truth-values cannot account for all of the roles that conceptual content must play. In particular, the truth-value alone tells us nothing of a sentence’s capacity to stand as that which is proved or proof-relevant. Consider, for example, that the conclusions of distinct proofs are not identical simply in virtue of sharing the same truth-value. Indeed, each distinct sentence possesses its own unique make up.

The same observation can be made of natural language. In “Function and Concept,” and later in “On Sense and Reference,” Frege introduces and illuminates a semantic distinction between the reference of a term or sentence and its sense. He begins his analysis with a discussion of referring expressions (“On Sense and Reference” 57f.). He asks: what does the term ‘evening star’ mean? Surely to know what ‘evening star’ means is (at least partly) to know what it designates—that is, to know that to which it refers. Frege agrees, but adds that if reference were all that mattered, then the sentences ‘The evening star is the evening star’ and ‘The evening star is the morning star’ would mean the same given that ‘the evening star’ and ‘the morning star’ designate the same object. But ‘The evening star is the evening star’ is true in virtue of its logical form, not in virtue of its content—that is, its truth is independent of any empirical considerations. By contrast, ‘The evening star is the morning star’ provides us with the information that ‘the evening star’ designates the same object as ‘the morning star’.

---

12 J. Alberto Coffa, in *The Semantic Tradition from Kant to Carnap*, points out that the words ‘*Sinn*’ and ‘*Bedeutung*’ (translated commonly as ‘sense’ and ‘reference’) occur in semantic systems before Frege’s, but adds that what distinguishes Frege’s use is that he regards these notions “as semantic categories, rather than as names of specific semantic objects or relations” (79).
Frege concludes that the difference lies in each expression's unique mode of representation, or *sense*. Though the expressions 'the evening star' and 'the morning star' refer to the same object, it is because of their different ways of referring to that object that the sentences 'The evening star is the evening star' and 'The evening star is the morning star' are semantically distinct. He concludes, "It is natural... to think of there being connected with a sign (name, combination of words, letter), besides that to which the sign refers, which may be called the reference of the sign, also what I should like to call the *sense* of the sign, wherein the mode of representation is contained" ("On Sense and Reference" 57).

Though Frege's discussion centres initially on words—or, more specifically, on referring expressions—because of their role in inference, it is indicatives, that interest him most. That is, though the doctrine of sense and reference is primarily a theory of referring expressions, because indicatives, alone are capable of being true or false, indicatives, are logically fundamental. The parts of the indicative—that is, the senses and references of the constituent expressions—are reached only by analysis.\(^\text{13}\)

What, then, according to Frege, constitute the sense and reference of an indicative? To begin, Frege contends that indicatives, express *thoughts*. For example, the sentence 'John Hinkley Jr. shot Ronald Reagan' expresses the thought that John Hinkley Jr. shot Ronald Reagan. But is the thought that a sentence expresses its sense or its reference? Frege argues that it cannot be the reference given that the substitution within an indicative, of terms with the same reference but different senses alters the

\(^{13}\) cf. Sluga, *Gottlob Frege* (134).
thought expressed: “we can see that in such a case the thought changes; since, e.g., the thought in the sentence ‘The morning star is a body illuminated by the Sun’ differs from that in the sentence ‘The evening star is a body illuminated by the Sun’” though the terms ‘The morning star’ and ‘The evening star’ designate the same object (“On Sense and Reference” 62). Thus, he concludes, since the thought expressed by a sentence cannot be its reference, it must be its sense.

What, then, constitutes the reference of a sentence? Frege first observes that we expect the reference of the sentence to be a function of the references of its constituent expressions: “The fact that we concern ourselves at all about the reference of a part of the sentence indicates that we generally recognize and expect a reference for the sentence itself” (“On Sense and Reference” 63). Further, because it is a function of the references of its parts, we expect the reference of an indicative to remain unaltered when we substitute co-referring expressions—say, ‘The tallest free-standing structure in the world’ for ‘The CN Tower’ in ‘The CN Tower is for sale’. Indeed, because of this, the reference of an indicative cannot be the thought it expresses (as indicated above, the thought expressed by an indicative may change when one substitutes co-referring expressions). The salient question, then, is what remains unchanged when one substitutes one co-referring expression for another in an indicative? Frege gives us the answer: the truth value of the indicative. To be sure, whatever is true of the CN Tower is true of the tallest free-standing structure in the world. He reasons, accordingly, that the

---

14 This is not to say that the substitution of one term for another with the same reference always results in a change of the thought expressed. For example, Frege writes, “It makes no difference to the thought whether I use the word ‘horse’ or ‘steed’ or ‘cart-horse’ or ‘mare’” (“The Thought” 23).
reference of an indicative, is its truth value.\textsuperscript{15}

Thus, Frege concludes, associated with each indicative, is a thought\textsuperscript{16} and a truth value. The thought is its sense, the truth value its reference. An indicative expresses a thought, it stands for or designates a truth value.

Importantly, Frege also differentiates between the sense of a sentence and its force. In the case of an assertion, for example, Frege distinguishes between the sentence asserted and the act of asserting it. The sense of the sentence is what is asserted and exists independently of the force attached to it by the speaker. Thus regardless of whether the an indicative is used to make an assertion, employed as the antecedent of a conditional, uttered in a play or used to issue a hypothesis, it expresses the same thought. Put simply, Frege argues that the thought expressed by an indicative exists independently of the particular use to which that indicative is put.

\textit{Frege's Anti-Psychologism Revisited}

How are we to understand the nature of Fregean senses in general and thoughts in

\textsuperscript{15} Strictly speaking, when we call an indicative true, we really mean that its sense—that is, the thought it expresses—is true. Indeed, Frege characterizes a thought as "something for which the question of truth arises" ("The Thought" 20). Rendering truth a property of thoughts, not of sentences, accounts for the many indicative sentences of natural language which, because of indexicals like 'him', 'now' and 'there', when removed from their context, fail to express a complete thought and, so, to refer. For example, strictly speaking, it is not the sentence 'There is someone who is wearing the same suit as you in the next room' that is true or false, but rather the thought expressed by it, which is supplemented accordingly with information about the identity of the speaker, the identity of the person to whom she is speaking, the spacial-temporal location of the utterance, and so forth. To Frege, indexicals are a further burden of natural language—they have no analogue in the concept-script.

\textsuperscript{16} Though Frege allows that an indicative, may, depending on the complexity of its structure, express more than one thought. A conjunction, for example, may express two thoughts.
particular? To begin, Frege asks for us: "Now is the sense of a sentence an idea?" ("The Thought" 19). That is, is the sense of a sentence the idea that I, as the speaker or as an interpreter, associate with it? In keeping with his ardent attack on psychologism, Frege's answer is a resounding 'No'. Just as he argues that the laws of truth are mind-independent, so too, he says, are senses in general, and thoughts in particular. Indeed, when he first introduces the concept of a thought, he adds in a footnote, "By thought I understand not the subjective performance of thinking, but its objective content, which is capable of being the common property of several thinkers" ("On Sense and Reference" 62).

In no uncertain terms, Frege argues that "the reference and sense of a sign are to be distinguished from the associated idea" ("On Sense and Reference" 59). He explains:

To assume that in the sentence 'The Moon is smaller than the Earth' the idea of the Moon is in question, would be flatly to misunderstand the sense. If this is what the speaker wanted, he would use the phrase 'my idea of the Moon'. (On Sense and Reference: 61)

Ideas change, both in clarity and in association, and are saturated with feelings—that is, they are subjective. In contrast, Frege's senses are objective and public; they are independent of our thinking. He writes, "A thought does not belong specially to the person who thinks it, as an idea does to the person who has it: whoever thinks it encounters it in the same way, as the same thought" ("Logic" 127). That is, unlike ideas, "one and the same thought can be grasped by many [people]" ("On Sense and Reference" 79, insert mine). Thus, the thought must be distinguished from the ideas that arise in people's minds as they grasp the thought.

Frege allows that "almost always, it seems, we connect with the main thoughts
expressed by us subsidiary thoughts which, although not expressed, are associated with our words, in accordance with psychological laws, by the hearer” (“On Sense and Reference” 75). He notes, for example, that a painter, a horseman and a zoologist will probably associate different thoughts with the name ‘Bucephalus’, but insists that nonetheless its sense remains constant (“On Sense and Reference” 59). Language encodes not ideas, conceived as mental objects, but thoughts and concepts conceived as mind-independent Platonic entities. Indeed, Frege argues that senses belong to a third realm: “The thought belongs neither to my inner world as an idea nor yet to the outer world of material, perceptible things” (“The Thought” 35). Moreover, the very existence of thoughts is entirely independent of human influence: “We grasp thoughts but we do not create them” (Frege, “A brief survey of my Logical Doctrines” 198).¹⁷

That Frege argues that senses are objective should not surprise us. The success of his concept-script depends importantly on establishing the objectivity not only of logical laws, but of conceptual content. Indeed, it is here that he locates the fundamental error of many of his predecessors. “Psychological treatments of logic,” he writes, “arise from the mistaken belief that a thought (a judgment as it is usually called) is something psychological like an idea” (“Logic” 144). According to Frege, objectifying senses is the only way to explain adequately how we manage to have meaningful discussions in the first place. Thoughts must be conceived as objective, mind-independent entities,

Otherwise two people would never attach the same thought to the same

¹⁷ Note that this entails the (perhaps troubling) corollary that before the discovery of quarks and before the invention of rollerblades and cheeseballs there existed thoughts waiting to be expressed by sentences employing the terms ‘quark’, ‘rollerblades’ and ‘cheeseballs’.
sentence... A dispute about the truth of something would be futile. There would be no common ground to fight on. ("Logic" 127)

Frege reasons that to conceive of senses as ideas is to render the meanings of our utterances inaccessible to one another, in which case meaningful discussion is impossible. He adds,

If something were true only for him who holds it to be true, there would be no contradiction between the opinions of different people... There would be no science, no error and no correction of error. ("Logic" 132-33)

Frege argues that if senses were determined by the mental states of particular individuals, then understanding and, so, verifying the truth of your claim that 'Water present at sea level boils at one hundred degrees celsius' would be as difficult as understanding and, so, verifying your claim that 'The Mona Lisa is beautiful'. Rigorous enquiry could not get off the ground.

Frege identifies further advantages to conceiving of thoughts as objective, Platonic entities. First, it explains how knowledge is preserved over time. He writes, "For one can hardly deny that [humankind] has a common store of thoughts which is transmitted from one generation to another" ("On Sense and Reference" 59, insert mine). Second, Frege contends, conceiving of thoughts as objects explains how we understand sentences for which the author is unknown: "If we see the sentence '2+3=5' written on a wall, we have no difficulty at all in recognizing the thought expressed by it, and we do not need to know who has written it there in order to understand it" ("Logic" 134). Third, thoughts conceived as objective Platonic entities gives us a criterion of successful translation, namely, the preservation of the thought expressed ("Logic" 141).
Frege’s Literal Language

How does Frege’s analysis of language bear on our investigation into the distinction between literal and non-literal language? I want to suggest that his account of meaning in terms of the doctrines of sense and reference constitutes the first sustained attempt by a philosopher to demarcate clearly between that in language which is literal and all else. If I am right about this, then his entire analysis of natural language operates on the implicit assumption that it is literal meaning that is his subject matter. To appreciate this, we need to re-evaluate the features of his analysis of natural language salient to our investigation.

First, though, it is instructive to note that Frege nowhere discusses the concept of non-literal language. In fact, in none of his major works on language does he even mention figurative devices such as metaphor, irony, hyperbole, and the like. He comes closest with the following comment about poetry: “The poet does not really depict anything: he only provides the impetus for others to do so, furnishing hints to this end, and leaving it to the hearer to give his words body and shape” (“Logic” 140). It may appear at first that here Frege is alluding to figurative devices such as metaphor, but what he says immediately following this indicates otherwise: “And in this connection it is useful to the poet to have at his disposal a number of different words that can be substituted for one another without altering the thought, but which can act in different ways on the feelings and imagination of the hearer (“Logic” 140). Apparently, Frege

18 Frege also claims that in poetry we leave aside the question of the proper reference of fictional names such as ‘Odysseus’ (“On Sense and Reference” 63). But this is a point about fictional names rather than about poetic devices, so it is of no help to us here.
intends only to observe that synonymous expressions might possess different evocative powers, and that the poet does well to bear this in mind. But this has nothing to do with metaphor and the like. To substitute the metaphor ‘No man is an island’ for the literal ‘No man exists in isolation’ is to alter the thought expressed, from a banal truth about the impossibility of identifying men and islands to a philosophical claim about human existence. In the situations about which Frege is thinking, though substituting one synonym for another may conjure different images and emotions, it nonetheless leaves the thought untouched. Frege has in mind the difference between, say, saying that someone is ‘sad’ and saying that they are ‘morose’.

Given Frege’s preoccupation with the concept-script, his conspicuous inattention to non-literal language should not surprise us. If queried, Frege would likely claim that there is no need, nor is there a place, for non-literal devices in any language suited for rigorous enquiry. Indeed, he likely thought that in virtue of being other than literal, non-literal language has no place in a rigorous language like the concept-script—that is, in a system which seeks to assign one interpretation to each of its constituent sentences. His apathy toward non-literal language is indicated by the examples that he gives to support his analysis of natural language. These are almost exclusively of short, literal sentences that employ uncontroversial referring expressions—‘The Moon is smaller than the Earth’, ‘5 is a prime number’, ‘Kepler did not die in misery’, ‘Ice is less dense than water’, ‘I smell the scent of violets’, ‘All horses are four legged animals’, ‘Whatever is a mammal

---

19 It is critical to Frege’s point that the substitution leave the thought unaltered. To ensure this, one cannot simply substitute co-refering expressions for one another; rather, the expressions must share the same reference and sense—though, and this is the point, the resulting sentence may evoke different psychological effects than the original sentence.
has red blood’, ‘Caesar conquered Gaul’, ‘Oxygen is combustible’, and so on. Granted, there are exceptions, but they arise only when Frege wants to illustrate the logical imperfections of natural language. Moreover, these examples are never of figurative language, but rather of, say, expressions found in fiction that fail to refer.

Frege’s stated interest in natural language lies not with the distinction between literal and non-literal language, nor with the concepts of literal and non-literal language themselves, but rather with that which he deems relevant to logical inference, namely, conceptual content. But the notions of conceptual content and of literal language are closely related. As we have seen, according to Frege, conceptual content consists of the thought expressed by an indicative, and its truth value. At the foundation of this analysis are the theses that thoughts are objective and shareable and that truth values are invariable and independent of a speaker’s beliefs. As Frege is well aware, these theses do not hold up if it is assumed that all features of an indicative, utterance constitute equally good guides to its conceptual content. Put differently, these theses depend on the assumption that only certain features of indicative, utterances serve as clues to the identity of the thoughts they express. These features include the meanings of the component expressions and the logical structure of the sentence. They do not, however, include the tone of voice, the use of words like ‘ah’ and ‘unfortunately’, the difference between the connotations of the conjunctions ‘but’ and ‘and’, and so forth. Demarcating what belongs to the thought expressed by an indicative, is the very role Frege intends the semantic concepts of sense and reference to fill. In other words, to posit a sense and a reference

---

20 See “On Sense and Reference” (62).
for each expression is to posit an objective, independent, invariable, shareable meaning. Indeed, the very idea that thoughts "do not belong to the individual mind... but are independent of our thinking and confront each one of us in the same way" presupposes the availability of standard meanings (Frege, "Logic" 149).

Frege's analysis of natural language depends importantly on being able to distinguish between that which is relevant to conceptual content and that which is not. Put differently, Frege's analysis presupposes the availability of a distinction between the thought(s) expressed by an indicative, and those that are merely suggested by it. Indeed, as Hans Sluga explains, "This distinction is what guarantees the separation of logic from psychology" (Gottlob Frege 76). Critical to Frege's analysis, then, is the claim that conceptual content does not depend in any way on the psychological particulars of any one speaker of the language. Indeed, Frege intends to demonstrate by means of the concepts of sense and reference that a theory of meaning is possible according to which the meaning of a sentence is determined solely by its logical form and component expressions, irrespective of any particular occasion of its utterance.

Frege grants that separating the psychological from the logical in natural language is not easy. Unlike the concept-script, natural language, he observes, is saturated with psychological distractions:

Right from the start we experience the series of sounds themselves, the tone of the voice, the intonation and rhythm with feelings of pleasure or displeasure... This is quite independent of the aim of the words to express thoughts. ("Logic" 139)

Moreover, we must keep distinct the thoughts expressed by a speaker's sentences and those which "the speaker leads others to take as true although he does not express them"
Again, of what relevance is this to our investigation? I submit that when Frege says that the sense and reference of a sentence (or, in the context of a sentence, of a word or phrase) together constitute its meaning, he means its literal meaning. Put differently, Frege's distinction between what a sentence says and what it merely suggests (or what we take the speaker to be suggesting by uttering it) just is the distinction between literal language and all else. A thought, then, is to be understood as the literal interpretation of an indicative.\textsuperscript{21} Though Tennyson's metaphor 'Sleep is the brother of death' might suggest to us a close relation between sleep and death, it says only that sleep is the brother of death (which is absurd). Similarly, to borrow an example from H.P. Grice, though you might take my reply 'It's Sunday' to your question about the location of the nearest gas station to indicate that the all gas stations are currently closed, my utterance means literally that it is Sunday.

But what of non-indicatives? That a sentence expresses a thought is a sufficient condition of its literalness, but is it also a necessary condition? That is, all thoughts are literal, but do all instances of literal language express thoughts? Given what we know of Frege's analysis, the answer must be no. On Frege's account, there are numerous sentences in natural language which, though they are literal, nonetheless do not express

\textsuperscript{21} Whether literalness, like truth, is a property of senses rather than of sentences depends on how one chooses to incorporate an account of non-literal language into a Fregean framework. Consider, for example, the case of a metaphorical utterance. On the one hand, if one were to argue that in addition to the literal thought expressed by the sentence there is also a metaphorical sense, then it would follow that literalness is a property of senses, not of sentences. On the other hand, if one were to argue that each sentence of natural language expresses just one sense (its literal sense), then it would be redundant to call the sense expressed by a natural language sentence 'literal'.
thoughts. These include optatives (‘Would that I were dancing’), interrogatives (‘Would you please pass the soap?’), imperatives (‘Put away that book!’) and exclamations (‘What a beautiful day!’), all of which are capable of being literal. These also include indicatives which contain one or more referring expressions that fail to refer (‘Santa Claus fed Rudolph some cookies’) and, perhaps, value judgments.

Demarcating the Bounds of Sense

How, according to Frege, can we distinguish between that which is part of the sense expressed by a sentence, and that which is not (but which nonetheless may be associated with it)? That is, how, according to Frege, do we demarcate the bounds of the literal meaning of a sentence? That Frege has available to him an answer to this question is of the utmost relevance to the success of his analysis of natural language. Without a clear criterion for identifying sense content, Frege’s application of the notion of conceptual content to the analysis of natural language stalls. In the concept-script, identifying sense

---

22 Frege’s remarks on this are not always consistent. At one point he writes that “an interrogative sentence and an indicative one contain the same thought” (though the indicative “lays it down as true” whereas the interrogative contains, in addition to the thought, “a request”) (“The Thought” 21). Elsewhere he argues that the difference between assertions, interrogatives, optatives, imperatives, and so forth, resides in their different senses. That is, he argues that just as assertions express thoughts, so interrogatives express questions, imperatives express commands, optatives express wishes, and so forth: “A command, a request, are indeed not thoughts, yet they stand on the same level as thoughts” (“On Sense and Reference” 68). In any case, this latter position is most consistent with the rest of his theory.

23 See footnote 11.

24 Although, according to Frege, the sense and reference of a sentence together constitute its meaning, the sense is to be considered primary: “It is the striving for truth that drives us always to advance from the sense to the reference” (“On Sense and Reference 63). Thus, establishing the meaning of a sentence is first and foremost a matter of establishing its sense (cf. Baker and Hacker, Frege 304; Dummett, Frege 159).
content is easy—there the sense of each expression, and so of each sentence, is expressly
defined. In the case of natural language, however, things are not quite so clear.\footnote{Frege’s most explicit remark on this is found not in one of his oft cited works on language, but rather in “Logic.” There, in reference to the fact that in the concept-script issues of sense content are decided in advance by agreement, he adds, “In [natural] language common usage takes the place of such agreements” (“Logic” 141). Unfortunately, this comment constitutes the extent of his attention to the matter.}

This is not to say that Frege is without a response to the question of demarcating
sense. In “On Sense and Reference,” he invites us to consider the example, ‘Napoleon, who recognized the danger to his right flank, himself led his guards against the enemy position’ (75). He says that each clause expresses a thought: the first, that Napoleon recognized the danger to his right flank; and the second, that Napoleon himself led his guards against the enemy position. But, he observes, one might also be tempted to attribute to the sentence the thought that the knowledge of the danger was the reason why Napoleon led the guards against the enemy position. How do we decide whether this thought is in fact expressed by the sentence or merely associated with it? Frege answers:

Let the question be considered whether our sentence be false if Napoleon’s
decision had already been made before he recognized the danger. If our
sentence could be true in spite of this, the subsidiary thought should not be
understood as part of the sense. (“On Sense and Reference” 75)

Frege urges us to agree with him that the original sentence, ‘Napoleon, who recognized
the danger to his right flank, himself led his guards against the enemy position’, would
\textit{not} be false if Napoleon’s decision had already been made before he recognised the
danger. If this is the case, then the thought that the knowledge of the danger was the
reason why Napoleon led the guards against the enemy position cannot be part of what is
expressed by the sentence. Otherwise, the sentence would express a conjunction of three
thoughts, one of which is false. This, in turn, would render the sentence false as a whole.

It appears, then, that Frege provides us with a means of determining sense content. First, we assume that the sentence, or, rather, that the thought expressed by the sentence, stripped of the thought about which we are enquiring, is true. Then we ask: if the thought about which we are enquiring were false, would the original, pared down thought also be false? If the answer is no, then the thought under consideration is not in fact expressed by the sentence. If the answer is yes, then it is. Indeed, this method is really just an extension of Frege’s thesis that the meaning of a sentence is identified with its inferential power. In the above example, he might well have asked, ‘Does the sentence “Napoleon, who recognized the danger to his right flank, himself led his guards against the enemy position” entail that the knowledge of the danger was the reason why Napoleon led the guards against the enemy position?’.

We might very well find ourselves in agreement with Frege about the above example. But consider a second example for which it is less clear that Frege’s solution meets with our intuitions. In “Logic,” Frege asks us to consider the sentences, ‘This dog howled the whole night’ and ‘This cur howled the whole night’ (140). Though the word ‘dog’ is “neutral” and the word ‘cur’ has “unpleasant associations,” Frege argues that the thought expressed by each sentence is the same—that is, the first sentence “tells us neither more nor less than the second” (“Logic” 140). Frege acknowledges that there may be resistance to this characterization of the matter: “It might be thought that the second sentence does nevertheless tell us more than the first, namely that the speaker has
a poor opinion of the dog” (Logic 140). If this were the case, he adds, then the word ‘cur’ would contain the thought that this is an unkept dog. He continues, “We can put this to the test in the following way” (“Logic” 140). As with the Napoleon example, he asks us to suppose whether the second sentence, ‘This cur howled the whole night’, would be false if the dog were not in fact unkept. That is, if it were true that this dog howled the whole night, would the sentence ‘This cur howled the whole night’ be false given that this dog is not in fact unkept? Frege does not think so: “The use of the word ‘cur’ does not prevent us from holding that [‘This cur howled the whole night’] is true as well” (“Logic” 140, insert mine).26

Frege is clearly satisfied by this solution to the problem of identifying sense content. Upon completion of its exposition in “Logic,” he writes, “[W]e have a means of deciding what is and what is not part of the thought” (141). But considered carefully his solution is imperfect both in scope and in method. First, consider its scope. Because his proposal is tied closely to the notion of truth, it applies only to those senses capable of being true or false, namely, thoughts. Thus, its scope is limited to senses expressed by indicatives. But indicatives are not the only sentences of natural language which express literal senses. Indeed, interrogatives are capable of expressing literal questions, imperatives literal commands and optatives literal wishes.27 In addition, sentences that employ non-referring expressions are capable of being literal. According to Frege,

26 This example emphasizes the difficulty of drawing the line between the so-called psychological and the so-called logical. Presumably it will be evident to some speakers of the language and questionable to others that part of the meaning of the word ‘cur’ is ‘unkept’.

27 See “On Sense and Reference” (68). There Frege says that commands, requests, and so forth “stand on the same level as thoughts.”
however, all such sentences are incapable of being true or false. They cannot, then, be subjected to the above test.

One might respond that because Frege's interest in natural language lies almost exclusively with indicatives, it is irrelevant that his proposal does not apply to other kinds of sentences. But, in so far as he attributes the concepts of sense and reference to all meaningful sentences of natural language, including imperatives, interrogatives, optatives, and the like, we should ask how sense content is identified. Indeed, we expect that the meanings of 'The Blue Jays won the World Series', 'Would that the Blue Jays were to win the World Series' and 'Did the Blue Jays win the World Series?' are closely related—that is, we expect that there is a content common to all of them and that all matters pertaining to this common content, including the determinants of its bounds, will bear equally on each sentence.

Perhaps Frege's proposal can be modified to broaden its scope to include sentences other than indicatives. One might argue that prior to applying his test, we must extract from the non-indicative its corresponding indicative. In our example above, we extract from the imperative 'Stop this cur from howling!', the indicative 'This howling cur is to be stopped' (or something like this) and then ask the question whether it entails that this dog is unkept. This response might seem promising, but it is not without its own problems. In particular, there are sentences for which it is not clear that there is an appropriate corresponding indicative. These will include interrogatives which do not solicit the response 'yes' or 'no'. For example, it is not clear what indicative might be extracted from the question, 'What is your favourite colour?' or from 'Who was the first
Prime Minister of Canada?’. Moreover, it is not clear that there are indicatives, to be extracted from indicatives which contain one or more referring expressions that fail to refer. That is, presumably one cannot extract indicatives, from the sentence ‘Humpty Dumpty fell off the wall’.

The question of scope aside, there is a further, more fundamental problem with Frege’s proposal, one that reaches to the core of his analysis of natural language. In accord with his anti-psychologism, Frege must avoid reducing the question of identifying sense content to the question of what seems to us to be the case. In the above case, for example, the question he does not want to ask is, ‘If it is true that this dog howled the whole night, and that this dog is not unkept, then does the sentence “This cur howled the whole night” seem to us to be false?’ Rather, he wants to ask, ‘Would it be false?’ Nevertheless, he leaves the reader with a sense that subjective judgment may be the only means of assessment available. In his discussion of the Napoleon example, he says, “If our sentence could be true in spite of [the falsity of the subsidiary thought], the subsidiary thought should not be understood as part of the sense,” and continues, “one would probably decide in favour of this” (“On Sense and Reference” 75, insert and emphasis mine). But we want to ask Frege, how does one decide, if not by what seems correct? Indeed, Frege’s proposal for identifying sense content risks circularity—one must already know what the utterance means in order to assess whether it entails the subsidiary thought.

Frege’s concepts of sense and reference are intended to serve as markers for all that is relevant to the objective, shared meaning of a sentence—that is, all that is relevant
to the literal meaning of a sentence. He argues persistently for the importance of recognising and maintaining the distinction between sense and reference—i.e., literal meaning—and all else; that is, between what a sentence means (literally) and what it suggests in a particular context. But he says too little about how sense content is identified (and, so, about how reference is determined). Without a theory of how words and phrases and, so, sentences acquire their sense and, so, their reference, we are left without a complete account of how the meanings of sentences are determined. Put simply, though Frege advances the claim that we can, indeed must, demarcate the literal meaning of a sentence, he says too little about how this is to be done. As we will find, the influence of his approach to linguistic analysis renders the issue of the bounds of literal meaning and, so, of the distinction between literal and non-literal language important to certain projects in analytic philosophy. If we are unable to answer adequately the question left to us by Frege—How are the boundaries of literal meaning determined?—then we are left with a picture of language that is, at best, incomplete.
Chapter 3
The Orthodox Position

Frege’s logicist programme generates a certain vision of natural language in general, and of literal meaning in particular. His concept-script provides analytic philosophers with a particular conception of the underlying structure of language, and even of thought and of reasoning. His programme implies that a language is, beneath all its appearances, a calculus of hidden rules and that speaking and understanding a language is a matter of operating according to these rules.

As we discovered in the last chapter, Frege strips meaning (and inferential rules) of any psychological content. As Dummett writes, Frege repudiates “any supposedly essential connection between the sense of the word and the psychological process which may accompany or precede its acquisition, and which may, within the realm of psychological law, be a necessary condition of that acquisition” (Frege 678). Language, Frege argues, does not encode ideas, conceived as psychological entities, but thoughts and concepts conceived as mind-independent Platonic entities. Meanings—i.e., senses—exist independent of our grasping of them. Indeed, it is Frege’s firm belief that a systematic theory of meaning is possible only if we first expel psychology from logic and, so, from the philosophy of language.

Frege’s Influence

Many of Frege’s theses about language are taken up in one form on another by subsequent philosophers in the analytic tradition. Indeed, the picture of language that
emerges from Frege’s work dominates subsequent philosophical enquiry into linguistic meaning. Whether it is Bertrand Russell’s theory of descriptions, Ludwig Wittgenstein’s theory of logical necessity, Rudolf Carnap’s work on the structure of languages, or the verificationism of the Logical Positivist’s, certain suppositions about linguistic meaning (implicit and explicit), suppositions borrowed from Frege, are virtually ubiquitous.

For example, Russell’s theory of descriptions, once touted as a paradigm of logical analysis, owes much to the Fregean analysis of sentence meaning (Russell, “On Denoting”). Briefly, Russell is troubled by the apparently forced conclusion that a sentence is meaningless if it contains a referring expression that fails to refer—for example, ‘The present King of France is bald’. Russell’s triumph is to show that such sentences are not, despite their appearance, of the form ‘The F is G’. Rather, their true logical form is ‘There exists exactly one thing which is F and that thing is G’. Hence, he argues, to say ‘The present King of France is bald’ is not to say of the present King of France that he is bald, but rather that there exists exactly one present King of France and he is bald. Thus understood, the sentence is not meaningless; rather, because the first part is false (i.e., because it is not the case that there exists exactly one present King of France), the sentence as a whole is false.

Frege’s influence on Wittgenstein also is apparent. In the Preface to his Tractatus Logico-Philosophicus Wittgenstein declares explicitly his debt to “The magnificent works of Frege and the writings of my friend Bertrand Russell.” The influence of Frege’s (and Russell’s) work on the views of the Tractatus is, as Sluga declares, “pervasive” (Sluga, Frege 181). In particular, Wittgenstein adopts whole-heartedly Frege’s anti-psychologism
about meaning. Indeed, the *Tractatus* is perhaps the most powerful example of the shift in philosophers' conception of their subject matter to language. It is, as Dummett describes, "a pure essay in the theory of meaning, from which every trace of epistemological or psychological consideration has been purged" (Dummett, *Frege* 679). Influenced by the Fregean approach to analysis, Wittgenstein argues that the function of philosophy is to regulate the bounds of sense, to elucidate philosophically problematic sentences, and to show that attempts to cross the bounds of sense are pointless.

Carnap in turn credits Frege and Wittgenstein with providing the logical framework within which he and other Logical Positivists construct their accounts of verificationism. Roughly, the aim of Carnap and the Logical Positivists is to develop a systematic theory of literal meaning according to which the meaning of a sentence is specified by the actual steps we take to verify its truth or falsity. According to this strict approach, if these steps cannot be provided—as in the case, say, of a metaphysical claim—the sentence is meaningless.

Frege provides not only the vision with which these and other analytic philosophers construct their accounts of language, but also the tools. His function/argument analysis provides the basic form for the analysis of sentences. In particular it provides key to the structure of sentences involving generality. Moreover, his distinction between *sense* and *force* is generally taken to be a requirement of any systematic theory of meaning. Further, his concepts of *sense* and *reference* provide the

---

1 The objective of Carnap's *The Logical Syntax of Language*, for example, is to give a formal account of the structure of *any* possible language. In his "Intellectual Autobiography" Carnap credits his interest in logical syntax and semantics to Frege (12-13).
framework for most analytic theorising about language.

*Compositionality and Truth Conditions*

Of concern to our investigation are two particular suppositions with which Frege operates, two suppositions that pervade subsequent work in analytic philosophy. The first of these is the joint supposition that the sentence, not the word, is the basic unit of meaning, and that sentence meaning is compositional. As we discovered, according to Frege, a complete theory of meaning for a language explains how the meanings of sentences are derived from the meanings of their constituent parts and logical structure. Stemming from this is the supposition that it is the sentence type, not token, that is the primary bearer of meaning. According to this view, sentence meaning is determined entirely independent of context—that is, the interpretation of a particular utterance of a sentence is independent of knowledge of its extra-linguistic context. This supposition is treated as a basic assumption in many subsequent analytic projects. Indeed, the commonly held notion that the literal meaning of a sentence is the meaning it has independent of any particular use to which it is put stems directly from Frege’s principle of compositionality.

The second of Frege’s suppositions that pervades analytic philosophy and is of

---

2 Though Frege offers no adequate definition of a sentence. Perhaps he comes closest with this suggestion: “What does one call a sentence? A series of sounds; but only when it has a sense, by which it is not meant that every series of sounds that has a sense is a sentence” (“The Thought” 19).

3 In chapter four we will discuss certain apparent barriers to this supposition, including the existence of ambiguities and of indexicals.
interest to us is his link between meaning and truth conditions. Frege's truth-conditional semantics is what makes possible the extension of his logical framework to natural language. An indicative,\(^4\) purports to describes the world as being a certain way. Its meaning, then, is given by stating the conditions the world must meet in order for things to be as the sentence says they are. Thus by linking meaning to truth, Frege attributes a special status to \textit{literal} language. This supposition pervades, virtually un-remarked, throughout various works in the analytic tradition. That is, though on the face of it the assumption that proper philosophical enquiry necessarily involves the analysis of language says nothing about the distinction between literal and non-literal language, hidden beneath this assumption is the further assumption that it is \textit{literal language} that is the object of analysis. Although non-literal devices, most notably metaphor, recently have received widespread respect as important philosophical devices, no analytic philosopher awards them status similar to literal language when it comes to their role in the representation of states of affairs. Indeed, none of the aforementioned analytic philosophers claims that there are important philosophical truths to be uncovered in the analysis of non-literal language.

Put simply, \textit{literal} meaning is what natural language semantics is about. That is, the distinction between literal and non-literal language is at least implicit in \textit{any} analysis of meaning whatsoever. As a result, natural language semantics assigns to the concept of literal meaning a kind of philosophical importance not found in pre-twentieth century

\footnote{\(^4\) Recall that an indicative,\(r\) is an indicative sentence that contains no referring expression that fails to refer.}
philosophising about language. Indeed, Frege, Russell, the Positivists and others raise the stakes in the effort to draw distinction between literal and non-literal language: making a principled distinction is critical to the success of natural language semantics.

The Orthodox Position

The picture of meaning that emerges within (and in part defines) the strand of analytic philosophy with which we are interested is not a theory per se, but rather a framework within which various analytic philosophers formulate their competing semantic and pragmatic theories of natural language. In its most general form, the picture is this: a language is a set of precise and specifiable syntactic and semantic rules; verbal communication depends on speaker and interpreter sharing the ability to operate according to the rules of the language being spoken, and it requires no more than this. Within this broad framework, these analytic philosophers share a series of more specific suppositions about language, suppositions that are relevant to our investigation:

1. There is a meaning that is associated with each sentence taken independent of context (i.e., independent of who utters it and of their purpose in uttering it). This is called sentence meaning. A sentence may have more than one meaning, in which case it is ambiguous, or its meaning may be defective in such a way that it is nonsensical.

2. Sentence meaning is determined entirely by the meanings of the component expressions and logical structure of the sentence.

3. Literal meaning and sentence meaning are equivalent.

4. The literal meaning of a sentence needs to be distinguished sharply from the uses to which the sentence might be put (i.e., speaker meaning may depart from literal sentence meaning in a number of ways, as with figurative meaning).
5. Of literal meaning, metaphorical meaning, speaker's meaning, etc., only literal meaning is a property of sentences.

6. For sentences in the indicative, literal meaning determines a set of truth conditions such that a literal utterance of the sentence will be true if and only if those conditions are satisfied. (Some philosophers in this tradition equate the literal meaning of a sentence with its truth conditions.)

Together, these suppositions form the framework within which the semantic analysis of language is conducted. Together they form what I will call the orthodox position.\(^5\)

---

\(^5\) The term *orthodox position* is borrowed from Robert Wilensky (*Primal Content and Actual Content* 163).
Chapter Four
On The Notion of Context Independent Meaning

According to what I have termed the *orthodox* approach to the analysis of natural language, associated with each sentence of a language is a meaning which exists independently of any particular use to which that sentence might be put. This meaning is compositional—that is, it is determined entirely by the logical structure of the sentence and the meanings of the component expressions of the sentence. According to the *orthodox position*, meaning, or rather *literal* meaning, is determined entirely independent of context. Granted, the sentence ‘It is cold in here’ may be used to motivate someone to close an open window, or to comment metaphorically on the mood in a room, but it means (literally) that it is cold in here. Any implicatures, figurative interpretations, or the like that we wish to assign to a sentence on a particular occasion are, according to the *orthodox position*, distinct from its literal meaning.

To make sense of the *orthodox position*’s notion of literal meaning, then, we must be able to make sense of the notion of context-independent meaning. To be sure, this would not be without benefit to our investigation. Indeed, context-independent meaning is, *prima facie*, an ideal criterion for identifying literal meaning. In the aforementioned case, for example, deprived of any particular context—e.g., a room with an open window or a room in which the mood is tense—we would agree that the sentence ‘It is cold in here’ means simply that it is cold in here; we would not be tempted to assign to it the meaning either that I want you to close the window or that the mood in the room is tense.
What Does 'Context-Independent' Mean?

The term 'context-independent meaning' is somewhat misleading.⁠¹ Strictly speaking, there can be no such thing as meaning entirely independent of context. Devoid of any context whatever, the noises you make and the marks on this page will not even be recognised as attempts to communicate. Communication, and so meaning, is possible only against some kind of backdrop of shared beliefs about the world. But how might we characterise the nature and scope of this set of beliefs?

Jerrold Katz, an advocate of the orthodoxy position, has done as much as anyone to try to make sense of the notion of context-independent meaning. He characterizes the literal meaning of a sentence as that which would be understood by an ideal speaker of a language in the “zero” or “null” context, which is defined in turn as “a context whose features provide no relevant information for choosing a compositional meaning as the utterance meaning different from the compositional meaning of the sentence used” (“Literal Meaning and Logical Theory” 217n.). Left as it stands, however, this characterization of literal meaning is circular—Katz defines literal meaning as meaning in the null context, and he then defines the null context as one in which we would choose the literal (compositional) meaning as utterance meaning. But Katz elucidates the notion of a “null” context with what he terms “the anonymous letter situation,” described as “the case where an ideal speaker of a language receives an anonymous letter containing just one sentence of that language, with no clue whatever about the motive, circumstances of transmission, or any other factor relevant to understanding the sentence on the basis of

---

¹ Though perhaps not as misleading as the often used ‘context-free meaning’.
its context of utterance” *(Propositional Structure and Illocutionary Force 14).² Literal meaning is constituted by “those aspects of the meaning of a sentence that an ideal speaker-hearer of the language would know in such an anonymous letter situation” *(Propositional Structure and Illocutionary Force 14).³

*Prima facie,* Katz’s anonymous letter scenario does a laudable job of capturing our pre-theoretic understanding of what would constitute a sensible interpretation of the notion of context-independent meaning. Granted, it is perhaps less than ideal in that it retains the specific context that it is an anonymous letter. In addition to the wonder, confusion, excitement, or the like that this might elicit, Jonathan Berg notes that in the context of the anonymous letter scenario if the sentence were to read, say, ‘This letter was written anonymously’, we would likely not notice that the word ‘letter’ has more than one meaning (“Literal Meaning and Context” 408).⁴ Nonetheless, this is not itself reason enough to abandon the anonymous letter scenario. Indeed, Berg’s point serves to highlight that the anonymous letter scenario is simply a heuristic device, albeit a powerful


³ To be clear, what Katz has in mind is not the case in which one receives a letter which, though it is written anonymously, is nevertheless intended for a particular recipient (or a group to which the recipient belongs), as in the case, say, of an anonymous letter sent by a secret admirer. Rather, he has in mind the more fantastic, though nevertheless plausible, scenario in which the only salient characteristic of the recipient of the letter is that she speaks the language to which the sentence in the letter belongs.

⁴ Berg writes that in this case, It would be interpreted as concerning a certain piece of correspondence (the one at hand). But the sentence has another literal meaning: a letter is also a letter of the alphabet. Since the contextual information in the anonymous letter situation leads the interpreter to overlook or disregard this other literal meaning of the sentence, the situation is not really null. (“Literal Meaning and Context” 408)
and important one. Its appeal resides in the fact that it gives a down-to-earth spin to the highly theoretical notion of context-independent meaning. The importance of this role cannot be overstated easily. The anonymous letter scenario affords us an understanding of what context-independence could be, and so something against which to measure various of the orthodox position’s theoretical claims about literal meaning.⁵

*Traditionally Cited Barriers to Context-Independent Meaning*

Identifying the context-independent meaning of a sentence in a logical language like that of Frege’s *Begriffsschrift* or Russell and Whitehead’s *Principia* is straightforward; indeed, these languages were designed with this end in mind. Each of their constituent sentences has one and only one meaning with a determinate set of truth conditions which determines for any context whether the sentence is true or false. For many natural language sentences, however, there are apparent barriers to the determination of context-independent meaning. Perhaps the most commonly cited of these apparent barriers are ambiguity and indexicality. The pervasiveness in natural language of ambiguities and of indexicals appears to threaten the idea that each sentence has a context-independent meaning which determines for any context whether the sentence is true or false.

Ambiguous sentences—that is, sentences that invite two or more independent

---

⁵ Indeed, more broadly, one might even contend that a measure of the success of an account of literal meaning (though certainly not the only one) is that it explain how interpretation is possible in such a situation.
interpretations—are a commonplace in natural language.\(^6\) \textit{Prima facie}, the presence of ambiguities in natural language represents a barrier to the \textit{orthodox position}'s identification of literal meaning with context-independent meaning. For example, were we to receive an anonymous letter which read ‘Bill is a poor philosopher’, we might understand it to mean either that Bill is not very adept at practising philosophy or, alternatively, that Bill is a philosopher with very little money. But in fact ambiguities represent a mere inconvenience to the \textit{orthodox position}: an ambiguous sentence is simply one that has two or more context-independent meanings.

Indexicals seem to pose a greater threat to the viability of the notion of context-independent meaning than do ambiguities.\(^7\) Indexicals are a species of referring expressions, which include pronouns (e.g., ‘I’, ‘she’, ‘you’, ‘it’, ‘my’), demonstratives (e.g., ‘this’, ‘that’), certain time or place indicators (e.g., ‘here’, ‘now’, ‘over there’, ‘tomorrow’) and the words and aspects of words that indicate temporal position (e.g., ‘It is raining’, ‘It rained’). Of interest to our investigation is that the defining characteristic of an indexical is that it depends for its reference on a particular context of utterance. Given the \textit{orthodox position}'s assumption that meaning is assigned to sentence-types, not

\footnotesize{
\begin{itemize}
\item \(6\) Stipulating that the interpretations are independent ensures that figurative devices whose interpretations are parasitic on literal meaning are not mistaken for ambiguities. Also, note that according to the orthodox picture a sentence may or may not retain its ambiguity in context.
\item \(7\) Though, according to Frege, indexicals are to be understood as a particular species of ambiguity. Since ‘I’ has a different \textit{reference} in your utterance of ‘I am hungry’ than in my utterance of ‘I am hungry’, your utterance of ‘I’ has a different \textit{sense} than my utterance of ‘I’. This analysis solves one problem, namely, that of explaining how your use of the word ‘I’ differs from my use of it. But it fails to explain the similarities between our two uses of ‘I’. Indeed, Frege’s position forces the counter-intuitive conclusion that ‘I’ has an altogether different meaning when you use it and when I use it. Many consider this an unacceptable corollary of Frege’s particular account of meaning. For a fuller discussion, see J. Perry, “Frege on Demonstratives.”
\end{itemize}
}
tokens, sentences that employ indexicals appear to threaten the claim that sentences have context-independent meaning or, at the very least, that context-independent meaning alone generates truth conditions. The presence of indexicals implies that the relation between meaning and truth cannot be straightforward because indexicals depend in part for their meaning—specifically, for their reference—on the context of the utterance. A sentence which employs one or more indexicals may be true in one context and false in another—for example, the truth of the sentence ‘You are here now’ depends on who says it, where they say it, when they say it and to whom they say it. Nor is this issue isolated to a few anomalous cases. Very many of our natural language sentences employ one or more indexicals. Indeed, by including verb tense among the indexical features of natural language, we ensure that the above considerations apply to virtually every natural language sentence.

What, then, of the orthodox position’s quest for context-independent meanings? Because according to the orthodox position the relation of meaning to truth involves truth-conditions as an intermediary, there is room to manoeuvre. One solution is to relativize the truth of a sentence to its particular context of utterance. According to this approach, one might argue that sentences that employ indexicals retain a context-independent meaning which determines a set of truth conditions, but whether those truth conditions are satisfied is determined only relative to a context of utterance—for example, an utterance of the sentence ‘I am hungry’ means that I am hungry and is true if and only

---

8 Donald Davidson, for example, describes truth as “a relation between a sentence, a person, and a time” (“Truth and Meaning” 319). Davidson’s account of linguistic meaning will be discussed in detail in chapter six.
if the speaker is hungry. Accordingly, meaning is a property of sentence-types but truth is a property of sentence-tokens.

Indeed, that the *orthodox position* has available a way to treat indexicals should neither surprise nor impress us. Though we might not know who wrote the anonymous letter containing, for example, the sentence ‘I arrived from New York yesterday’, we would not say that because of this we cannot understand what it means, nor the conditions under which an utterance of it would be true or false.

_Searle’s Challenge to Context-Independent Meaning_

In a series of papers, John Searle challenges the view that the literal meaning of any sentence can be construed as the meaning it has independent of any context whatever.⁹ He argues that “for a large class of sentences there is no such thing as the zero or null context” (“Literal Meaning” 207). He contends, rather, that in general the notion of the literal meaning of a sentence “only has application (it only, for example, determines a set of truth conditions) against a background of assumptions” (“The Background of Meaning” 221). Moreover, Searle adds that unlike the case of indexicals, these background assumptions are not all and could not all be realized in the semantic structure of the sentence (“Literal Meaning” 210; “Background of Meaning” 221).

Searle’s strategy is to consider sentences that appear to have unproblematic context-independent meanings, sentences like ‘The cat is on the mat’, and then

---

⁹ Searle first presents this argument in “Literal Meaning,” and reiterates it in “The Background of Meaning” and in “Metaphor.”
demonstrate that even in such cases we can make sense of the notion of the sentence's
literal meaning only relative to a set of contextual assumptions. He argues, contra the
orthodox position, that it is not possible to assign to the sentence 'The cat is on the mat' a
literal meaning independent of context; rather, its literal meaning will vary depending (in
part) on background assumptions specific to each context of its utterance. Searle supports
his claim by citing cases in which the purported literal meaning of a sentence does not
appear to determine its truth or falsity, even once all ambiguities and indexical features
are taken into account. Searle asks us to envision various abnormal contexts in which the
truth of an utterance of the sentence 'The cat is on the mat' is uncertain. In the first
case, he directs our attention to Figure 1 below and invites us to suppose that the cat and
mat are as depicted but floating freely in outer space where the vertical orientation of the
earth's gravitational field no longer applies.

\[\text{Figure 1}\]

\[\text{To be sure, 'The cat is on the mat' contains indexical elements. To understand an utterance of it we would need to know to which cat and to which mat the speaker is referring and at what time and place the cat is said to be on the mat. But, Searle allows, these features "are already realised in the semantic elements of the sentence" ("Literal Meaning" 210).}\]

\[11\text{ This represents a clever move by Searle. By referring our attention to his drawing (rather than writing, say, "Suppose that the cat is on the mat but floating freely in outer space...", or even giving a more general description of the physical relation between the cat and the mat), he avoids making any presuppositions about whether the cat is (or is not) on the mat, or about what might constitute sufficient conditions for saying that the cat is (or is not) on the mat. The question 'Is the cat on the mat?' (in this context) remains wide open.}\]
Is the cat still on the mat? Or is the earth’s gravitational field a necessary background assumption for determining whether it is true that the cat is on the mat? Searle concludes that in this context, because we do not know what to say about the truth or falsity of the sentence, the notion of the literal meaning of the sentence “does not have a clear application” (“Literal Meaning” 211). But what if we were to add to the truth conditions of the sentence that to be true the cat and mat must be at or near the earth’s surface? Searle argues that for at least two reasons this will not work. First, we can imagine contexts in which we are not at or near the earth’s surface and yet we would want to say that an utterance of ‘The cat is on the mat’ is true (“Literal Meaning” 212). Second, and more important, even if we were to add successfully to the truth conditions of the sentence all the necessary assumptions about gravitational fields, there are an indefinite number of other contextual assumptions that would have to be dealt with in the same way. To illustrate, Searle invites us to consider the case in which that cat and mat are on Earth, but the cat is suspended by invisible wires so that the cat “though slightly in contact with the mat, exerts no pressure on it” (“Literal Meaning” 212). ‘Is the cat still on the mat?’, he asks. Again, he argues that the question does not have a clear answer, and that “this is just another way of saying that the meaning of the sentence ‘The cat is on the mat’ does not have a clear application in the context as so far specified and hence it does not yet determine a clear set of truth conditions” (“Literal Meaning” 212).

Next, Searle asks us to consider a situation in which different speakers are

---

12 Note that this is not an issue of vagueness. If the cat were (on earth) half on and half off the mat we might not be able to agree whether it is on the mat, but this does not jeopardize the notion of context independent meaning in the way that Searle believes his space example does.
equipped with different sets of background assumptions. He asks us to suppose that the cat is drugged into a stupor and is drooped over the edge of a mat that has been stiffened and placed so that one end is raised and supported at an angle. The cat’s owner, unaware of the odd circumstances, asks, ‘Where is the cat?’ to which Searle replies ‘The cat is on the mat’. Searle asks, “Have I told the truth? My inclination is to say that my answer is misleading at best and probably should be described as an ingenious lie, since I know that that is not what the owner understands when he hears and gives a literal interpretation to the utterance of the sentence, ‘The cat is on the mat’” (“Literal Meaning” 213). But then Searle asks us to consider a variation of the same example whereby the mat is part of a row of objects sticking up similarly at odd angles. Moreover, these facts are known both to the speaker and to the audience. The cat, which was at first on one of the other objects, is moved to the mat. Searle concludes, “it is pretty obvious what the correct answer to the question ‘Where is the cat?’ should be... ‘The cat is on the mat’” (“Literal Meaning” 214).

Elsewhere Searle asks us to consider the different contributions of the word ‘cut’ in the sentences ‘Bill cut the grass’ and ‘Sally cut the cake’ (“The Background of Meaning” 223). Searle argues that the semantic content of the word ‘cut’ contributes differently to the truth conditions of each sentence because “the sort of thing that constitutes cutting the grass is quite different from, e.g., the sort of thing that constitutes cutting a cake” (“The Background of Meaning” 223). Searle argues that if someone asks us to cut the grass and we run out and stab it with a knife, or if we are asked to cut the cake and we run over it with a lawn mower, in each case we will have failed to obey the
order in that “that is not what the speaker meant by his literal and serious utterance of the sentence” ("The Background of Meaning" 223). Searle contends that his examples call into question the orthodox position’s thesis that every (unambiguous) sentence has a literal meaning which is absolutely context independent and which determines for every context whether or not an utterance of that sentence in that context is literally true or false ("Literal Meaning" 214). Further, his examples are intended to support the alternative thesis that “for a large class of unambiguous sentences such as ‘The cat is on the mat’, the notion of the literal meaning of the sentence only has application relative to a set of background assumptions” ("Literal Meaning" 214). Moreover, Searle concludes, the truth conditions of many sentences vary given changes in these background assumptions, as shown by the example of the drugged cat that is drooped over the edge of the raised, stiffened mat. Further, he argues, these variations in background assumptions “have nothing to do with indexicality, change of meaning, ambiguity, conversational implication, vagueness or presupposition as these notions are standardly discussed in the philosophical and linguistic literature” ("Literal Meaning" 214). Searle does not deny that sentences have literal meanings.

---

13 Searle intends his argument to extend beyond indicatives and their truth conditions to imperatives and their obedience conditions and to optatives and their fulfilment conditions. He tells of going to a restaurant and saying ‘Give me a hamburger, medium rare, with ketchup and mustard, but easy on the relish’ ("Literal Meaning" 215):

Suppose for example that the hamburger is brought to me encased in a cubic yard of solid lucite plastic so rigid that it takes a jack hammer to bust it open, or suppose the hamburger is a mile wide and is ‘delivered’ to me by smashing down the wall of the restaurant and sliding the edge of it in. Has my order ‘Give me a hamburger, medium rare, with ketchup and mustard, but easy on the relish’ been fulfilled or obeyed in these cases? My inclination is to say no, it has not been fulfilled or obeyed because that is not what I meant in my literal utterance of the sentence. (216)
Rather, he argues for the relativity of meaning: "the thesis of relativity of meaning is the thesis that one can only make these connections [i.e., connecting meaning to truth conditions, entailment, inconsistency, etc.] relative to some coordinate system of background assumptions" ("Literal Meaning" 220, insert mine). Put simply, Searle rejects the tenability of the notion of context-independent meaning as a criterion of literal meaning.

**Searle’s Background Assumptions**

It is important to be clear as to the relation between the shared beliefs referred to in our original presentation of the anonymous letter scenario—that is, the general set of beliefs held by someone in a situation in which they receive an anonymous letter—and Searle’s ‘background assumptions’. This is important because ultimately they are two very different notions, although Searle sometimes conflates them.

In “The Background of Meaning” Searle explains that the reason why the same semantic content, for example, of ‘cut’, determines different sets of truth conditions in, say, ‘Bill cut the grass’ and ‘Sally cut the cake’, is because

as members of our culture we bring to bear on the literal utterance and understanding of a sentence a whole background of information about how nature works and how our culture works. A background of practices, institutions, facts of nature, regularities, and ways of doing things are assumed by speakers and hearers when one of these sentences is uttered or understood. (226-227)

Similarly, in “Literal Meaning” Searle explains that when he says, ‘Give me a hamburger, medium rare, with ketchup and mustard, but easy on the relish’, “entire institutions of restaurants and money and of exchanging prepared foods for money” are in
place (215). Surely Searle is right about these cases. But in neither of his explanations does Searle invoke anything more than the shared set of beliefs available to an individual who finds herself receiving an anonymous letter. Searle's notion of background assumptions, thus far characterised, does not represent a challenge to the orthodox position.

Elsewhere, however, Searle makes the stronger claim that what he means by background assumptions are beliefs and expectations that may vary from one context of utterance to another: "Given one set of [background assumptions], a sentence or expression may determine one set of truth conditions and given another set of assumptions and practices the same sentence or expression with the same meaning can determine a different set of truth conditions" ("Background of Meaning" 227, insert mine). Thus, for example, according to Searle, it is in part because both speakers are aware of the context of the series of objects standing similarly at odd angles, one of which, the mat, has the neighbour's cat drugged and drooped over it, that the response 'The cat is on the mat' to the question 'Where is the cat?' is true. Thus, whereas the set of shared beliefs available to someone in the anonymous letter scenario may include non-contextually specific beliefs such as those about the fact that objects tend to fall toward the earth's surface and that it is often the case that people with lawns mow them, Searle's background assumptions incorporate beliefs specific to the context of the utterance. Indeed, Searle argues that these background assumptions are not specifiable as part of the semantic content of the sentence just because they are "not fixed and definite in number and content" ("Literal Meaning" 214-215).
Responses to Searle

Searle’s argument is a direct challenge to the orthodox position’s supposition that literal meaning, understood as context-independent meaning, is an unproblematic notion. It is no surprise, then, that proponents of the orthodox position take issue with his conclusions. Their responses follow one of two strategies. On the one hand, Berg argues that despite Searle’s claim to the contrary, the sorts of issues his examples raise are in fact ones of indexicality, ambiguity, and the like, in which case they do not threaten the orthodox position’s notion of context-independent meaning. On the other hand, Katz argues that although we might grant that the kinds of background assumptions Searle discusses cannot be realised semantically, this nonetheless does not threaten the orthodox position’s notion of context-independent meaning—that is, he argues that Searle is mistaken about the nature and role of context-independent meaning. I will consider each of these responses in turn.

Berg’s Response to Searle

Berg argues that although Searle repeatedly insists that his examples have nothing to do with indexicality, ambiguity or the like, this is “far less evident than Searle supposes” (“Literal Meaning and Context” 398). He agrees that the truth of ‘The cat is on the mat’ may vary with respect to vertical orientation—that is, that “whether the specified cat is on the specified mat depends (or at least could depend) on which way is up” (“Literal Meaning and Context” 398). But Berg asks, why should this not be regarded as an example of indexicality? He writes,
While explicitly recognising the sentence's indexicality with respect to elements such as time, Searle neglects the indexicality of the preposition 'on' (in the relevant sense) with respect to upwardness. In both cases the truth conditions vary systematically with a particular aspect of the context of utterance, in a way that is understood by any speaker of the language prior to his interpreting any particular utterance of the sentence.¹⁴ ("Literal Meaning and Context" 399)

According to Berg, just as the spatial-temporal location of a speaker is semantically relevant to the meaning of a sentence that employs indexicals such as 'here' and 'now', so too is upwardness to a sentence that employs the preposition 'on' (in the relevant sense). Berg agrees with Searle that we might not know what to say about the truth of an utterance of 'The cat is on the mat' in the context of a cat on a mat in outer space. Contra Searle, however, Berg claims that the question of vertical orientation is realised semantically, not as part of some set of non-semantic background assumptions. That is, just as the truth conditions of the sentence 'Tomorrow it will rain in Toronto' depend in part on the time of utterance, the truth conditions of 'The cat is on the mat' depend in part on vertical orientation. In both cases, Berg would argue, the context dependence in question is of no threat to the notion of context-independent meaning because the expressions 'tomorrow' and 'on' are indexicals.

Further, Berg contends that just as Searle's examples of utterances of 'The cat is on the mat' in outer space "establish no more than the sentence's indexicality," his

¹⁴ Berg's claim that these indexicalities are "understood by any speaker of the language prior to his interpreting any particular utterance of the sentence" is unfortunate and unnecessary (emphasis mine). Indeed, unlike the expressions 'today', 'this', 'now', 'she' and the like, it is arguable that not every competent speaker would identify 'on' (in the relevant sense) as similarly indexical. Later in his paper, Berg acknowledges this: "This indexicality goes easily unnoticed, because it does not come from one of the standard indices, such as time or place" ("Literal Meaning and Context" 403).
examples of utterances of the 'The cat is on the mat' with regard to a cat and a mat
suspended by wires "illustrate nothing more than the sentence's lexical ambiguity"

("Literal Meaning and Context" 399):

The preposition 'on' in one sense indicates only a certain spatial relation, relative to upwardness, while in another sense it indicates a relation concerning pressure exerted as a result of gravitational force ("resting on" as opposed to "on" simpliciter in some broader sense. ("Literal Meaning and Context" 399)

In virtue of these different senses of 'on', the sentence 'The cat is on the mat' is actually ambiguous. As is the case with any ambiguous sentence, we have no reason for choosing one meaning over another when considering the sentence independent of context.

Further, as argued above, ambiguity does not pose a genuine threat to the tenability of the notion of context-independent meaning—an ambiguous sentence simply has two or more context independent meanings. If Berg is right, and Searle's examples can be accounted for as examples of indexicality and/or ambiguity, then Searle's case against context-independent meaning stalls. It is important, then, to be clear as to the force of Berg's argument.

To begin, Berg's suggestion threatens the plausibility of the anonymous letter scenario. This does not seem to bother him, but it will bother advocates of the orthodox position who wish to retain the anonymous letter scenario as a heuristic device for marketing the notion of content-independent meaning. Berg would have us believe that were we to receive an anonymous letter containing just the sentence 'The cat is on the mat', in addition to realising that its truth value depends on the identity of the cat and of the mat and on the time of the utterance, we would also realise that it is ambiguous with
respect to 'on', and that, depending on which of two or more meanings of 'on' is intended, it may depend for its truth on spatial orientation. But that a typical speaker would recognise this is unclear. Indeed, recall that Searle chooses this particular sentence precisely because its literal interpretation is seemingly straightforward.

To be sure, Berg is right to point out that there are many ways in which we make use of words like 'on'. But this alone does not threaten the force of Searle's argument. That is, no matter how precisely we specify various meanings of 'on', there will be additional interpretations not covered by our lexicon. For example, although we might agree that in both 'He has on his favourite cologne' and 'He has on his favourite shirt' a certain sense of 'on' as 'wearing' is employed, our interpretations of the two sentences are quite different. Robert Wilensky echoes this point with respect to the word 'cut': "[W]hile we might postulate a specific sense of 'cut' that means 'slice', our interpretation of 'Cut the salami' and 'Cut the cake' will be rather different, even though the same sense seems to be in play" ("Primal Content and Actual Content" 169). Thus we can go along with Berg and acknowledge that words such as 'on' and 'cut' may have many different meanings, but deny that this threatens the essential strength of Searle's

---

15 This last point is interesting. Berg identifies the indexicality of 'on' with only one of its senses: that of indicating a certain spacial relation. What he proposes, then, is a highly irregular case in which an ambiguous word is an indexical on one of its interpretations, and not on others. If he is right, then this appears to set it apart from other indexicals which, unless they are appropriated for use as proper names, titles or the like, have no non-indexical senses.

16 Indeed, Berg identifies various senses of 'on' as in 'The fly is on the ceiling', 'The assignment is on the blackboard' and 'This paper is on the subject of literal meaning' ("Literal Meaning and Context" 399). To these we can add 'The hockey game is on television', 'The boss is away on a trip', 'There is a stain on the carpet', 'The missile is on target' and 'The house is on fire'. Doubtless there are others.
argument. As Wilensky puts it, "acknowledging polysemy affects where precisely [Searle’s] argument should be exploited, but it does not endanger its essential force" ("Primal Content and Actual Content" 169, insert mine).

**Katz's Responses to Searle**

Katz makes it his goal in “Literal Meaning and Logical Theory” to reply to Searle on behalf of the view that “sentences of natural language have a meaning independent of the social contexts in which their utterances occur” (203). He argues from two fronts that Searle misunderstands the nature and role of context-independent meaning according to the *orthodox position*.

First, Katz argues that in fact Searle defeats a position that no one holds: “His examples cast doubt on a strawman, but miss entirely the position on which ‘the literal meaning of a sentence is the meaning that it has in the “zero context” or the “null context’”’ (“Literal Meaning and Logical Theory” 219). Searle’s error, Katz argues, is to tack onto the principle claim that every sentence has, in virtue of its composition, a context-independent meaning, the further claim that this meaning must establish for all contexts whether an utterance of that sentence is literally true or false. Katz insists that the real position denies this claim. Instead, he argues, the *orthodox position* allows that in virtue of reference failure, for example, a sentence may be neither true nor false. Thus it is within the scope of the *orthodox position* to claim that the sentence ‘The cat is on the mat’ as uttered in outer space is neither true nor false because, “given the absence of up and down in outer space, neither the sentence-world relation required for truth nor
the sentence-world relation required for falsehood obtains" (Katz, "Literal Meaning and Logical Theory" 220).

Katz is right about this. Searle mistakenly supposes that the *orthodox position* is necessarily committed to the claim that all utterances are either true or false. Nonetheless, Katz is mistaken about the force of his criticism, likely because he focuses so much of his attention on Searle’s example of the cat floating freely on the mat in outer space, claiming erroneously that “Searle’s examples are all variants of [this] initial example” ("Literal Meaning and Logical Theory" 218, insert mine). To illustrate, consider instead Searle’s example of the cat that is drugged and drooped over the edge of the raised, stiffened mat. What Searle means to demonstrate by this example is that given one set of background assumptions, i.e., those of the neighbour who is unaware of the odd circumstances his cat faces, to tell him ‘The cat is on the mat’ is surely to mislead him. This, for Searle, is another way of saying that the utterance is neither clearly true nor clearly false ("Literal Meaning" 212). But, Searle claims, given the same scenario with one difference—both speaker and interpreter are aware of the unusual circumstances, i.e., they share the same set of background assumptions (but different than that of the neighbour)—an utterance of ‘The cat is on the mat’ is true. In both cases, the same ‘sentence-world relations’ obtain. Thus Searle’s point is not, as Katz suggests it is, simply that certain unusual circumstances might render the utterance of a sentence neither clearly true nor clearly false, but also that given the right set of background assumptions an utterance of the same sentence in the same unusual scenario might be clearly true or clearly false. The critical factor for Searle, then, is not the unusual nature of the
circumstances, but the background assumptions which we bring to bear on the situation. Indeed, the central point of Searle’s challenge is that we cannot evaluate the truth or falsity of an utterance without taking into account our background assumptions, including contextually specific ones.

Despite what Katz seems to think, Searle can grant that he was mistaken to attribute to all proponents of the orthodox position the claim that all utterances of all sentences are either true or false. To be sure, Searle’s argument does not depend fundamentally on his attributing to the orthodox position a strict stance on logical bivalence. He might as well have characterised the context-independent meaning of a sentence as determining for all contexts whether the sentence is true, false, or neither. Indeed, we need only look to Searle’s other works on language to find that he acknowledges the possibility of utterances that are neither true nor false.

Katz’s second criticism of Searle’s approach to literal meaning is that Searle expects too much of semantic content. Whereas both philosophers agree, contra Berg, that the semantic content of ‘cut’ in ‘Sally cut the cake’ and ‘Bill cut the grass’ is the same—that is, that the sorts of issues raised by Searle are not fundamentally issues of indexicality, ambiguity or the like—they disagree as to how this semantic content contributes to the truth conditions of a sentence in which it appears. On the one hand, Searle argues that despite employing the same semantic content, because of our background assumptions, in ‘Sally cut the cake’ and ‘Bill cut the grass’, ‘cut’ yields

\[ \text{indeterminate}. \]

\[ \text{17 Berg makes a similar point ("Literal Meaning and Context" 402).} \]

\[ \text{18 Searle's account of meaning in Speech Acts, for example, tolerates utterances whose truth is indeterminate.} \]
different sets of truth conditions such that if Sally were to cut the cake using a lawn mower, she would have failed to understand the literal meaning of the request. On the other hand, Katz claims that Searle is simply wrong that the semantic content of ‘cut’ makes a different contribution to the truth conditions of ‘Bill cut the grass’ and ‘Sally cut the cake’ (“Literal Meaning and Logical Theory” 223). Katz agrees that someone who ran over a cake with a lawn mower may have misunderstood the speaker’s intentions in uttering ‘Cut the cake’, but adds that speaker’s intentions are not the issue (“Literal Meaning and Logical Theory” 223). Rather, he contends, the way that the cutting is done is not part of the meaning or truth conditions of ‘Cut the cake’ or ‘Cut the grass’:

“The semantic contribution of ‘cut’… is the same, namely, the concept of dividing something (the pieces of grass in the one case and the cake in the other) with a sharp-edged instrument” (Katz, “Literal Meaning and Logical Theory” 223). According to Katz, then, whether one cuts the cake with a lawn mower, a hatchet or a knife one has satisfied the truth conditions set out according to the minimal semantic content of ‘cut’.19 This notion of minimal semantic content is, it would appear, a version of what we have called strict literal meaning.

Katz argues that Searle’s examples show only that the meaning of a sentence is but one of the factors entering into the meaning of an utterance. He contends that the supporter of context-independent meaning does not deny that background assumptions shape the meaning of an utterance of a sentence on a particular occasion, rather “the

19 The term ‘minimal semantic content’ is mine. It is meant to capture the essence of Katz’s view that we should attribute to words the barest semantic content possible. Thus, as we have found, for example, according to Katz ‘cut’ (in the appropriate sense) does not mean anything as robust as ‘slice’, but rather simply ‘to divide with a sharp-edged instrument’.
supporter denies only that such background assumptions are relevant to the meaning of sentences in the language” (Katz, “Literal Meaning and Logical Theory” 222).

According to Katz, Searle’s claim that the same sentence can be used, on the one hand, to deceive and, on the other, to convey one’s intentions accurately is of no threat to the orthodoxy position. Advocates of the orthodoxy position can argue that in both cases the sentence’s meaning is the same, but its contribution to utterance meaning is different.

Put differently, contra Searle, the orthodoxy position does not assume that the literal meaning of a sentence determines the conditions of satisfaction of use of that sentence as an utterance. Accordingly, there is no reason that sentence meaning alone should have to determine compliance with a speech act. Katz adds, “what counts as obeying particular orders cannot be supposed to reflect directly the meaning of the sentences that the speaker used to issue the orders” (“Literal Meaning and Logical Theory” 223). Thus, according to Katz, it is only because Searle is conflating what a speaker’s utterance of a sentence means with the separate matter of what the sentence used means (in the language) that he thinks that his examples cast doubt on the notion of context-free sentence meaning. Katz cites Searle’s frequent formulation, “‘What I meant in my literal utterance of the sentence’” as a “dead give away” that Searle has made this conflation (“Literal Meaning and Logical Theory” 222, Katz’s emphasis).

Katz’s invoking of a notion of minimal semantic content might address successfully the details of the challenge posed by Searle, but it does so only at the cost of introducing certain tensions into Katz’s own account. To begin, Katz himself seems to have trouble maintaining a consistent application of the notion of minimal semantic
content. On the one hand, for example, he argues that the truth conditions of 'Sally cut the cake' are satisfied even if the cutting is done in a non-standard way—for example, with a lawn mower. But, on the other hand, the truth conditions that Katz himself proposes for 'The cat is on the mat'—"that some (contextually specified) cat is vertically positioned over some (contextually specified) mat and that the aforementioned cat is also positioned so that its bottom is in contact with the top of the mat" ("Literal Meaning and Logical Theory" 220)—seem to preclude non-standard ways in which the cat might be 'on', i.e., supported by, the mat. Contra his proposed truth conditions, it would seem that according to Katz's stance on minimal semantic content he would want to allow that the sentence would be true if the cat were, for example, on the mat in the manner that a notice is on the bulletin board.

More important, Katz's appeal to minimal semantic content threatens the applicability of the anonymous letter scenario to his own account of literal meaning. Recall that according to Katz, the anonymous letter scenario is intended to elucidate the thesis that the null context is a context "whose features provide no relevant information for choosing a compositional meaning as the utterance meaning different from the compositional meaning of the sentence used" (Katz, "Literal Meaning and Logical Theory" 217n.). The appeal of Katz's anonymous letter scenario lies in its capacity to stand as a heuristic device with which to make sense of the orthodox position's claim that the literal meaning of a sentence is its meaning independent of any particular context of utterance. It offers us an understanding of what context-independent meaning is: the context independent meaning of a sentence is the meaning that a speaker of the language
would assign to that sentence were she to receive an anonymous letter containing it and only it. In virtue of his response to Searle, however, Katz commits himself to the conclusion that a seemingly straightforward sentence such as ‘Sally cut the cake’ has a meaning that is different than that which the typical speaker would attribute to it in the anonymous letter scenario. Consider the example, ‘When the neighbours were away on vacation, Bill took in the mail and cut the grass for them’.\textsuperscript{20} Were this to appear in an anonymous letter, one might conclude reasonably that, among other things, when the neighbours were away on vacation, Bill mowed their lawn. But according to Katz this is not part of what the sentence means. Thus, it seems, compositional meaning and the meaning that a typical speaker would ascribe to the sentence in the anonymous letter situation will sometimes diverge. Indeed, given the apparent uncontroversial nature of the examples thus far cited, one might even argue the stronger claim that sentence meaning and anonymous letter meaning will often diverge.

To be fair, Katz might protest that by invoking the notion of the ideal interpreter in his presentation of the anonymous letter scenario, he manages to avoid the above criticism.\textsuperscript{21} That is, Katz might argue that the ideal interpreter will recognise in accordance with his account that, for example, ‘Sally cut the cake’ means simply that Sally divide the cake with a sharp-edged instrument, \textit{however}. Likewise, although the typical interpreter might not recognise that it would be erroneous to conclude from the

\textsuperscript{20} The example is Berg’s ("Literal Meaning and Context" 409).

\textsuperscript{21} It is interesting to note that in its original formulation, Katz and Fodor formulate the anonymous letter scenario by appealing to “a number of English speakers” ("The Structure of a Semantic Theory" 478). It is later that Katz changes it to the “ideal speaker” of a language (Propositional Structure and Illocutionary Force 14).
meaning of the sentence 'When the neighbours were away on vacation, Bill took in the mail and cut the grass for them' that when the neighbours were away on vacation Bill mowed their lawn, the ideal interpreter would recognise this. Thus, Katz can argue, compositional meaning and 'anonymous-letter meaning' do not come apart after all.

Granted, by invoking the notion of an ideal interpreter, Katz avoids the above criticism. Nonetheless, by the same token he invites a different, perhaps more damaging criticism. Even assuming that Katz can give a non-circular formulation of the notion of an ideal interpreter, and it is not clear that he can, this response, in conjunction with the appeal to minimal semantic content, exposes the anonymous letter scenario, as characterised by Katz, as explanatorily inert.

Again, the purpose of the anonymous letter scenario is to provide a heuristic device with which to make sense of the orthodox position's concept of context-independent meaning. It succeeds in part because, prima facie, it does not assume that linguistic understanding takes place in isolation from general beliefs about the world. That is, the appeal of the anonymous letter scenario lies largely in that it does not invoke the strong, erroneous claim that when we read a sentence in isolation we use only linguistic knowledge to interpret it, but rather the weaker, less objectionable claim that when interpreting a sentence in isolation we have no reason to depart from sentence

---

22 That is, it is not clear that it is possible to explain the notion of an ideal interpreter without appealing (at least in part) to his or her ability to discern all lexical ambiguities, in which case the explanation risks circularity.
meaning. That the null context is not to be understood as a context in which one has access only to linguistic knowledge is an idea that Katz (in places) supports explicitly. For example, in response to a critic who attributes to him the position that in the null context there is nothing but the sentence itself, he explains that when he and Fodor first introduced the notion of a 'null' context, "'Null' was used to mean not void, but merely lacking in information on which to base a departure from sentence-meaning" ("Literal Meaning and Logical Theory" 217n.). But this is where the tension enters Katz's account. Unlike the typical interpreter, Katz's ideal interpreter is someone who computes meaning based solely on linguistic facts, without regard for general beliefs. Thus, although knowledge of the world provides the requisite clues to interpret 'Sally cut the cake' as 'Sally sliced the cake', and to conclude from 'When the neighbours were away on vacation, Bill took in the mail and cut the grass for them' that when the neighbours were away on vacation Bill mowed the lawn, Katz's ideal interpreter ignores this.

In light of his appeal to the ideal interpreter, Katz's anonymous letter scenario is in fact ill-suited to serve as a heuristic device for making sense of the claim that literal meaning is context-independent meaning. Put simply, it cannot do the work that Katz thought it could. Katz might concede that the anonymous letter scenario is no longer useful to him, but deny that this matters much. After all, it was intended primarily as a heuristic device—that is, he might contend that his account does not depend on its availability. But this would be to underestimate the role that it plays in making sense of

---

23 Although even this weaker claim might not always prove true. That is, even in the context of an anonymous letter, a single sentence might be taken as a malaprop, a metaphor, a 'slip of the tongue', an idiom, or the like.
the *orthodox position*’s claim that the literal meaning of a sentence is the meaning that it has independent of any particular context. Indeed, the reason we chose to direct our attention to Katz’s particular formulation of the *orthodox position* was that he appeared to provide, by means of the anonymous letter scenario, a device with which to make sense of the *orthodox position*’s concept of context-independent meaning. By invoking the notion of an ideal interpreter, however, the explanatory power which Katz’s presentation of the anonymous letter scenario appeared to provide, is lost.

*Searle and the Orthodox Position*

Searle’s examples serve to illustrate the fact that literal interpretation is a complex process. Further, he thinks, it is a process that often depends on invoking contextually specific background assumptions. Of course, this latter claim is unacceptable to proponents of the *orthodox position* who argue that literal meaning is to be identified with meaning independent of context. As we have found, responses from the *orthodox position*’s camp have been of two sorts. On the one hand, Berg, though acknowledging the complexity of literal interpretation, believes that all the factors that settle literal interpretation can be accounted for semantically in terms of complex and perhaps subtle ambiguities and indexicalities. On the other hand, and in contrast to Berg, Katz’s strategy is to strip semantics of all but the barest content. Both responses threaten the explanatory power of the anonymous letter scenario, without which the *orthodox position*’s notion of context-independent meaning remains abstract and perhaps unintuitive. To be sure, to deny, as Berg and Katz do, that the literal meaning of a
sentence is the meaning that a typical speaker of the language would assign to that sentence were it to appear by itself in an anonymous letter is to diminish significantly the appeal of the *orthodox position*.

Further, the strategies employed by Berg and Katz exhaust the responses to Searle that are available to advocates of the *orthodox position*. Searle argues that often when we interpret an utterance literally, we bring to the interpretation a set of contextually specific background assumptions which help determine the truth conditions of the utterance, and which cannot be incorporated semantically. For example, because of our assumptions about how people cut cakes, we interpret ‘Sally cut the cake’ as ‘Sally sliced the cake’ rather than, say, as ‘Sally ran over the cake with a lawn mower’. But the *orthodox position* denies that context plays a role in determining the truth conditions of a sentence, except, of course, where ambiguities or indexicalities occur. Thus a successful response to Searle on behalf of the *orthodox position* must argue either that the features of literal interpretation Searle’s examples illustrate can in fact be accounted for semantically—that is, that they are cases of ambiguity and/or indexicality—or that they cannot be accounted for semantically, but that this is of no threat to the *orthodox position* because truth conditions are not as robust as Searle claims they are. The first strategy is Berg’s; the second is Katz’s.

This is not meant to suggest that the *orthodox position* is untenable. It does, however, reveal that the *orthodox position*’s approach to the analysis of language is more abstract than it first appears and, so, it might be profitable to explore other possible approaches to understanding literal meaning.
Searle on Meaning

What of Searle's own account of literal meaning? One might expect to find a positive account of literal meaning in a paper entitled "Literal Meaning." Indeed, although he denies the orthodox position's identification of literal meaning with context-independent meaning, Searle insists that he is "not denying that sentences have literal meanings" ("Literal Meaning" 220). Nonetheless, Searle makes few positive claims about literal meaning. The thesis he advances is that "for a large class of sentences the speaker, as part of his linguistic competence, knows how to apply the literal meaning of a sentence only against a background of other assumptions" ("Literal Meaning" 222). But this says very little about the nature and origin of the literal meanings of words and sentences.

What we can piece together from Searle's other works on language is less than helpful. In Speech Acts Searle argues, contra the orthodox position, that linguistic meaning is partly conventional and partly intentional (Speech Acts 42-50). He presents his position as follows:

S utters sentence T and means it (i.e., means literally what he says) = S utters T and
(a) S intends (i-1) the utterance U of T to produce in H the knowledge (recognition, awareness) that the states of affairs specified by (certain of) the rules of T obtain. (Call this effect the illocutionary effect, IE)
(b) S intends U to produce IE by means of the recognition of i-1.
(c) S intends that i-1 will be recognised in virtue of (by means of) H's knowledge of (certain of) the rules governing (the elements of) T. (Speech Acts 49-50)

Leaving aside the question of how (and how convincingly) Searle argues this position, what is of immediate interest to our investigation is that it is inconsistent with the thesis
he touts in “Literal Meaning.” To begin, in the above formula there is no mention of the background assumptions to which Searle repeatedly refers in “Literal Meaning.” Indeed, according to the Searle of Speech Acts, the factors that decide meaning are, first, the rules within a linguistic community that govern the use of a sentence (or its components)—i.e., linguistic conventions—and, second, the recognition by both speaker and hearer that the speaker is employing those rules in her utterance of the sentence—i.e., speaker’s intentions and the recognition thereof. Conspicuously absent from this account is any mention of the variability of truth conditions given different contexts or, more specifically, different sets of background assumptions. Indeed, according to the formulation above, and contra Searle’s position in “Literal Meaning,” it is “(certain of) the rules of T”—i.e., the rules of use for the sentence uttered—that specify “the states of affairs” which render the utterance of T true or false. This precludes the “relativity of meaning” thesis Searle argues for in “Literal Meaning”.

What of Searle’s other more recent works? In “The Background of Meaning,” which he claims to be “a continuation of [the] line of investigation [found] in “Literal Meaning”,” Searle makes some brief but odd claims about the literal meaning of the word ‘cut’ (“The Background of Meaning” 221, inserts mine). There he denies that ‘cut’ occurs literally in ‘Sam cut two classes last week’, ‘The President cut the salaries of the employees’ and ‘The Raiders cut the roster to 45’ (222). Instead, Searle argues that the sense (or senses) in which ‘cut’ is used in these sentences represents a figurative extension of the literal meaning of ‘cut’ as it appears in, say, ‘The barber cut Tom’s hair’ and ‘The tailor cut the cloth’ (222). He cites the difficulty of translating directly into
another language, say, ‘Sam cut two classes last week’, and the absurdity of saying ‘Bill cut the grass, and Sam two classes’ as evidence that a non-literal sense of ‘cut’ is at work (“The Background of Meaning” 222).

Though Searle claims that the above observations are part of a continuation of the argument found in “Literal Meaning,” it would seem rather that they represent a retreat from his context-dependence stance on literal meaning. Indeed, Searle seems to appeal here to something like the orthodox position’s conception of context-independent meaning. Though his stated objective is to deny the orthodox position’s identification of literal meaning with context-independent meaning, and though he argues alternatively that literal meaning is relative to background assumptions and, so, to context, here he issues claims about the literal interpretations of sentences though these sentences are presented apart from any particular context of utterance.24 Further, one might find his claims about the literal meaning of the term ‘cut’ unusual. Does not ‘cut’ occur literally in ‘Sam cut two classes last week’, ‘The President cut the salaries of the employees’ and ‘The Raiders cut the roster to 45’? To be sure, the requisite ingredients of meaning that Searle cites in Speech Acts, intention and convention, are (or, rather, could be) present in an utterance of each of these sentences. Even according to the revised approach to meaning found in “Literal Meaning,” there is no reason to suppose that an utterance of any of these sentences could not qualify as literal. Indeed, Searle offers little in the way of argument to support his claim. Moreover, his contention that ‘Sam cut two classes last

24 One might respond that they are offered in the context of his discussion of literal meaning. But this is a non-starter, for in the context of his discussion of literal meaning they are intended to be interpreted apart from any particular application.
week’ does not translate easily into other languages is puzzling. It may be that one would not translate ‘cut’ in ‘Sam cut two classes last week’ using the same term as one would use for ‘cut’ in ‘The barber cut Tom’s hair’, but to cite this as evidence that the literal meaning of ‘cut’ in our language is limited to the sense employed in the latter sentence is to beg the question against the possible ambiguity of ‘cut’. It is not uncommon for a word to do double duty in one language but not in another. Similarly, the ambiguity of ‘cut’ is why we find it odd to say ‘Bill cut the grass, and Sam two classes’.

Without a philosophically sound account of literal meaning with which to back his claims, Searle’s argument is susceptible to various criticisms. First, though he denies it ("Literal Meaning" 221), by linking literal meaning to context, Searle risks failing to maintain the distinction between sentence meaning and speaker meaning. But this distinction is necessary to distinguish between what our sentences mean and what we do with them. Though Searle claims that to maintain the distinction we need only to “distinguish the special role of context of utterance in these cases from the role that background assumptions play in the interpretation of literal meaning,” he remains silent about how this is done ("Literal Meaning" 221). Second, by linking literalness to background assumptions, Searle’s account seems to allow that the same utterance may be true for one person and false or indeterminate for another. That is, Searle appears to associate truth with individual interpretations rather than with utterances. To illustrate, recall the example of the cat that is drugged and drooped over the edge of the raised, stiffened mat. Searle argues that on the one hand, to tell the unaware neighbour ‘The cat
is on the mat’ is to mislead him and, so, to issue an utterance that is neither clearly true nor clearly false. On the other hand, Searle argues that the same utterance is true if the audience is savvy to the unusual circumstances. Now suppose that both the unaware neighbour and the savvy audience pose the same question at the same time: ‘Where is the cat?’ According to Searle, the neighbour’s interpretation of the response, ‘The cat is on the mat’ is neither true nor false whereas the savvy audience’s interpretation is true. But then is the utterance ‘The cat is on the mat’ true or not?

Getting Back To Linguistic Basics

The orthodox position forces us to accept various disagreeable conclusions about language, not the least of which is that what a sentence means literally is not the interpretation that the typical speaker of a language would attribute to it in the absence of any particular context. But the alternative, Searle’s relativity of meaning thesis, though perhaps more inviting, lacks the support of a philosophically sound account of meaning. In light of this, I want to propose that a more plausible analysis of literal language will emerge if we change two of the basic assumptions with which we have been operating. First, rather than assume, as the orthodox position does, that meaning consists in something abstract and external (Fregean senses or the like), we will assume that meaning is the result of individual acts of communication (or, more precisely, individual attempts at communication). Second, rather than assume, as the orthodox position does, that every sentence has a context-independent meaning which is identified with its literal meaning, we will assume that every sentence, in its context, has a literal meaning. Together these
will permit us the freedom to tell a fuller story about linguistic interpretation and, so,
about literal meaning, a story that treats linguistic interpretation as but one inseparable
element of the broader phenomenon of interpersonal communication and, indeed, of
knowing our way around the world generally.
Chapter Five
Grice and the Foundations of Linguistic Communication

Up to this point in our investigation we have considered in detail only one approach to the analysis of language, the *orthodox position*. What is lost in this approach, however, is that language is only one inseparable element of interpersonal communication. At the end of the last chapter I proposed that a more plausible analysis of language would emerge if we were to change two of the assumptions with which we have been operating. First, the *orthodox position* assumes that meaning consists in something abstract and external. I propose that we supplant this conception of meaning with one that treats meaning as the result of individual acts (or attempted acts) of communication. Second, the *orthodox position* assumes that every sentence has a context-independent meaning which is identified with its literal meaning. I propose that we reject this assumption in favour of the idea that every sentence, in its context, has a literal meaning. Though this latter change frees us from the burden of providing an adequate explanation of the notion of context-independent meaning, nevertheless, we are still obliged to explain how it is that we understand sentences in cases like the anonymous letter scenario.

Thus, included among the criteria for our new account of literal meaning are the following: (i) that it treat meaning as originating with individual acts of communication, not abstract ‘meanings’, (ii) that it acknowledge that context plays a critical role in all cases of linguistic interpretation, and (iii) that it treat linguistic communication as an unisolatable element of the broader phenomenon of interpersonal communication. Moreover, the success of this new approach will depend in part on its ability to replicate
the orthodox position’s success at explaining important features of natural language such as meaning relations and logical structure, our ability to create and interpret novel utterances and how the same sentence can mean the same thing for different people.

*The Foundation: Grice*

We find the foundation for our new approach to understanding literal meaning in the works of H.P. Grice. Grice’s approach to understanding meaning is unique because, unlike the orthodox position’s approach, it is not directed toward an analysis of literal sentential meaning in particular, but rather toward an analysis of how people can and do mean things by their actions in general. Indeed, Grice’s approach presupposes nothing about language.

Grice’s work is divided roughly into two branches. On one, he explores a distinction between *saying* and *implicating*.1 On the other, he explores a distinction between *utterance-type meaning* (roughly, sentence meaning) and *utterer’s occasion meaning* (speaker meaning).2 Each branch is of interest to our investigation in its own right. And, as we will find, an exploration of the relation between the two reveals more clearly what is necessary to construct a suitable account of the distinction between literal and non-literal language.

---

1 This branch originates with “Logic and Conversation” and is continued in “Further Notes on Logic and Conversation.”

2 This branch finds its origin in Grice’s seminal piece on meaning, “Meaning,” and is taken up in detail in “Utterer’s Meaning and Intentions” and “Utterer’s Meaning, Sentence-Meaning and Word-Meaning.”
Branch One: Saying and Implicating

In “Logic and Conversation,” Grice observes that sometimes we utter the sentence ‘p’ in order to communicate a message other than that p. Thus he claims that for a large class of utterances we must distinguish between what is said (in a certain favoured sense of ‘said’), and what is implicated. On the route to explaining the distinction, he asks the reader for some latitude: “I shall, for the time being at least, have to assume to a considerable extent an intuitive understanding of the meaning of say in such contexts,” though he adds, “I intend what someone has said to be closely related to the conventional meaning of the words (the sentence) which he has uttered” (“Logic and Conversation” 66). Grice intends the term ‘implicate’ (and the related noun ‘implicature’) to stand for a range of terms, including ‘suggest’, ‘imply’ and ‘indicate’ (“Further Notes on Logic and Conversation” 42).

Dan Sperber and Deirdre Wilson illustrate Grice’s distinction well with the example of someone who asks, ‘Would you like more coffee?’ to which one responds ‘Coffee keeps me awake’ (Relevance: Communication and Cognition 34). In one context, this utterance may convey the message that you do not want any more coffee. In a different context, say, one in which you are obliged to stay awake during a boring lecture, it may convey the message that you do want more coffee. In any case, neither of these messages is part of what you said; in each case you said only that coffee keeps you awake. In each case the message that you intend to communicate by means of your utterance is not said, it is implicated.

That we arrive easily at the understanding, in one context, that you mean to
convey by your utterance that you would like more coffee, and, in another, that you
mean to convey that you would not like any more coffee, is explained by Grice in terms
of certain maxims that govern conversational exchanges. Grice observes that

conversations are cooperative events:

Our talk exchanges do not normally consist of a succession of disconnected
remarks, and would not be rational if they did. They are,
characteristically, to some degree at least cooperative efforts. ("Logic and
Conversation" 66)

Further, Grice argues, each member of a (cooperative) conversation abides by certain
implicit principles or maxims which promote success in any conversational exchange. He
groups this set of maxims under the heading the Cooperative Principle: "Make your
conversational contribution such as is required, at the stage at which it occurs, by the
accepted purpose or direction of the talk exchange in which you are engaged" ("Logic
and Conversation" 67). To this general principle Grice adds nine maxims in four
categories:

Maxims of quantity:
1. Make your contribution as informative as is required (for the current
purposes of the exchange).
2. Do not make your contribution more informative than is required.

Maxims of quality:
Supermaxim: Try to make your contribution one that is true.
1. Do not say what you believe to be false.
2. Do not say that for which you lack evidence.

Maxim of relation:
Be relevant.

Maxims of manner:
Supermaxim: Be perspicuous.
1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief (avoid unnecessary prolixity).
4. Be orderly. ("Logic and Conversation" 67)³

According to Grice, conversational implicatures are worked out by using the Cooperative Principle and its maxims in conjunction with information about the sentence used, contextual information and other background assumptions. Roughly, if I say sincerely that \( p \)—that is, I say that \( p \) and am abiding by the Cooperative Principle and its attendant maxims—then the conversational implicature that \( q \) may be inferred by the audience if the supposition that I believe that \( q \) is required to make my saying that \( p \) consistent with the presumption that I am abiding by the Cooperative Principle and its maxims.⁴ Grice explains, "[W]hat is implicated is what it is required that one assume a speaker to think in order to preserve the assumption that he is observing the Cooperative Principle (and perhaps some conversational maxims as well), if not at the level of what is said, at least at the level of what is implicated" ("Utterer’s Meaning and Intentions" rev. 86). At the boring but important lecture you notice that I am nodding off and ask whether I would like some coffee, to which I respond 'Coffee keeps me awake'. Because you assume that I am abiding by the Cooperative Principle and its maxims, in particular the maxims 'Be relevant' and 'Make your contribution as informative as is required', you infer that yes, I would like some coffee.

³ Grice adds, “There are of course all sorts of other maxims (aesthetic, social, or moral in character) such as 'Be polite,' which are also normally observed by participants in talk exchanges, and these may also generate nonconventional implicatures” ("Logic and Conversation" 67). Further, the observation of some maxims is clearly more important than the observance of others—for example, it seems less shocking that someone should use an obscure expression than that they should assert something which they believe to be false.

⁴ Of course, if one opts not to abide by the Cooperative Principle, then all implicatures are cancelled.
Implicatures are generated in various ways. They may involve a failure to fulfil a maxim by violating, flouting or opting out of it. For example, Grice offers the case of a professor who writes in a reference letter for a student applying for a job, “Dear Sir, Mr. X’s command of English is excellent, and his attendance at tutorials has been regular” ("Logic and Conversation" 71). The reader infers from the obviously lacklustre appraisal that the author is trying to convey information that he is reluctant to write down, i.e., that he thinks Mr. X is not very good at philosophy. Implicatures also can occur when no maxim is violated. Grice offers the following exchange as an example:

A: “Smith doesn’t seem to have a girlfriend these days.”
B: “He has been paying a lot of visits to New York lately.”
("Logic and Conversation" 70)

In this case B implicates that Smith has (or may have) a female companion in New York, though no maxim is obviously violated.5

Grice’s model is not perfect. Martin Davies notes, for example, that it underdetermines implicatures because it does not rank maxims, nor account for how they interact ("Pragmatics: Conversational Implicature and Relevance" 129). Nor does it acknowledge that whereas some implicatures are conveyed strongly, others are conveyed weakly. Nonetheless, for the purposes of our investigation, our interest is in Grice’s general approach rather than in its details.

---

5 Indeed, even the supermaxim ‘Be perspicuous’ does not seem to be violated given that the connection between B’s remark and A’s is so obvious.
**Saying vs. Implicating and the Literal/Non-Literal Distinction**

Despite its imperfections, Grice’s analysis of conversational implicature is of interest to our investigation because it suggests conceptualising the literal/non-literal distinction in terms of a distinction between what is *said* and what is *conversationally implicated*. That is, it suggests that we might identify the literal element of an utterance with what a speaker *says* and all non-literal elements with what the speaker *conversationally implicates*.

Indeed, this is how Grice treats figurative language. He discusses irony, metaphor, meiosis and hyperbole under the heading *Examples in which the first maxim of Quality [Do not say what you believe to be false] is flouted* (“Logic and Conversation” 71, insert mine). He uses the example of A whose friend X has recently betrayed him. A says (ironically) ‘X is a fine friend’. Grice explains,

> It is perfectly obvious to A and his audience that what A has said or has made as if to say is something which he does not believe, and the audience know that A knows that this is obvious to the audience. So, unless A’s utterance is entirely pointless, A must be trying to get across some proposition other than the one he purports to be putting forward... [T]he most obviously related proposition is the contradictory of the one he purports to be putting forward. (“Logic and Conversation” 71)

Similarly, metaphors often involve flouting the maxim ‘Do not say what you believe to be false’. Grice discusses the example, ‘You are the cream in my coffee’: “The most likely supposition is that the speaker is attributing to his audience some feature or features in

---

6 In “Further Notes on Logic and Conversation” Grice admits that this treatment of irony is too brief. Often one employs an ironical utterance to convey more than just the contradictory of the sentence uttered. In this case, Grice writes, the interpreter might take the speaker to mean something like *He is, usually, a fine friend: how could he have treated me like that?* (53).
respect of which the audience resembles (more or less fancifully) the mentioned substance” ("Logic and Conversation" 71). We can explain similarly how hyperbole works. Suppose that I turn off the radio in disgust proclaiming ‘I’ve heard that song a million times’. What I have said is obviously false. Nonetheless, because you assume that I am abiding by the Cooperative Principle, you infer that I mean that I have heard that song too many times to want to listen to it any more.

Note, however, that we cannot identify particular figurative devices by the way in which they generate conversational implicatures. Though many metaphors, for example, flout the maxim Do not say what you believe to be false, others do not—for example, ‘No man is an island’. Also, interestingly, according to this approach simile is not clearly a non-literal device. In the case of simile, it would seem, conversational implicature is not necessarily involved. A simile relies not on provoking us with a literal untruth or apparently irrelevant claim, but rather by making a surprising (literal) comparison. If I say, for example, ‘The dark clouds are like a cold blanket’ I have said, and I mean, that the dark clouds are like a cold blanket.

**Literal/Non-Literal vs. Literal/Figurative**

Of further interest to us is that conceptualising the literal/non-literal distinction in terms of what is said and what is conversationally implicated vindicates our earlier decision to characterise our topic as that of the distinction between literal and non-literal language as opposed to the distinction between literal and figurative language. Recall that we characterised a literal utterance as an utterance in which the principal intention of the
speaker is to communicate what she says. Grice's work on conversational implicature reveals that not all utterances which do not convey as the intended message their literal meaning convey a figurative meaning. When you ask whether I will attend the party, I might respond enthusiastically, 'Cathy will be there!', thus communicating my affirmative answer without actually saying 'yes'. Accordingly, figurative language is but one kind of implicature. Put differently, the literal/non-literal distinction might be best understood as the distinction between the literal meaning of an utterance and the use to which that literal meaning is put, of which figurative devices are but one class of uses of literal language.

Conventional Implicature

The utility of Grice's account of implicature to our investigation is further enhanced by the finer distinction he makes between conversational and conventional implicature. He observes, "Sometimes one can say that the use of a certain form of words in an utterance would normally (in the absence of special circumstances) carry such-and-such an implicature or type of implicature" ("Logic and Conversation" 73). Though he admits that it may be difficult to find non-controversial examples, he has in mind, for example, words which are natural language analogues to logical constants but which appear to diverge in meaning from the related constant. 'But' is a good example. Though 'but' is symbolized in formal systems as a conjunction, in natural language sometimes it carries with it the sense that the second conjunct is intended to qualify the first ('I will go to the party but I won’t drink any alcohol', ‘The contractors appeared competent but their work was shoddy'). Similarly, some uses of 'and' suggest temporal order ('He took off his
clothes and went skinny dipping in the pool’, ‘They were married and had a baby’). In such cases, reversing the order of the conjuncts, though a harmless alteration in logic, changes how each sentence is understood. Grice adds that conventional implicatures also can arise from the stress of certain words—for example, ‘I know that’ (implicating that I did not merely believe it) (“Further Notes on Logic and Conversation” 52). A conventional implicature, he argues, is the result of a regularized conversational implicature. That is, when members of a linguistic community regularly conversationally implicate the same thing by the same word or phrase or sentence, the implicature becomes a conventional implicature. Though, he admits, identifying when exactly this occurs is impracticable.

Thus ultimately Grice’s analysis yields three possible elements of every utterance: what is said, what is conventionally implicated and what is conversationally implicated, though for any given utterance one or both kind of implicature may be absent.

**Implicature and ‘Strict’ and ‘Communicated’ Literalness**

In the Introduction we observed that there are at least two ways in which we employ the term ‘literal meaning’. We illustrated the distinction with the example of a sentence with noticeable roots in metaphor: ‘Ian’s boss gave him the boot’. On the one hand, because it is idiomatic, we might agree that this sentence means literally that Ian was fired by his

---

7 The so-called exclusive use of ‘or’ is another example. Interestingly, there also appear to be cases of ‘or’ that suggest temporal order. Suppose, for example, that today is Wednesday. The statements, ‘We should meet on Friday or on Monday’ and ‘We should meet on Monday or on Friday’ seem to suggest two distinct sets of possibilities—the first suggests that we meet either in two days or in five; the second suggests that we meet either in five days or in nine.
boss. On the other hand, we might also agree that it means (strictly) literally that Ian's
boss gave him the boot—i.e., his boss gave him a certain item of footwear. The first
interpretation is what I have called the *communicated literal* meaning—that is, the
meaning that the speaker intends the interpreter to assign to her utterance. The second
interpretation is what I have called the *strict literal* meaning—roughly, how a logician or
an unforgiving lawyer might interpret the utterance.

*Prima facie*, Grice's distinction between what someone says and what she
conventionally implicates might serve as an explication of our distinction. The idea is
this: *strict literal* meaning is identified with what someone *says*; *communicated literal*
meaning is identified with what someone says plus what her words conventionally
implicate. Consider our earlier examples. Someone who utters 'He took off his clothes
and went skinny dipping in the pool' has *said* 'He took off his clothes and went skinny
dipping in the pool'. This, it would seem, corresponds to its *strict literal* meaning (the
conjunction of the two events, taking off his clothes and skinny dipping in the pool). But
by saying 'He took his clothes off and went skinny dipping in the pool', one
conventionally implicates a temporal order of the events—first he took off his clothes and
then he went skinny dipping in the pool. This, it would seem, is its *communicated literal*
meaning. Similarly, one might claim that the *strict literal* meaning of 'The contractors
appeared competent but their work was shoddy' is that the contractors appeared
competent *and* their work was shoddy, which corresponds to what is said, but that its
*communicated literal* meaning is something like that the contractors appeared competent
and so it was to the speaker's surprise that their work was shoddy.
If this is right, then it can be used to explain why we attribute to some metaphors a kind of literal status. A metaphor acquires literal status when the implicature which it is regularly used to communicate becomes conventionally embedded in the utterance of the metaphor. Thus, although ‘Ian’s boss gave him the boot’ means strictly literally that Ian’s boss gave him a certain item of footwear, it has the communicated literal meaning that Ian was fired by his boss because the expression ‘give him the boot’ and its variants have been used regularly by members of our linguistic community to conversationally implicate that someone has been (or is being, or will be) fired. In contrast, the implicatures generated by the utterance of a fresh metaphor cannot be said to contribute to the literal meaning (strict or communicated) of the sentence uttered because such implicatures are non-conventional—rather, they are conversational.  

It would be of great utility to our investigation if the above proposal proved sound, if only to provide a framework within which to explore the notions of strict and communicated literal meaning. But there is a certain artificiality to Grice’s distinction. Grice’s insistence that what is conventionally implicated by a sentence will never be part of what someone says by uttering the sentence signals a failure to understand the causal

---

8 Having said this, Grice seems to intend his notion of conventional implicature to extend beyond the sorts of cases that fit with our distinction between communicated and strict literal meaning. The first example he puts forth to illustrate the concept of a conventional implicature is the following: “Anyone who uses a sentence of the form ‘X is meeting a woman this evening’ would normally implicate that the person to be met was someone other than X’s wife, mother, sister, or perhaps even close platonic friend” (“Logic and Conversation” 73). It is doubtful that we would want to say that part of the communicated literal meaning of ‘X is meeting a woman this evening’ is that the woman in question is not X’s mother, sister, or close platonic friend. Put differently, we would not be inclined to find the utterance false (on its strict or its communicated literal interpretation) if we found that the woman that X was meeting was in fact his mother or sister or close platonic friend.
mechanism (or at least a causal mechanism) by which languages evolve, namely, metaphor. Much so-called literal language finds its causal history in metaphor—indeed, how else can one make a novel observation except by comparing it to that which is familiar? Given that this is the case, and that Grice precludes the possibility of dead metaphors acquiring 'says status', it is unclear exactly how Grice conceives of the origin and nature of what is said.

Indeed, this pin-points a weakness in Grice's analysis as it is thus far presented: he has not yet explained adequately his favoured sense of say. Grice knows this. In fact, it is his desire to meet this concern that provides the link to his other branch of work on language, that on meaning.

Branch Two: Meaning

Grice's first significant contribution to the study of meaning is to distinguish between the concept of meaning relevant to communication and the sort of meaning involved in claims like 'Those clouds mean rain' and 'That frown means he is unhappy' ("Meaning"). Meaning of the former sort he terms non-natural; meaning of the latter sort he terms natural. Natural meaning involves causal co-variation between two kinds of states of affairs. That is, if $x$ means naturally that $p$ then $x$ is a symptom of, or evidence for, $p$. In such cases, that $x$ means $p$ is independent of our interpretation. For example, certain kinds of spots indicate measles whether or not anyone takes them to mean measles. In the case of non-natural meaning, however, no causal co-variation exists. We say $x$ means non-naturally that $p$' only if $x$ originates with an individual, rather than with an
event or natural state of affairs. That is, for \( x \) to mean non-naturally that \( p \), \( x \) must be a verbal utterance, a gesture, a facial expression, or the like.\(^9\) Grice does not mean to argue that “all our uses of ‘mean’ fall easily, obviously, and tidily into one of the two,” rather that in most cases we would be “fairly strongly inclined” to categorise a particular use of ‘mean’ as either natural or non-natural (“Meaning” 22).

Meaning as Intending

Having distinguished between natural and non-natural meaning, Grice’s primary interest becomes that of understanding what is essential to non-natural meaning.\(^{10}\) Meaning, Grice contends, is a kind of intending: “[T]he meaning (in general) of a sign needs to be explained in terms of what users of the sign do (or should) mean by it on particular occasions” (“Meaning” 24). Thus Grice takes the orthodox position’s supposition that meaning gives life to understanding and turns it on its head. As a result, as Jonathan Bennett notes, Grice’s approach “treats of what an individual means by an utterance on a particular occasion, without reference to what anybody did, does, will or would mean on any other occasion” (Linguistic Behaviour 9). Of interest to us, then, is that this opens the way to an analysis of linguistic meaning whereby meaning originates with individual

---

\(^9\) Grice intends this distinction to improve upon the more traditional distinction between ‘natural’ and ‘conventional’ signs. He argues that the natural/conventional distinction is inadequate because it fails to acknowledge that “some things which can meanNN [i.e., mean non-naturally] something are not signs (e.g., words are not), and some are not conventional in any ordinary sense (e.g., certain gestures); while some things which mean naturally are not signs of what they mean” (“Meaning” 23, insert mine).

\(^{10}\) From this point forward, when I use the term ‘mean’ and its variants in an unqualified way, I mean non-natural meaning.
acts of communication rather than with rules or conventions or abstract 'meanings'.

What is Grice's analysis? As a first approximation, he says that for someone to mean something by $x$ she must intend to induce by $x$ a belief in the audience, and intend her utterance to be recognised as so intended. Suppose that I have just returned from vacation and my neighbour and I are standing on my lawn. My neighbour is puzzled by the fact that my lawn has been mowed recently because he knows that I just returned from out of town. I tell him 'Bill cut the grass'. According to Grice, I mean that Bill mowed the lawn because I intend that my neighbour believe, and that he recognise that I intend by my utterance that he believe, that Bill mowed the lawn. Similarly, to borrow an example from Bennett, suppose that I am at the opera and I look across at a friend sitting several seats away; our eyes meet and she grimaces in an exaggerated manner and holds her nose (Bennett, *Linguistic Behaviour* 12-13). Obviously she means to communicate that she abhors the performance. According to Grice's account, her gesture leads me to believe that she abhors the performance just because that is what I take her to be trying to get me to believe.\(^\text{11}\)

Generalising from examples like these, Grice settles on the following approximation of a criterion for meaning: "'A meant$_{\text{NN}}$ something by $x$' is (roughly) equivalent to 'A intended the utterance of $x$ to produce some effect in an audience by means of the recognition of this intention'" ("Meaning" 27). More perspiciously, we can say that $U$ utters $x$ to audience $A$ meaning that $p$ is analyzed by Grice as:

\(^\text{11}\) In both cases, of course, there are the Gricean conversational constraints.
U uttered x intending:
(i) A to believe that p;
(ii) A should be aware that U intends (i);
(iii) the awareness mentioned in (ii) should be part of A's reason for believing p.12

I will call this the *Gricean analysis of understanding*. It is important to note, as Steven Schiffer does, that the *Gricean analysis of understanding* "does not provide a criterion for determining what S [the speaker] meant, but only a criterion for determining what must be determined if one is to determine what S meant" (*Meaning* 13-14, insert mine). In any case, it is of interest to us because it does not itself make use of the notion of literal meaning and, so, is available for use in a non-circular analysis of this notion.

Grice adds that though we might intend by an utterance a certain effect by which we intend to produce some further effect, his analysis applies only to the first intended effect. That is, he intends the connection between meaning and recognition of intention to apply only to what he calls the primary intention of the utterer: "[I]f I utter x, intending (with the aid of the recognition of this intention) to induce an effect E, and intend this effect E to lead to a further effect F, then insofar as the occurrence of F is thought to be dependent solely on E, I cannot regard F as in the least dependent on recognition of my intention to induce E" (*Meaning* 28). Though Grice does not offer an example to elucidate the distinction between the primary and subsidiary intended effects of an utterance, he seems to have in mind something like this. I might intend you to believe by my utterance 'He has a gun' that I believe that the man facing us has a gun,

---

12 The account needs to be tweaked to handle non-indicatives. For imperatives, for example, the utterer intends the hearer to intend to do something ("Utterer's Meaning, Sentence-Meaning, and Word-Meaning" 230).
and intend you to believe further that I believe that because he has a gun you should hand over your wallet promptly and peacefully. Though you may recognise both of these intentions, according to Grice, his analysis applies only to the former, not to the latter.

Refinements to Grice's account have been suggested both by others and by Grice himself. Grice, for example, amends his account to read not that the intended effect of an utterance is that $A$ is to believe that $p$, but rather that $A$ should believe that $U$ believes that $p$ ("Utterer’s Meaning, Sentence-Meaning, and Word-Meaning" 230). This allows him to admit of utterances by which we intend the audience to believe only that we believe that $p$, as distinguished from those by which we intend the audience to notice not only that we believe that $p$ but also, by means of our utterance, to induce in the audience the belief that $p$.

Taking Stock: How Well Does Grice's Model Fit Our Criteria?

Thus far Grice's analysis of meaning meets our criteria for a new account of literal meaning. To begin, it fits with our supposition that meaning originates with individual acts of communication. According to the Gricean analysis of understanding, meaning originates with intention, and so each utterance is treated individually. In the case of Bennett’s friend’s grimace at the opera, for example, nothing like conventional meaning or senses is involved: “Gestures like my friend’s are in common use; but their success as vehicles of meaning does not depend on that, for someone might understand such a

---

gesture perfectly on his first encounter with this sort of behaviour” (Bennett, *Linguistic Behaviour* 14). Also, Grice’s analysis of meaning allows a critical role for context. Unlike the *orthodox position*, it does not treat sentences in isolation from their occasions of utterance. Grice explains, “[I]n cases where there is doubt, say, about which of two or more things an utterer intends to convey, we tend to refer to the context (linguistic or otherwise) of the utterance and ask which of the alternatives would be relevant to other things he is saying or doing” (“Meaning” 29). Moreover, Grice’s analysis does not treat linguistic meaning in isolation from other means of communicating. Indeed, his analysis applies equally to non-verbal behaviour. Grice himself uses the term ‘utterance’ broadly to include gestures, diagrams, and other meaningful non-verbal behaviour (or, in the case of diagrams and the like, *results* of non-verbal behaviour). He writes, “surely to show that the criteria for judging linguistic intentions are very like the criteria for judging nonlinguistic intentions is to show that linguistic intentions are very like nonlinguistic intentions” (“Meaning” 29). Thus Grice’s account allows us to situate knowing how to understand one another’s utterances in the broader context of knowing how to make our way around the world generally. That is, it acknowledges that interpretation is not achieved in a vacuum; rather, it involves attention to previous conversation, the attribution of beliefs and desires, the use of background information, and so forth.

Nevertheless, it is worth noting that there are two kinds of cases which resist the *Gricean analysis of understanding*. The first is the case in which there is no intended audience—for example, when I shout (a sentence) in frustration though I believe that no one else is within earshot. The second is the case in which the author’s intentions are
unknowable, as in, for example, the anonymous letter scenario.

**Non-Natural Meaning, Speaker Meaning and Sentence Meaning**

But Grice’s analysis is not yet complete. His account of non-natural meaning provides us with an analysis of speaker meaning (or what Grice prefers to call *utterer’s occasion meaning*), but not of sentence meaning (roughly, what Grice prefers to call *utterance-type meaning*). What of sentence meaning? In “Utterer’s Meaning, Sentence-Meaning, and Word-Meaning” Grice directs his attention “towards an explication of the favoured sense of ‘say’ and a clarification of its relation to conventional meaning” (“Utterer’s Meaning, Sentence-Meaning, and Word-Meaning” 225). Here he aims to elucidate the connection between the notion of meaning which he regards as basic—i.e., non-natural meaning—and the notions of meaning involved in saying that a certain sentence (or word) has a particular meaning (“Utterer’s Meaning, Sentence-Meaning, and Word-Meaning” 225). His analysis of non-natural meaning fixes utterer’s occasion meaning, or the sense of meaning involved in locutions of the sort ‘U (utterer) meant that...’, but not locutions of the sort ‘X (utterance-type) means “...”’. Grice is concerned with locutions of the latter sort because they are closely related to his favoured sense of ‘say’.

Grice’s strategy is as follows: from an adequate account of how it is that someone means $p$ by a particular utterance $X$ on a particular occasion, it should be a short step to an account of how $X$ means in general that $p$ (for some linguistic community). Grice begins with a discussion of a friend whose policy it is to use a particular hand wave (H-W) on occasions when he intends to communicate to his audience that he knows the
route. We might specify the policy as follows: ‘I (that is, utterer U) shall utter H-W if I intend some audience A to think that I think I know the route’. Grice adds, “Now, if U is ever to have the particular intentions which will be involved in every implementation of this policy, he must (logically) be in a position... to suppose that there is at least some chance these intentions will be realised” (“Utterer’s Meaning...” 232). But on what basis would this supposition be justified? Grice answers: “[A]s U well knows, a given audience A must be aware of U’s policy and must suppose it to apply to the utterance of H-W with which U has presented him” (“Utterer’s Meaning, Sentence-Meaning, and Word-Meaning” 232). Grice cautions that because U’s H-W might have another meaning, and because U might have other ways of getting A to think that he thinks that he knows the route, the definition of what H-W means for U cannot be simply that it is his habit to utter H-W if, for some audience A, he intends A to think that he thinks that he knows the route. To contend with these difficulties, Grice appeals to the idea of “having a certain procedure in one’s repertoire” (“Utterer’s Meaning, Sentence-Meaning, and Word-meaning” 233). Thus equipped, he offers the following definition:

For U, H-W means ‘I know the route’ =ₕ. U has in his repertoire the following procedure: to utter a token of H-W if U intends A to think that U thinks that U knows the route. (“Utterer’s Meaning, Sentence-Meaning, and Word-meaning” 233)

Thus it is a short step to the following definition of utterance-type meaning for H-W for a group or class of individuals:

For group G, utterance-type H-W means ‘I know the route’ =ₕ. At least some (?) many) members of group G have in their repertoires the procedure of uttering a token of H-W if, for some A, they want A to think that they think that they know the route. (“Utterer’s Meaning, Sentence-Meaning, and Word-Meaning” 233)
Grice acknowledges that explicating the notion of *having a procedure in one's repertoire* is considerably difficult, and abandons the attempt to provide a definition. He does, however, suggest a clue. He says that one may speak legitimately of an established procedure in respect of utterance-type $X$ for group $G$ if to utter $X$ in such-and-such circumstances is part of the practice of many members of $G$ ("Utterer's-Meaning, Sentence-Meaning, and Word Meaning" 234). Recent efforts to further Grice's programme attempt an explication of *having a procedure in one's repertoire* in terms of a notion of linguistic convention as a rationally self-perpetuating regularity in behaviour, drawing on David Lewis's work in *Convention*. According to Lewis, a convention is a regularity $R$ in action, or in action and belief, that meets the following conditions:

1. everyone conforms to $R$;
2. everyone believes that everyone else conforms to $R$;
3. the belief in (2) furnishes each person with a reason to conform to $R$;
4. there is a general preference for general conformity to $R$, rather than slightly less than general conformity;
5. there is at least one alternative regularity which would serve reasonably well; and
6. the facts listed in (1)-(5) are matters of common knowledge.

(Lewis, *Convention*)

Thus the solution is, roughly, that an utterance-type $X$ means $p$ if it is a convention (understood as above) of the members of a linguistic community (or the habit of an individual in the case of idiomatic utterances) to utter $X$ with the intention of inducing the belief that $p$ (or at least the intention of inducing the belief that the speaker believes that $p$), based upon the intention-based mechanism which governs non-natural meaning (i.e., the *Gricean analysis of understanding*). Utterance-type meaning, as Simon Blackburn

---

puts it, is "fossilized one-off meaning" (Spreading the Word 112).

**Connecting the Branches: ‘Say’ and Utterance-Type Meaning**

How does Grice’s analysis of utterance-type meaning relate to his favoured sense of ‘say’? One might suppose that what someone says on a particular occasion is to be identified with the meaning of the utterance-type of which that utterance is a token. Indeed, recall that in “Logic and Conversation” Grice claims that the sense in which he uses the term ‘say’, what someone has said “is closely related to the conventional meaning of the words (the sentence) which he has uttered” (66). Nevertheless, this is not quite right. Grice’s definition of ‘say’ suggests that we need to be more careful.15 He proposes for discussion the claim that ‘U (utterer) said that p’ entails

\[ U \text{ did something } x \text{ (1) by which } U \text{ meant that } p \text{ [and] } \]
\[ \text{ (2) which is an occurrence of a type } S \text{ which means ‘} p \text{’ in some linguistic system.} \]
\[ \text{ (“Utterer’s Meaning and Intentions” (rev) 88, insert mine).} \]

But he argues that this definition is too wide. Consider the sentence ‘She was poor but she was honest’.

What \( U \) meant, and what the sentence means, will both contain something contributed by the word ‘but’, and I do not want this contribution to appear in an account of what (in my favoured sense) \( U \) said (but rather as a conventional implicature). (“Utterer’s Meaning and Intentions” (rev) 88)

Thus Grice introduces the notion of the central meaning of a sentence. Roughly, Grice

---

15 Interestingly, this definition is absent from the original version of “Utterer’s Meaning and Intentions,” but appears in the revised version found in Studies in the Way of Words (referred to in the text as “Utterer’s Meaning and Intentions” (rev)). In any case, there is a similar though less detailed discussion of ‘say’ in “Utterer’s Meaning, Sentence-Meaning, and Word-Meaning” (227-228).
wants to associate the *central meaning* of a sentence with what someone *says* and the 
(non-central) meaning of an utterance-type with what someone says plus what her words 
conventionally implicate.\(^{16}\) Thus ‘*U* said that *p*’ finally comes out as meaning

\[
\begin{align*}
U \text{ did something } & x \quad (1) \text{ by which } U \text{ centrally meant that } p \quad [\text{and}] \\
(2) \text{ which is an occurrence of a type } S \text{ part of the meaning of which is } 'p'. \\
& \text{ (“Utterer’s Meaning and Intentions” (rev) 88, insert mine)}
\end{align*}
\]

*The Unstated Connection: Utterer’s Occasion Meaning and Implicature*

Connecting Grice’s branches, thus far we have something like this:

<table>
<thead>
<tr>
<th>Branch One</th>
<th>Branch Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is <em>said</em></td>
<td>Central meaning (utterance-type meaning less</td>
</tr>
<tr>
<td></td>
<td>conventional implicature)</td>
</tr>
<tr>
<td>What is <em>implicated</em></td>
<td>Utterer’s occasion meaning</td>
</tr>
</tbody>
</table>

Grice links the two branches of his work on language by arguing that what someone says 
is identified with utterance-type meaning less any conventional implicatures (perhaps what 
we have called *strict literal* meaning). But he says nothing of the connection between 
conversational implicature and utterer’s occasion meaning. Nevertheless, it would 
appear, this connection is significant in the case of figurative language. In the case of 
metaphor, irony, hyperbole, and the like, utterer’s occasion meaning just *is* what is 
conversationally implicated. Consider one of Grice’s own examples. He explains that a 
speaker who says ironically ‘*A* is a fine friend’ implicates that *A* is *not* a fine friend. 
Likewise, we might describe the situation in terms of the *Gricean analysis of* 

\(^{16}\) Again, this brings to light the similarities between Grice’s notion of what someone *says* and 
our notion of *strict literal meaning*. 
understanding: the speaker utters ‘A is a fine friend’ with the intention that the audience recognise that the speaker intends the audience to believe that the speaker believes that A is not a fine friend. Put simply, ‘A is a fine friend’ (utterance-type) means that A is a fine friend whereas U (utterer) meant that A is not a fine friend. One could construct a similar account for metaphor, hyperbole, meiosis, and the like. As we have already noted, in the case of figurative language, what the utterer means and what her utterance means are necessarily distinct.

That conversational implicature and utterer’s occasion meaning coincide is not, however, true of all utterances. Consider, for example, Grice’s example of the following exchange between A, who is standing beside his obviously immobilised car, and B, who approaches him:

A: ‘I am out of petrol’
B: ‘There is a garage around the corner’

Though we can agree with Grice that B implicates that the garage is open and has petrol for sale, we would not say that this is what B means by his utterance.

Problem: Indeterminacy of Utterer’s Occasion Meaning and Figurative Language

The coincidence of utterer’s occasion meaning and implicature in the case of figurative language is problematic for Grice’s analysis of utterance-type meaning. Recall that according to Grice an utterance-type has its meaning as a result of being used regularly by various members of a linguistic community to convey the same intention—that is, utterance-type meaning is conventionalised utterer’s occasion meaning. It is not always the case, however, that when a speaker issues an utterance he does so with a determinate
intention as to how he is to be understood. The truism ‘War is war’ and Shakespeare’s metaphor ‘Juliet is the sun’ both generate an indefinite number of implicatures. In the context of a discussion of the distinction between what someone ‘says’ and what she ‘implies’, this kind of indeterminacy is unproblematic. Indeed, Grice acknowledges that sometimes we are led to infer not some particular implication from an utterance, but to explore possible implications:

Since, to calculate a conversational implicature is to calculate what has to be supposed in order to preserve the supposition that the Cooperative Principle is being observed, and since there may be various possible specific explanations, a list of which may be open, the conversational implicatum in such cases will be [a] disjunction of such specific explanations; and if the list of these is open, the implicatum will have just the kind of indeterminacy that many implicata do in fact seem to possess. (“Logic and Conversation” 39-40, insert mine)

This is particularly true in some instances of poetry where the author intends the audience to explore a multitude of possible implications.

Now consider this case in the context of the relation between utterer’s occasion meaning and utterance-type meaning. Again, an utterance-type acquires meaning in virtue of being used regularly within a linguistic community with the same utterer’s occasion meaning. In the case of a metaphor, utterer’s occasion meaning and conversational implicature coincide. But, as we noted above, Shakespeare’s metaphor ‘Juliet is the sun’ generates an indefinite number of implicatures. According to the Gricean analysis of understanding, to understand how we interpret Romeo’s (Shakespeare’s) utterance, Grice needs to tell a story something like this: Romeo utters ‘Juliet is the sun’ intending us to believe (or at least believe that he believes that) $p$ in part by our recognition of his intention that we should believe (or at least believe that he
believes) that \( p \). But what do we substitute for \( p \)? That Romeo’s day begins and ends with Juliet? That she is the source of his life? That his world revolves around her? As with all potent metaphors, there is no determinate answer—indeed, a carefully constructed metaphor gets its value in part from its lack of a determinate reading. It would appear that there is no definite utterer’s occasion meaning associated with Shakespeare’s metaphor. How, then, does the utterance-type ‘Juliet is the sun’ acquire its meaning? \(^{17}\)

The Deeper Problem: Interpreting Figurative Language on the Gricean Model

Aside from the issue of the indeterminacy of utterer’s occasion meaning in some instances of figurative uses of language, there is a deeper problem associated with Grice’s definition of ‘say’ and its relation to utterance-type meaning. The problem arises from Grice’s focus on cases in which utterance-type meaning and utterer’s occasion meaning coincide—i.e., the case of literal utterance. Indeed, Grice’s definition of his favoured sense of ‘say’ draws explicitly and exclusively on this type of case.

By his definition of ‘say’ Grice appears to equate saying with asserting. There is evidence of this elsewhere in his writings. In “Further Notes on Logic and Conversation” he writes, “On my account, it will not be true that when I say that \( p \), I

\(^{17}\) This signals that, at the very least, Grice needs to supplement his account with an account of how the meanings of words contribute to the meanings of sentences. Indeed, Grice’s failure to provide an adequate analysis of the meaning of subsentential components (he gives an inadequate sketch in “Utterer’s Meaning, Sentence-Meaning, and Word-Meaning” (235f.)) is an oft cited complaint about his overall approach (see, for example, Schiffer, Meaning ch. 6 and Loar, Mind and Meaning 253f.). Without such an analysis, Grice cannot account for the meanings of sentences that are never uttered (or, given his commitment to conventionality, even for the meanings of sentences that are rarely uttered).
conversationally implicate that I believe that $p$ . . . the natural thing to say is that [I] expressed (or at least purported to express) the belief that $p$ ” (42, insert mine). And, in the Preface to *Studies in the Way of Words*, Grice writes that his concern lies with “two closely linked ideas... which may be loosely characterised as that of assertion and implication and that of meaning” (emphasis mine). In the case of literal language, this apparent equation of saying and asserting is unproblematic. Consider, for example, an utterance of ‘The leaves on the maple tree are turning red’ in an uncontroversial context in which the speaker intends to convey that she believes that the leaves on some contextually specified maple tree are turning red. In this case, utterance-type meaning and utterer’s occasion meaning coincide. Indeed, we can say that the speaker has *asserted* that the leaves on the maple tree are turning red. But now consider the case of figurative language—that is, the case in which utterance-type meaning and utterer’s occasion meaning diverge. Suppose that of three friends who plan to meet at a pub after work, only two arrive on time. One asks the other about the whereabouts of the third friend who is an accountant. The second responds, ‘He is still chewing numbers’. Though his *utterance* means that the contextually specified friend is still chewing numbers (literally), *he* (the speaker) means that the contextually specified friend is still at work (or something like this). Clearly, he has not asserted ‘He is still chewing numbers’. But has he not *said* ‘He is still chewing numbers’? Grice, in places, writes as though this is how he means to be understood. Indeed, recall that he says that what someone says is “closely related to the conventional meaning of the words (the sentence)... uttered” (“Logic and Conversation” 66). Further, in his own account of the figurative devices
irony, metaphor and meiosis he talks this way:

X, with whom till now A has been on close terms, has betrayed a secret of A's to a business rival. A and his audience both know this. A says 'X is a fine friend'...
I say 'You are the cream in my coffee', intending the hearer to reach first the metaphor interpretant 'You are my pride and joy'...
Of a man to have broken up all the furniture, one says 'He was a little intoxicated'. ("Logic and Conversation" 71, emphases mine)

So, although what someone says should, it would appear, be connected to the notion of utterance-type meaning, it is not clear that it should, or can, be tied to the notion of utterer's occasion meaning. Saying, like asserting, is a speech act, but more basic.

Indeed, it would seem that saying is the foundation of all speech acts. One cannot assert, promise, lie, ask a question, issue a command, or the like, without saying something—that is, without making a meaningful utterance.18

Why, then, in his definition of 'say', would Grice connect saying not only with the central meaning of an utterance, but also with utterer's occasion meaning? Given his broader analysis, he has good reason. Since, according to Grice, utterance-type meaning is derivative of utterer's occasion meaning—that is, because utterance-type meaning is conventionalized utterer's occasion meaning—he must suppose that utterances of a type will be made, as a general rule, with the same basic Gricean intentions. Because he focuses exclusively on literal utterances—that is, on utterances in which the principal intention of the speaker is to communicate what her utterance means—this appears to him to be unproblematic. But, as Grice is well aware given his work on implicature, literal

18 This is not quite right. In a particular context one might, for example, issue a gesture that serves to make an assertion (such as pointing or nodding). Such acts, however, are not speech acts.
utterances are but a subset of actual cases. In all cases of figurative language, for example, utterer’s occasion meaning and utterance-type meaning diverge—indeed, this is a necessary condition of a figurative device. More generally, any sentence, from the most straightforward to the most fantastic, might be employed for any number of purposes: to make an assertion, to issue a promise, to tell a lie, to make a metaphor, and so forth. Thus the lesson is this: utterance-type meaning cannot be determined by the uses to which that utterance-type is put.

So, aside from the issue of the indeterminacy of utterer’s occasion meaning in some instances of figurative uses of language, a more fundamental issue has surfaced: the Gricean analysis of understanding cannot explain adequately how figurative utterances are interpreted. In the case of figurative language, the intention with which the utterer issues her utterance cannot be determined unless the audience already knows the meaning of the utterance. One cannot, for example, begin to know what Shakespeare’s Romeo means by ‘Juliet is the sun’ without first understanding what ‘Juliet is the sun’ means (literally). Figurative interpretations, as a rule, are parasitic on utterance meaning. Indeed, it is questionable that in general one can even know that an utterance is to be understood figuratively without foreknowledge of the meaning of the utterance. But, according to Grice’s programme, an account of utterer’s occasion meaning (non-natural meaning) is primary, and an account of utterance-type meaning is derived from this primary account. Grice’s dilemma, then, is this: what a speaker S means by issuing the utterance X cannot

---

19 Though there may be cases, for example, when one knows by the tone of voice that an utterance is intended ironically, though one does not understand what the utterance means.
be determined, even partially, by the meaning of \( X \); for if it were, then this would render circular Grice’s account of what \( X \) means in terms of what \( S \) means by uttering \( X \).\(^{20}\) But Grice slips into this circularity. In a number of places he says openly that knowledge of utterance-type meaning is necessary to determine utterer’s occasion meaning. Consider this remark about conversational implicature:

\[
\text{[I]n order that a nonconventional implicature should be present in a given case, my account requires that a speaker shall be able to utilize the conventional meaning of a sentence. (“Further Notes on Logic and Conversation” 49)}
\]

An interpreter cannot infer an implicatum without knowledge of the meaning of the speaker’s utterance. In the case of figurative language, because they coincide, anything that is required to determine what is implicated by an utterance will also be required to determine utterer’s occasion meaning. Thus effectively Grice admits that in the case of figurative language one cannot work out utterer’s occasion meaning without foreknowledge of utterance-type meaning. Further, in “Utterer's Meaning and Intentions” Grice writes, “[W]hen the vehicle of meaning is a sentence (or the utterance of a sentence) the speaker’s intentions are to be recognised, in the normal case, by virtue of a knowledge of the conventional use of the sentence” (160-161).\(^{21}\) Indeed, at the end of his discussion of the H-W gesture, Grice writes that he intends to cover “both the case in which U meant by H-W that he knew the route (coincidence of meaning ‘…’ and meaning that …), and also for the case in which, for example, U (a criminal) has lured a


\(^{21}\) What Grice means by the “normal case” is unclear (statistically most regular? literal?). In any case, the admission is damaging to his account of utterance type meaning.
victim into his car and signals (non-literally, so to speak) to his accomplice that he knows how to handle the victim” ("Utterer’s Meaning, Sentence-Meaning, and Word-Meaning" 235). In both cases, Grice adds, the audience’s understanding of the utterance H-W “will be based on its knowledge that U has a certain procedure (to utter H-W if U wants an audience to think that U thinks that U knows the route)” (235).

What, then, of the supposed power of the Gricean analysis of understanding to explain our understanding of non-conventional utterances? And what of Grice’s claim that to be understood an utterance “need not be a linguistic or even a conventionalised performance” ("Utterer’s Meaning, Sentence-Meaning, and Word-Meaning" 226)?

**Summing Up**

In our quest for an adequate analysis of the distinction between literal and non-literal language, Grice’s programme provides us with some important ingredients, and one important lesson. The Gricean analysis of understanding provides a persuasive basic description of what actually occurs during the interpretive process. Further, his account treats individual acts of communication as the source of meaning, provides a critical role for context, and allows us to situate linguistic interpretation in the broader context of interpersonal communication. Moreover, he provides new tools for understanding the distinction between literal and non-literal language in the form of his concepts of saying and implicating. But his analysis falters on one crucial point: he tries to construct an account of literal sentential meaning based on speaker meaning, failing to recognise that the same sentence can be used to perform an indefinite number of different kinds of
speech acts. Ideally, then, what we need is an account of *sentence* meaning that makes use of something akin to the *Gricean analysis of understanding*. The substance of a successful continuance of Grice's work—that is, one that embraces its virtues and rejects its vices—is found in the works of Donald Davidson, to which we now turn.
The distinction between sentence meaning and speaker meaning is implied directly by the fact that the same sentence with the same meaning can be used for a variety of different ends—to make an assertion, to tell a story, to issue an invitation, to create a metaphor, and so forth. The bounds of use, it would seem, are constrained only by the creativity of speakers and interpreters. It is essential, then, to distinguish between what our utterances mean and what we use them to accomplish. Davidson contends, “[I]t is not an accidental feature of language that the ulterior purposes of an utterance [i.e., uses to which the speaker puts the utterance] and its literal meaning are independent, in the sense that the latter cannot be derived from the former: it is of the essence of language” (“Communication and Convention” 274, insert and emphasis mine).¹ Davidson calls this feature of language the principle of the autonomy of meaning, and he credits Grice with bringing it to our attention (“A Nice Derangement of Epitaphs” 437).

Despite the impossibility of generating the meanings of sentences from what speakers mean by uttering them, there is a widespread belief that something like the

---

¹ Indeed, Davidson observes that the use to which a sentence is put is not even constrained by its grammatical mood. In his paper “Moods and Performances” he notes that although prima facie we might suppose that there is a correlation between the moods of sentences (indicative, imperative, optative, interrogative) and the uses to which we put those sentences (to make assertions, to give orders, to express wishes, to ask questions), this is not the case. An indicative might be used to ask a question (‘I’d like to know where you bought that watch’), as might an imperative (‘Tell me who gave you that’). Similarly, interrogatives may be used to make assertions (‘Did you notice that the neighbours are building a new fence?’). Davidson adds that the response that indicatives are normally used to make assertions is “dubious” if normally means statistically more frequent (“Moods and Performances” 111).
Gricean analysis of understanding explains what is essential to interpretation. The present objective, then, is to give an account of sentence meaning that makes use of something like the Gricean analysis of understanding—that is, an account of sentence meaning that retains a central role for speaker’s intentions. Such an account would coincide with our suppositions, first, that meaning is the result of individual acts of communication rather than the precursor to communication, second, that interpretation is context dependent and, third, that linguistic communication is but one element of the broader phenomenon of interpersonal communication. And, importantly, such an account would provide us with a non-circular analysis of literal meaning. We find the substance of such an account in the works of Davidson.

On the Apparent Regularity of Language Use

Both the Gricean analysis of literal utterance-type meaning and the orthodox position are motivated by the apparent regularity of language use within particular linguistic communities. Generally, it might be thought, we understand one another because we use words in the same way—that is, we adhere to the same linguistic rules or conventions. Indeed, the supposition that language use is regular or conventional is what gives proponents of the orthodox position a foothold for their claims about the existence of context-independent meanings; and, as we have found, it is what Griceans appeal to in order to transform his account of non-natural meaning into an account of utterance-type meaning. However, when attention is turned from sterile academic discourse to unchecked everyday discourse, we find far less regularity than is first supposed.
Davidson draws our attention to the empirical fact that in addition to the proliferation of figurative devices, ordinary conversations are teeming with incomplete or otherwise ungrammatical sentences, malaprops, slips of the tongue, idioms, and other so-called ‘misuses’ of language (“A Nice Derangement of Epitaphs” 437). Nevertheless, typically, understanding does not break down. Indeed, he contends,

Using a word in a non-standard way out of ignorance may be a faux pas, but it is a faux pas in the same way that using the wrong fork at a dinner party is, and it has as little to do with communication as using the wrong fork has to do with nourishing oneself, given that the ‘wrong’ word is understood and the fork works. (“The Social Aspect of Language” 5)

Davidson thinks that this observation can teach us something about what is essential to successful communication.

_Mrs. Malaprop and Conventional Meaning_

Davidson shares our supposition that language is essentially an instrument of communication and, so, that literal meaning must be explained within the broader context of the phenomenon of interpersonal communication. He wants to erase the boundary “between knowing a language and knowing our way around in the world generally” (“A Nice Derangement of Epitaphs” 445-446). Thus the starting point of his quest for an understanding of literal meaning lies not with an appeal to abstract entities such as ‘languages’ or ‘meanings’. Indeed, he observes, “[W]e all talk so freely about language or languages that we tend to forget that there are no such things in the world; there are only speakers and their various written and acoustical products” (“The Second Person” 256). He thinks that this point, though obvious in itself, “is nevertheless easy to forget,
and it has consequences that are far from universally recognized” (256).

For instance, it is commonly supposed and, indeed, consistent with the orthodox position, that one knows what an utterance means literally just in case one knows what it means conventionally. In “A Nice Derangement of Epitaphs” Davidson’s objective is to show that this supposition is mistaken—that is, that literal meaning cannot properly be identified with conventional meaning. He constructs an argument based on the example of Mrs. Malaprop’s utterance ‘This is a nice derangement of epitaphs’. In summary, the argument is this.² Although Mrs. Malaprop does not use the words ‘derangement’ and ‘epitaphs’ according to the conventions of her linguistic community, we have no trouble understanding her utterance. In what, then, does our understanding consist? Davidson argues that since an interpreter could assign to her utterance its conventional meaning—that is, that the demonstrated item is a nice derangement of epitaphs—and still fail to understand her, understanding cannot consist in this knowledge. But then, it follows, if we assume that linguistic communication succeeds when and only when the interpreter grasps the literal meaning of the speaker’s utterance, literal meaning cannot be identified with conventional meaning.

But if knowledge of conventional meaning does not play an essential role in our understanding of Mrs. Malaprop’s utterance, what does—that is, how is it that we understand her? One may be tempted to say that we understand Mrs. Malaprop just because we recognise that she intends by her utterance that we should believe that she

² My outline of Davidson’s argument owes much to C.J.L. Talmage’s excellent summary in “Literal Meaning, Conventional Meaning and First Meaning” (214).
believes that the demonstrated item is a nice arrangement of epithets. But this is Grice’s explanation, and so it will not work—that is, it conflates the distinction between what Mrs. Malaprop’s utterance means and what she intends to accomplish as a result of uttering it. What this answer identifies, Catherine J.L. Talmage explains, “is not what is grasped in grasping the literal meaning of this utterance, but rather what is grasped in grasping a certain purpose Mrs. Malaprop had in making it” (“Literal Meaning, Conventional Meaning and First Meaning” 215).

Though the Gricean explanation does not work, Davidson thinks that there is something to be learned from Grice’s appeal to speaker’s intentions. Indeed, Davidson wishes “to emphasise, following Grice, the central role of intention in communication” (“The Second Person” 258). His objective, then, is to construct an account of literal meaning that permits a central role for speaker’s intentions, thus retaining the essential link between meaning and individual acts of communication.

*First Meaning*

Davidson uses the example of malaprops to argue that to preserve the distinction between sentence meaning and speaker meaning, we must pry apart what is literal in language from what is conventional. In place of the notion of *literal meaning*, which he contends “is too incrusted with philosophical and other extras to do much work,” he introduces the notion of *first meaning* (“A Nice Derangement of Epitaphs” 434). Distinct from conventional meaning, first meaning applies to words and sentences as uttered by a particular speaker on a particular occasion. The *first meaning* of any utterance is
specified by what Davidson calls the speaker’s *first intention*. He offers the following example as a general explanation of what he means by *first intention*:

Suppose Diogenes utters the words ‘I would have you stand from between me and the sun’ (or their Greek equivalent) with the intention of uttering words that will be interpreted by Alexander as true if and only if Diogenes would have him stand from between Diogenes and the sun, and this with the intention of asking Alexander to move from between him and the sun, and this with the intention of leaving a good anecdote to posterity... In general, the first intention in the sequence to require this feature specifies the first meaning. (“A Nice Derangement of Epitaphs” 435)

In general, then, according to Davidson, the first meaning of an utterance is specified by the speaker’s first intention, i.e., the intention to utter words that will be interpreted in a certain way, as having certain truth conditions. Regardless of the use to which a speaker puts her utterance (e.g., to make a promise, to tell a lie, to issue a command), if communication succeeds, her first intention must necessarily be grasped by the interpreter.\(^3\) Roughly, then, first meaning comes first in the order of interpretation. In the case of Mrs. Malaprop’s utterance, for example, linguistic communication succeeds when and only when the interpreter realises that she intends her words to be interpreted as true if and only if the demonstrated item is a nice arrangement of epithets.\(^4\) Since first intentions specify first meaning, her utterance means that the demonstrated item is a

---

\(^3\) Thus, following Quine, Davidson disposes of any fundamental role for abstract ‘meanings’. To say that a word or sentence has a meaning is not to posit some separate entity, but rather it is like saying that a building has a certain height or a stone a certain weight. Thus we might say that understanding an utterance consists not in attaching to it the correct sense or meaning, but in knowing the conditions under which it would be true.

\(^4\) Davidson adds that this characterisation can be improved by admitting the Gricean condition that the speaker intends the interpreter to arrive at the right truth conditions through the interpreter’s recognition of the speaker’s intention to be so interpreted (“A Nice Derangement of Epitaphs” 265, n.5).
nice arrangement of epithets.⁵

Like Grice, then, Davidson depends on the notion of speaker’s intention for his account of utterance meaning (i.e., first meaning). Unlike Grice, however, Davidson has in mind a very specific kind of intention, namely, the speaker’s (first) intention that the interpreter assign certain truth conditions to her utterance. And, indeed, this is what allows Davidson to keep distinct the concepts of utterance meaning and speaker meaning. Whatever other intentions a speaker may have, for communication to succeed, speaker and interpreter must assign the same truth conditions to the speaker’s words. Davidson adds that if the grasping of first meaning is the mark of successful linguistic communication, then the literal meaning of an utterance must be its first meaning, not its conventional meaning.

This is not to say that Davidson is committed to the radical view that one can assign to one’s utterance a certain meaning simply by intending that the utterance be

---

⁵ Davidson knows that we do not interpret utterances one at a time, independent of our general and accumulated specific beliefs and expectations about a speaker and her behaviour. Rather, we, if you will, construct theories about the languages that those with whom we communicate speak, and we continually revise these theories in light of new evidence. Thus, in order to describe more perspicuously how his notion of first meaning fits into a general account of interpretation, Davidson introduces the notions of prior and passing theories. For the interpreter, the prior theory expresses how he is prepared in advance to interpret an utterance of the speaker; his passing theory is how he actually interprets the utterance. For the speaker, the prior theory is what he believes the interpreter’s prior theory to be; his passing theory is the theory he intends his interpreter to use. Davidson intends his malaprop argument to show that what must coincide for linguistic communication to succeed are passing, not prior, theories. Mrs. Malaprop utters ‘This is a nice derangement of epitaphs’ expecting the interpreter to apply a certain non-conventional (prior) theory to her utterance and so attribute to it the meaning that is conventionally expressed by the sentence ‘This is a nice arrangement of epithets’. But Mrs. Malaprop is wrong about the interpreter’s prior theory—he is prepared to interpret her utterance according to the conventions of our linguistic community. To understand her utterance, then, the interpreter must instead apply a theory of meaning that assigns to ‘derangement’ and ‘epitaphs’ the semantic roles assigned conventionally to ‘arrangement’ and ‘epithets’. 
understood to have that meaning. Rather, he argues, a speaker must provide an interpreter with enough clues reasonably to expect that the interpreter will interpret his utterance as he intends. Lewis Carroll’s Humpty Dumpty cannot mean ‘There’s a nice knock-down argument for you’ by saying ‘There’s glory for you’ because he knows that there is no way that Alice can interpret it as meaning this (“A Nice Derangement of Epitaphs” 440).\(^6\) Thus, according to Davidson, interpreters constitute a kind of constraint on what a speaker’s words mean literally. Nevertheless, to be clear, Davidson is not committed to the equally radical view that interpreters determine the literal meanings of our utterances. Rather, his account of first meaning allows a role both for the speaker and for the interpreter: an utterance has a first meaning only if the speaker believes the interpreter will, or at least might, interpret it as having this first meaning.

**A Conventionalist Response**

Davidson’s malaprop argument is designed to show that our ability to interpret malaprops, slips of the tongue, ungrammatical sentences, and the like threatens the idea that interpretation depends essentially on advance learning of meanings. But one might protest that although his argument shows that knowledge of conventional meaning is not sufficient to interpret malaprops, it does not show that knowledge of conventional meaning is not necessary to interpret malaprops. Put differently, one might respond to

---

\(^6\) Though, Talmage points out, in “The Structure and Content of Truth” Davidson admits of a case in which “the audience for the purpose of the interpretation is... just the speaker himself,” thus apparently committing himself to the absurdly individualistic view that one can mean what one says one means without regard for whether others have been provided with sufficient clues for successful interpretation (qtd. in Talmage, “Davidson and Humpty Dumpty” 540). This, it would seem, is an error on Davidson’s part.
his argument by claiming that we understand malaprops and the like precisely *because* we recognise that there is a conventional meaning from which the speaker is diverging. One might contend, therefore, that Davidson's account of how we interpret malaprops and their kin is not the only one available.

Davidson is drawn to the case of malaprops because he is concerned that their presence threatens the distinction between what speakers, on given occasions, mean, and what their words mean, since in these cases the intended meaning seems to take over from the standard meaning. Indeed, this is the reason he calls for "a deeper notion of what words, when spoken in context, mean"—that is, his concern about how we interpret malaprops provides the impetus for his account of *first meaning* ("A Nice Derangement of Epitaphs" 434). To be sure, we want to keep utterance meaning and speaker meaning distinct. Why not, though, treat malaprops as a kind of non-literal language? According to this line, in the case of malaprops, as in the case of non-literal language, what the speaker's words mean, and what the speaker means to convey by using those words, are distinct. In both cases, the absurdity of the conventional interpretation of the speaker's words (given the context and our background assumptions) prompts us to attribute to the speaker a meaning different than that which we attribute to her words.7

Thus, one might respond to Davidson that his examples do *not* show that we need to pry apart the notions of conventional meaning and literal meaning, but only that

---

7 The difference between, say, a metaphor and a malaprop would be that in the case of a metaphor the clues to speaker meaning come from the meaning of the utterance (and, of course, the context) whereas in the case of malaprops, the clues come from the similarity in sound of the speaker's words to the words that would conventionally express what he means to convey by his words.
sometimes (perhaps often) speaker meaning and utterance meaning diverge. We can
grant Davidson’s point that since an interpreter could assign to Mrs. Malaprop’s utterance
its conventional meaning and fail to understand her, understanding cannot consist in this
knowledge. But to show that our understanding does not consist in our knowledge of the
conventional meaning of her words is not to show that this knowledge does not play an
essential role in the interpretative process. Indeed, one might retell Davidson’s example
of Mrs. Malaprop’s utterance as follows: we assign to the words in her utterance ‘This is
a nice derangement of epitaphs’ their conventional meanings, but, because of the
absurdity of her utterance given the context and our background assumptions (including
our knowledge of the similarity in sound between ‘epitaph’ and ‘epithet’ and between
‘derangement’ and ‘arrangement’), we understand that she means to convey that the
demonstrated item is a nice arrangement of epithets.⁸

Thus Davidson’s malaprop argument by itself does not prove that shared linguistic
conventions are not necessary to successful linguistic interpretation. Rather, his argument
shows only that we cannot understand a malaprop based solely on our knowledge of what

---

⁸ Interestingly, Davidson himself proposes something like this explanation, though, in keeping
with his argument, he is careful to stop short of saying that knowledge of conventional meaning is
necessary to understanding:

[In the case of malaprop and its kin] the hearer realises that the ‘standard’
interpretation cannot be the intended interpretation; through ignorance,
 inadvertence, or design the speaker has used a word similar in sound to the word
that would have ‘correctly’ expressed his meaning. The absurdity or
inappropriateness of what the speaker would have meant had his words been taken
in the ‘standard’ way alerts the hearer to trickery or error; the similarity in sound
tips him off to the right interpretation. (“A Nice Derangement of Epitaphs” 434,
insert and emphasis mine)

The suggestion currently under consideration is that whereas Davidson thinks that ‘the right
interpretation’ applies to the speaker’s words, one might argue that it applies to what the speaker
means to convey by uttering those words.
the words mean conventionally. To show that shared conventions are not an essential element of the interpretive process, Davidson needs to show that we could understand Mrs. Malaprop's utterance without knowing what her words mean conventionally. And, indeed, he has an argument for this. But this argument is not found in his account of malaprops; rather, it is found in his account of the methodology of radical interpretation.

**Radical Interpretation**

To sharpen our focus on what is and what is not essential to linguistic communication, Davidson invites us to envision a scenario in which we have the task of interpreting someone whose utterances are wholly unfamiliar to us—this he calls the case of radical interpretation. In the case of radical interpretation, the interpreter has nothing to go on but what she sees the native speaker do and the sounds she hears him make. How is it, then, that one goes about interpreting someone in such a case?

To begin, Davidson claims, the interpreter must be able to identify when the speaker holds his sentences to be true ("Radical Interpretation" 135). That is, the interpreter must assume that she is observing a creature that makes assertions, and she must be reasonably good at determining when he is engaged in this particular activity (even when she does not know what is being asserted). Though, to be sure, assertion is not the only case in which the interpreter might detect whether the speaker holds his

---

9 It is worth noting that Davidson intends the methodology of radical interpretation as a theoretical description of linguistic competence, not as a description of an actual procedure. Also, it will be clear that for his account of radical interpretation Davidson borrows heavily from Quine's account of radical translation. There are, however, significant differences. Importantly, for example, Davidson's methodology of radical interpretation does not allow for indeterminacy of logical form (see "Radical Interpretation" 136 n.16).
utterance to be true—"Lies, commands, stories, irony, if they are detected as attitudes, can reveal whether a speaker holds his sentence to be true" ("Radical Interpretation" 135).

Further, Davidson argues, the interpreter must assume that when the native speaker holds his sentences to be true, he does so largely correctly. This is often called the principle of charity, but, as Bjørn Ramberg points out, this term is misleading in so far as it suggests that it is a principle we choose to employ as a kindness toward the native speakers; on the contrary, Davidson takes it to be "an indispensable methodological principle, without which... a theoretical description of our own linguistic competence would founder because of the inseparability of attributing beliefs to speakers and meaning to the sentences spoken" (Ramberg, Donald Davidson's Philosophy of Language 69). Put another way, because we cannot attribute meanings to the native's sentences without also attributing beliefs to the native, the role of the principle of charity is to hold belief constant (as far as possible) while solving for meaning.

Armed with the principle of charity, the interpreter, appreciating as best she can the perspective of the native speaker, notes correlations between sentences held true and the (external) circumstances in which they are uttered. For each utterance of this sort that the native speaker makes, the interpreter tries to isolate the salient features of the context in which the uninterpreted utterance is made—that is, he tries to isolate the features of the environment that cause the speaker to assent to the sentence. The interpreter continually adjusts his theory according to new evidence, evidence in the form of "knowledge of the character, dress, role, sex, of the speaker, and whatever else has
been gained by observing the speaker’s behaviour, linguistic or otherwise” (“A Nice Derangement of Epitaphs” 441).

Using this knowledge, the interpreter constructs hypotheses about what each sentence uttered means. Davidson suggests that these hypotheses might be put in the form of Tarski-like T-sentences, the formulations of which look like this: “‘S’ is true-in-L when uttered by x at time t if and only if p” where S stands for the sentence asserted and p for the truth conditions of the sentence asserted.\(^{10}\) Equipped with a stock of T-sentences, the interpreter then develops more fine-grained hypotheses about correlations between subsentential components (i.e., words and phrases) and external objects or events. Along the way, she tests her hypotheses by interacting with the native speaker in a game of query and assent/dissent without which no language learning could take place\(^{11}\), revising her hypotheses where necessary. As a result, the interpreter acquires increasing confidence that she has in her possession a theory which generates correct interpretations of the native speaker’s sentences, actual and potential:

The longer we interpret a speaker with apparent success as speaking a particular language the greater our legitimate confidence that the speaker is speaking that language, or one much like it. Our strengthening expectations are as well founded as our evidence and induction make them.

---

\(^{10}\) Though we shall not enter into the details here, it is worth noting that Davidson holds that imperatives, optatives and the like (i.e., non-indicatives) can be accounted for by T-sentences (Davidson, “On Saying That”). The perception that they cannot is the result of a failure to appreciate that each non-indicative is in fact an abbreviation of two sentences. The first indicates the mood; the second gives the content. For example, the imperative “Close the door!” is analysed by Davidson as something like this: “The next sentence is an order. You will close the door.” Troublesome indicatives, in particular, those that express indirect discourse (e.g., ‘Blair said that he would attend the conference’) and those that express propositional attitudes (e.g., ‘Chris believes that the market will recover’), are treated similarly.

\(^{11}\) cf. Quine, “Mind and Verbal Dispositions” (88).
155

("The Second Person" 258)

Davidson's objective in telling his story about radical interpretation is to teach us something important about what is necessary to successful communication. The lesson is this: "[T]here is no reason, in theory at least, why speakers who understand each other ever need to speak, or to have spoken, as anyone else speaks, much less as each other speak" ("The Social Aspect of Language" 4). Indeed, given that "[d]ifferent speakers have different stocks of proper names, different vocabularies, and attach somewhat different meanings to words," Davidson contends that something approaching this radically individualistic use of language is actually the case ("Communication and Convention" 277). Accordingly, Davidson intends his description of the methodology of radical interpretation to apply not only to cases in which the speaker's utterances are wholly unfamiliar to us, but also to cases in which we communicate with other members of our linguistic community: "The problem of interpretation is domestic as well as foreign... All understanding of the speech of another involves radical interpretation" ("Radical Interpretation" 125). Though, again, this is a theoretical point, intended to focus our attention on what is essential to successful linguistic communication: "I do not think we normally understand what others say by consciously reflecting on the question what they mean, by appealing to some theory of interpretation, or by summoning up what we take to be the relevant evidence" ("The Social Aspect of Language" 2).12 Indeed,

12 Even when we interpret the clerk at the supermarket, the stranger on the bus or the news anchorperson on television, Davidson would say that we are equipped with a robust prior theory of the language they speak. Indeed, the less we know about a speaker, assuming she belongs to our linguistic community, the more closely our prior theory will simply be the theory that describes the standards (or, more precisely, what we perceive to be the standards) of our linguistic community.
though Davidson argues that every act of interpretation is essentially an act of radical interpretation, he does not deny that linguistic conventions are of immense practical use. He acknowledges, “Knowledge of the conventions of language is... a practical crutch to interpretation, a crutch we cannot in practice afford to do without,” but adds that it is “a crutch which, under optimum conditions for communication, we can in the end throw away, and could in theory have done without from the start” (“Communication and Convention” 279). We speak as others do because we want to be understood, and speaking as others do assists greatly with this goal. But Davidson’s radical interpretation argument is designed to show that what matters to successful communication is not that we speak the same language (i.e., that we share linguistic conventions), but rather that we provide each other with something intelligible as a language. In any given interpretive situation, what must be shared is “the interpreter’s and the speaker’s understanding of the speaker’s words” (“A Nice Derangement of Epitaphs” 438). In theory, this does not necessitate following shared rules or conventions. What is conventional about language, then, is that people tend to speak much as their neighbours do. But, Davidson reiterates,

[I]n indicating this element of the conventional, or of the conditioning process that makes speakers rough linguistic facsimiles of their friends and parents, we explain no more than the convergence; we throw no light on the essential nature of the skills that are thus made to converge. (“Communication and Convention” 278)

Although meaning something requires that by and large one follow a practice of one’s own, a practice that can be understood by others, there is no fundamental reason why practices must be shared.
Thus, Davidson concludes, although successful interpretation depends on a speaker intentionally making herself interpretable to a hearer (she must go on more or less as the interpreter expects, or at least is equipped to interpret), it does not depend on shared conventions. Even when a speaker speaks in accordance with the standards of his linguistic community, "he speaks with the intention of being understood in a certain way, and this intention depends on his beliefs about his audience, in particular how he believes or assumes they will understand him" ("The Social Aspect of Language" 7). According to Davidson, then, the meaning of an utterance is determined not by convention, even when a speaker conforms to convention, but rather by the speaker's intention that her utterance be assigned a certain truth conditions, truth conditions specifiable by the appropriate T-sentence. "[I]f communication succeeds," Davidson contends, "there must be these intentions on the part of the speaker" ("The Second Person" 259).

First Meaning and Radical Interpretation

Does Davidson's methodology of radical interpretation yield theories of meaning that give first meanings? Davidson seems to think so—at least, he gives no indication that he considers first meaning and meaning generated by the methodology of radical interpretation distinct. Talmage is less certain. She argues that there are cases in which the first meaning of an utterance, as specified by the speaker's first intention, and the meaning of that utterance as specified by the theory of meaning generated by the methodology of radical interpretation, are in fact distinct. To argue this, she offers the following example:
[S]uppose that, unlike Mrs. Malaprop, I intend my hearer (whom I believe to be less adept than most in noticing similarities in sound) to assign to the utterance ['This is a nice derangement of epitaphs'] its conventional meaning, that I intend him, that is, to interpret it as being true if and only if this is a nice derangement of epitaphs. (“Literal Meaning, Conventional Meaning and First Meaning” 219-220, insert mine)

She argues that the meaning that would be arrived at through radical interpretation is that the demonstrated item is a nice arrangement of epithets and so does not coincide with the meaning she intends the hearer to assign to her utterance (i.e., that the demonstrated item is a nice derangement of epitaphs).

Actually, the matter is not so simple. Talmage knows that in saying that she intends her hearer to assign to her utterance its conventional meaning, she has not said all that is necessary to specify her utterance’s first meaning. One’s first intention—that is, the intention that one’s utterance be taken to mean what one wants it to be taken to mean—is not the only constraint on first meaning. As we found, Davidson argues that when a speaker speaks with the intention of being understood in a certain way, this intention depends on her beliefs about her audience, in particular how she believes or assumes she will be understood. Indeed, he says plainly,

[F]or me the concept of “the meaning” of a word or sentence gives way to the concepts of how a speaker intends his words to be understood, and of how a hearer understands them. Where understanding matches intent we can, if we please, speak of “the” meaning; but it is understanding that gives life to meaning, not the other way around. (“The Social Aspect of Language” 6, emphasis mine)

In Talmage’s example, however, she cannot reasonably expect that the hearer will interpret her utterance as meaning that the demonstrated item is a nice arrangement of epithets. Indeed, by her own description, she intends her hearer to “misinterpret” her
utterance ("Literal Meaning, Conventional Meaning and First Meaning" 220). But in the
case in which one intends an audience to misinterpret one's utterance, success depends on
the interpreter misidentifying the meaning of the utterance. Because first meaning arises
only when understanding matches intent, we are forced to conclude that Talmage's
utterance is actually devoid of a first meaning. Talmage knows this, and so she
concludes, "[T]he point is just that I neither expected nor wanted him to understand [my
utterance]—my purpose in uttering words that literally mean that this is a nice
arrangement of epithets was not to communicate linguistically with him" ("Literal
Meaning, Conventional Meaning and First Meaning" 220, insert mine). If her utterance
does not have a first meaning, and yet a radical interpreter would interpret it as true if
and only if the demonstrated item is a nice arrangement of epithets, these notions of
meaning, she concludes, must be distinct.¹³

Prima facie, Talmage's argument is persuasive. According to her analysis, on
Davidson's account we are forced to conclude that it does not make sense to say that a
speaker can intend her audience to misinterpret her utterance. To intend that someone
misinterpret one's utterance is to intend that they assign to the utterance a meaning other
than the meaning of the utterance. This is uncontroversial enough. But first meaning,
according to Davidson, is determined by the interpretation that the speaker intends that
the interpreter assign to the utterance. Thus we are led to an absurd conclusion: in the
case of intended misinterpretation, the intended interpretation is necessarily different than

¹³ Talmage's objective is to show that not only should literal meaning not be identified with
conventional meaning, but it should also not be identified with first meaning. Rather, she argues,
the literal meaning of an utterance is the meaning specified by the theory of meaning arrived at
through the methodology of radical interpretation.
the meaning that the speaker intends that the interpreter assign to her utterance.

But surely there are cases of intended misinterpretation. My lawyer might draft a document which she wants me to believe, were I to sign it, would mean that I was to receive ninety percent of a monetary settlement, and her, for her work, the other ten percent. Unknown to me, however, the document actually dictates that she is to receive ninety percent of the settlement, and I am to receive ten percent. Her hope, then, is that because of my lack of legal acumen, I will misinterpret the document and sign it. That is, it would appear that this is a case of intended misinterpretation. Given that the notion of intended misinterpretation makes pre-theoretic sense, it is worth delving further into Davidson’s notion of first meaning to determine whether we can incorporate such cases into his account in such a way as to avoid rendering them nonsensical.

This much seems indisputable about the case Talmage describes: the literal meaning that would be arrived at through the process of radical interpretation and what she means to convey to her interpreter by her utterance are distinct. What is not clear, however, is that in describing that she intends this particular interpreter to assign to her utterance its conventional meaning she has in fact identified her first intention and, so, begun to specify the first meaning of her utterance. To see this more clearly, consider a general analysis of intended misinterpretation. A speaker S intends an interpreter H to misinterpret her utterance X if and only if she utters X, which means $M_1$, with the intention that H instead interpret her utterance as meaning something else, say, $M_2$. According to this definition, then, the intention that X be taken to mean $M_2$ or, more broadly, the intention that one’s utterance be misinterpreted, is a higher level
intention—it is something we do with our words. This suggests that in Talmage’s example her intention that the hearer interpret her utterance as being true if and only if the demonstrated item is a nice derangement of epitaphs is not her first intention and, so, does not begin to specify her utterance’s first meaning. Rather, it suggests, her first intention is that her utterance be interpreted as being true if and only if the demonstrated item is a nice arrangement of epithets, and that she has the further intention that this first intention pass unrecognised by her interpreter, and instead that he interpret the utterance as true if and only if the demonstrated item is a nice derangement of epitaphs. Indeed, the intention that her first intention pass unrecognised just is the intention to be misinterpreted—to be sure, her intention to be misinterpreted will succeed if and only if the interpreter misidentifies her first intention.

But things are not so simple. One might retort that because, according to Davidson, meaning is specified by a speaker’s first intention, our above definitionunpacks like this: a speaker S intends an interpreter H to misinterpret her utterance X if and only if she utters X with the intention that H interpret X as true if and only if truth conditions TC₁ obtain, with the further intention that H instead interpret X as true if and only if TC₂. And this is nonsense—one cannot intend that one’s audience interpret one’s utterance as true if and only if TC₁ and instead interpret it as true if and only if TC₂ where TC₁ and TC₂ are ex hypothesi different.

In order to preserve the supposition that the intention that one’s utterance be misinterpreted is a higher level intention, then, we need an alternative account of how first meaning is specified in Talmage’s example. In particular, somehow we need to keep
separate what we want to call her first intention—that is, the intention that her utterance be taken as true if and only if the demonstrated item is a nice arrangement of epithets—and her intention that her utterance be interpreted by this particular interpreter as true if and only if the demonstrated item is a nice derangement of epitaphs. Generally, first meaning is specified by the speaker’s intention that her utterance be interpreted as she intends, provided that she equips the interpreter with enough clues to suppose that he will, or at least might, interpret it this way. In Talmage’s example, as we found, because the interpreter is “less adept than most in noticing similarities in sound,” and because Talmage knows this, she cannot be said to have provided him with enough clues to reasonably expect that he will interpret her utterance as being true if and only if the demonstrated item is a nice arrangement of epithets. Indeed, this is why we felt forced to conclude that her utterance does not, after all, have a first meaning. But critical to Talmage’s story is that it is this particular interpreter who cannot be expected to interpret her utterance as true if and only if the demonstrated item is a nice arrangement of epithets. If we consider a broader audience—specifically, one which includes a radical interpreter—we can say with confidence that she has in fact provided enough clues to reasonably expect that an interpreter—again, specifically, a radical interpreter—will interpret her utterance as true if and only if the demonstrated item is a nice arrangement of epithets. Why the confidence? Because an integral element of Talmage’s example is that a radical interpreter would understand her utterance this way—indeed, this is what gives content to her claim that first meaning and meaning generated via the methodology of radical interpretation do not always coincide.
Consider again the example in which my lawyer intends me to misinterpret our contract. Suppose I sign the contract. My lawyer then explains in considerably plainer language what the contract means. I protest and, feigning ignorance, she claims that I must have misinterpreted the document. But her claim that I have misinterpreted the document really matters only when a judge rules in her favour—that is, when he decides that the contract in fact means that she is to receive the bulk of the settlement. In Talmage's example, we might employ the mythical radical interpreter in the role that the judge plays in our contract example. That is, the radical interpreter can provide the yardstick against which we measure whether a speaker has provided her audience with enough clues to reasonably expect that her utterance will be interpreted as she intends. Thus we might adjust our characterisation of first meaning to reflect this: the first meaning of an utterance is specified by the speaker's first intention—that is, the intention that her utterance be understood a certain way—provided that she equips her audience with enough clues so that she might reasonably expect that a radical interpreter would interpret it as she intends.

According to the Davidson Talmage describes, first meaning arises only when the particular interpreter with whom the speaker is interacting identifies correctly the speaker's first intention. Certainly there is support for this interpretation in Davidson's writings. According to our revised understanding of first meaning, however, that a

---

14 Recall the quotation: 
[For me the concept of "the meaning" of a word or sentence gives way to the concepts of how a speaker intends his words to be understood, and of how a hearer understands them. Where understanding matches intent we can, if we please, speak of "the" meaning; but it is understanding that gives life to meaning, not the other way around. ("The Social Aspect of Language" 6)
particular interpreter misinterprets or does not understand a speaker’s utterance does not preclude the utterance from having a first meaning. What is critical, in addition to the speaker’s first intention, is that she equip the interpreter with enough clues to interpret her utterance as she intends—that is, that she be co-operative to the extent that a radical interpreter could interpret her utterance as she intends. Indeed, just as there are places where Davidson seems to attribute to particular interpreters the role of constraining first meaning there are also places that support our revised interpretation. For example, Davidson writes, if communication succeeds, “the speaker must intend the hearer to interpret his words in the way the speaker intends, and he must have adequate reason to believe that the hearer will succeed in interpreting him as he intends” (“Communication and Convention” 277). Here Davidson attaches importance to the reasonableness of the speaker’s expectation that she be understood as she intends, but not to whether the particular interpreter actually interprets the utterance as the speaker intends. If our revised analysis of first meaning is granted, then the notions of first meaning and the meaning generated from the methodology of radical interpretation are not in fact distinct.

Note also that according to our revised interpretation Davidson would have an answer to Dummett’s example of someone who issues a series of insults in English to an Italian police officer, not expecting to be understood (“A Nice Derangement of Epitaphs: Some Comments on Davidson and Hacking” 472). Dummett argues that the utterances clearly have meaning according to the conventions of English, but that we are forced on Davidson’s account to conclude that they are meaningless because the speaker does not want to be understood. On our revised account of first meaning, however, Dummett’s
speaker's words do have a first meaning because, though he is not being cooperative with the particular interpreter to whom his remarks are directed, he has provided enough clues that a radical interpreter would interpret his words as he intends them to be understood.

Non-Literal Language and Radical Interpretation

What of the case non-literal utterances and radical interpretation? Again, though Davidson applies the term radical interpretation to every act of interpretation, it will sharpen our focus if we consider the case in which a speaker's utterances are wholly unfamiliar to us. In such a scenario, there are, it seems, two barriers to the interpretation of non-literal utterances. First, recall that critical to the success of radical interpretation is that the interpreter be able to identify successfully when the speaker holds a sentence to be true, even if the content of the sentence is unknown. The identification of the attitude of holding a sentence to be true is critical to the success of radical interpretation. But many non-literal utterances, in particular many figurative utterances, are held by a speaker to be false on their literal interpretation (‘Sampras never loses!’, ‘My car is my life’, ‘The hills are alive with the sound of music’). Further, it is not clear that in the case in which a speaker's utterances are wholly unfamiliar to the interpreter, the interpreter will be able to identify when a speaker is employing a non-literal device.15

15 The exception being the case in which a non-literal attitude is detected by the interpreter, though he does not know the content of the utterance. It was suggested previously, for example, that in some instances one might detect an ironical tone in a speaker's voice though one does not know what the speaker's utterance means. In such a case, one might override the principle of charity and suppose the speaker to hold his utterance to be false.
In any case, there is a more serious issue. Recall that one cannot interpret a non-literal utterance without first knowing what the utterance means literally because any non-literal effects (i.e., conversational implicatures) are parasitic on the literal meaning of the utterance. In general, then, one cannot begin to interpret a speaker’s non-literal utterances unless one already knows much about the speaker’s language. Thus non-literal utterances are unhelpful at best and disruptive at worst to the process of radical interpretation.16

Davidson on the Distinction Between Literal and Non-literal

Thus far outlined, Davidson’s account suits us well. The notion of first meaning provides us with a clear domain of meaning. Each sentence has one and only one meaning, as specified by its author’s first intention (provided, of course, that he is being cooperative). According to Davidson, all ulterior purposes to which the sentence might be put are relegated to the realm of use. In Gricean terms, what a speaker says is identified with the meaning of her utterance whereas what she implicates by her utterance is confined to the realm of use. All utterances, literal or non-literal, mean what the words mean, and nothing else—figurative language and other implicatures are parasitic on first meaning. Accordingly, strictly speaking there is no such thing as, say, metaphorical meaning.17 Tennyson’s metaphor ‘Sleep is the brother of death’ means just that sleep is the brother of death (which is absurd), though we can add that Tennyson uttered this sentence with

16 Indeed, a native speaker who wishes to avoid being interpreted might succeed if she utters exclusively non-literal utterances.

17 cf. Davidson, “What Metaphors Mean.”
the (higher level) intention that we should notice the similarities and differences between
the physical states of sleep and death. According to Davidson, then, non-literal devices
such as metaphor, irony, hyperbole, and the like are relegated to the domain of use.
They are, like issuing a promise, telling a story or making a joke, things which we do
with language.

Davidson and Non-Verbal Communication

Having introduced his account of first meaning, Davidson writes, “Nothing said so far
limits first meaning to language” (“A Nice Derangement of Epitaphs” 436). Indeed,
according to our analysis thus far, it would seem that there is no principled reason why
the notion of first meaning could not be applied to cases of non-verbal communication.
As we learned from Grice, if a gesture, a non-verbal vocalisation, a drawing, or the
like succeeds in conveying a particular message, it is because we attribute to the author
of the gesture or sound or drawing the intention that we should understand it in the way
the author intends. Indeed, non-verbal acts of communication (or, rather, attempts at
communication) are generated by the same non-linguistic intentions as verbal utterances.
One might suppose, then, that we can assign first meaning to non-verbal utterances.
Thus, just as my verbal utterance has a first meaning because I intend you to interpret it
as having certain truth conditions, and you do, so might a wave of my arms, a groan or

18 I am thinking here of cases like ‘brrr’ (to express one’s uncomfortable chill), ‘ummm’ (to
mark a pause in one’s thought process), ‘hmmm’ (to express doubt), and the like. I am not
thinking of vocalisations whose purpose is not to communicate something—for example,
involuntary vocalisations, such as when one is startled, or noises such as those one makes to
startle someone.
an arrow inscribed on a poster.\textsuperscript{19} Indeed, treating non-verbal language this way appears to be consistent with the role Davidson attributes to the concept of a language. He writes, “The point of the concept of a language... is to enable us to give a coherent description of the behaviour of speakers, and of what speakers and interpreters know that allows them to communicate” (“The Second Person” 256-57). If this broad conception of a language is in fact at the core of Davidson’s approach, then, it would appear that non-verbal language should be incorporated into his account. That is, because non-verbal language is integral to interpersonal communication, one might argue that “a coherent description of the behaviour of speakers, and of what speakers and interpreters know that allows them to communicate” must include an account of non-verbal language. To be sure, this would come with certain advantages. To begin, incorporating non-verbal language into an account of first meaning would obviate the difficult (perhaps impossible) task of demarcating clearly between those noises and inscriptions which qualify as verbal and those which do not. Further, it would meet our supposition that verbal language is but one element of the broader phenomenon of interpersonal communication.

Nevertheless, it is clear that Davidson wants to keep separate verbal and non-verbal language. Verbal language is unique, he thinks, because it is utterly systematic:

\begin{quote}
[I]n the case of a language the hearer shares a complex system-or theory with a speaker, a system which makes possible the articulation of logical relations between utterances, and explains the ability to interpret novel utterances in an organised way. (“A Nice Derangement of Epitaphs” 436).
\end{quote}

\textsuperscript{19} Moreover, treating non-verbal language this way frees us to speak of non-literal uses of non-verbal language, as in the case when someone issues a ‘thumbs up’ sign ironically during a fierce rain storm.
Davidson argues that a competent speaker or interpreter of a language is able to create and interpret sentences that she has never before encountered, and that "[t]his is possible because [she] can learn the semantic role of each of a finite number of words and phrases and can learn the semantic consequences of a finite number of modes of composition" ("A Nice Derangement of Epitaphs" 437, insert mine). Indeed, he argues that we cannot describe the totality of the knowledge possessed by the speaker or interpreter of a language—that is, their knowledge of the truth conditions of an indefinitely large number of sentences—without employing a recursive theory ("The Second Person" 257). As we have found, he thinks that one model for such a machine is a theory of truth along the lines of a Tarski truth definition. The linguistic ability of a speaker is specifiable by the T-theory (i.e., collection of T-sentences and their entailment relations) which applies to the totality of utterances (actual and possible) of the language that the speaker is speaking at a particular time.\(^{20}\)

One might retort that we can construct T-sentences for our drawings, non-verbal vocalisations, gestures, and the like—for example, 'A 'thumbs-up' gesture is true for you at time \(t\) if and only if things are okay' (or something like this). But this would be to misconstrue the way that Davidson says T-sentences are generated. T-sentences are not generated one at a time. Indeed, though we may have inductive evidence for a particular T-sentence as it applies to a particular utterance, this is evidence for the meaning of that utterance only, and precisely in so far as, it is evidence for a truth-theory which allows

\(^{20}\) Davidson does not think that a T-theory describes a kind of propositional knowledge possessed by the speaker of the language; rather, a T-theory provides us with an adequate description of the linguistic abilities of the speaker, including the ability to create novel sentences ("A Nice Derangement of Epitaphs" 438).
us to derive the relevant T-sentence as a theorem. That is, evidence for a T-sentence is evidence for the meaning of the utterance only in conjunction with evidence for other T-sentences which, taken together, allow us to formulate a truth-theory. Thus, the evidence is evidence for the meaning of an utterance in so far as it permits us to situate each utterance in the structure that relates each semantically to one another via the principle of compositionality. Put another way, the meaning of a sentence just is this structural location in the language of which it is a part. Therefore, according to Davidson, inductively determining the T-sentence that gives the truth conditions of an utterance, and theoretically determining the meaning an utterance has, are distinct. The former involves empirical observation and developing theories about the speaker's beliefs, both of which require attending to the context of the utterance. The latter is a semantic claim which projects beyond any particular context which inductively supports it. Thus although context provides inductive evidence for which T-sentence applies to an utterance, the T-sentence itself is generated by the theory of meaning that describes the language of which the utterance is a part. Davidson would deny T-sentences to non-verbal language because non-verbal language is not systematic in the same way as verbal language. In particular, non-verbal language is not compositional. That is, in the case of non-verbal language, the requisite semantic structure that would permit us to situate a particular non-verbal utterance in a system—i.e., a language—that relates it semantically to other (verbal and non-verbal) utterances is absent.

Of course, we might regard it as a weakness of Davidson's account that it treats language as utterly recursive. Indeed, Ian Hacking challenges Davidson on this very
point ("The Parody of Conversation" 455-456). That is, though we might agree that language is compositional, we might question the necessity of accounting for this compositionality with one all-encompassing recursive system. Moreover, we might add to this doubt our concern that because T-theories apply only to verbal behaviour, Davidson artificially isolates linguistic meaning from the broader realm of interpersonal communication. One of the lessons Grice taught us was that to mean something by an action one must be in possession of a certain kind of intention, and that these intentions guide linguistic and non-linguistic actions alike. Perhaps, then, the inability of a T-theory to account for non-verbal language can be counted against Davidson’s account of language. Nevertheless, we should not be too quick to dismiss truth-theories. To be sure, a truth-theoretical account of linguistic ability has its virtues. Importantly, T-theories retrieve for Davidson much of the explanatory power of the orthodox position. Specifically, the appeal to T-theories provides Davidson with a way of accounting for our ability to create and interpret novel sentences, and a way of explaining entailment relations between sentences.

First Meaning and Literal Meaning

As useful as Davidson’s notion of first meaning is to our investigation, it is not without its problems. Davidson identifies literal meaning with first meaning. His point in doing so is to pry apart the notions of literal meaning and conventional meaning. But his subsequent identification of first meaning with literal meaning is problematic. Though we might agree with Davidson that Mrs. Malaprop’s utterance ‘This is a nice derangement of
epitaphs' means that the demonstrated item is a nice arrangement of epithets, nevertheless we might deny that it means this literally. That is, we might argue that although Mrs. Malaprop's utterance means, according to her first intention, that the demonstrated item is a nice arrangement of epithets, it means literally, i.e., means-according-to-the-standards-of-our-linguistic-community, that the demonstrated item is a nice derangement of epitaphs. Contra Davidson, then, it would seem that we should distinguish between first meaning and literal meaning.

Indeed, there are utterances for which a first meaning is absent and yet we do not want to say that they are devoid of meaning. A child learning English might transcribe a token of a written English sentence into her notebook without any idea of what it means in English. Depending upon her native language, she may not even recognise the letters—her task may be simply to copy a set of characters from one page to another. In such a case, we would say that she is not in possession of the requisite first intention to specify a first meaning of the sentence she inscribes. Thus, although we would say that the sentence means something—i.e., it means what it means according to the standards of English—it lacks a first meaning. Alternatively, we can imagine a slightly more fantastic case in which books fall from a shelf onto a type-writer producing a token of an English word or phrase. Though we would not hesitate to say that the word or phrase means something (again, it means what it means in English), to be sure, the books will not be in possession of the requisite first intention to specify a first meaning of the word or phrase. Keeping separate the notions of first meaning and literal meaning provides us with a means of explaining the above types of situations.
There are further advantages to this approach. To begin, enlisting a community-standard based notion of literal meaning answers the concern that 'This is a nice derangement of epitaphs' is a *malaprop* just because there are facts about our language against which it is a mistake. More broadly, this approach allows that in practice a notion of correctness arises only within the context of a linguistic community—it does this precisely by providing a role for that community. Davidson himself gestures in this direction: "Perhaps we could even agree... that without a social environment nothing could count as misapplying words in the way that it must be possible to misapply words in speech" ("The Social Aspect of Language" 3). A socially based notion of literal meaning has the additional advantage that it provides us with a means of explaining the anonymous letter scenario. In the anonymous letter scenario, faced with a dearth of clues as to the first intention of the author and, so, to the first meaning of her sentence, we attribute to the sentence the meaning (i.e., the truth conditions) that accord with our perception of the standards of our linguistic community. Indeed, this explanation applies equally to literature, graffiti and other cases in which we are unable to interact socially with the speaker.

What of the implications of this for our notions of *communicated* and *strict literal meaning*? Because, as in the case of malaprops and the like, the meaning that the speaker intends to communicate by her utterance is not always the same as the literal

---

21 cf. Hacking, "The Parody of Conversation" (450).

22 Recall that *communicated literal meaning* was characterised as the meaning that a speaker intends to communicate by her utterance, and that the *strict literal meaning* of an utterance is, roughly, how it might be interpreted by a logician or an unforgiving lawyer.
meaning of her utterance according to the standards of her linguistic community, i.e., because first meaning and literal meaning sometimes diverge, we need to adjust our notion of *communicated literal meaning*. The result is this. For any utterance the following classifications of meaning may apply: the utterance may have a *first meaning* as specified by the first intention of the speaker; further, the utterance may have a *communicated literal meaning* as specified by the standards of her linguistic community; finally, the utterance may have a *strict literal meaning*, also specified by the standards of her linguistic community. Although the *communicated* and *strict literal meaning* of the utterance both are specified by the standards of our linguistic community, nevertheless, they may be distinct.

No doubt there are cases in which one or more of these classifications does not apply. As we discovered above, there are utterances which have no first meaning, but have a *communicated* and *strict literal meaning*, as in the case of the word or sentence produced by the books that fall on the type-writer. Likewise, there are utterances that have a first meaning, but no *communicated* or *strict literal meaning*, as in the case of an idiomatic expression. Nevertheless, there is still a sense in which first meaning is basic. That is, although *communicated* and *strict literal meaning* belong, in a sense, to the language community as a whole, they depend for their origins on first meaning.

---

23 One might argue that a word or phrase $X$ acquires the *communicated* literal meaning $M$ if and only if some (many?) members of the linguistic community utter $X$ with the (first) intention that it be understood to mean $M$ (i.e., as true if and only if truth conditions $TC$ obtain). A similar account could be given for *strict* literal meaning. In both cases, however, we need to delve deeper into who sets the standards that determine literal meaning, and by what means. More will be said about this below. The point here is simply that first meaning is necessarily prior to *communicated* or *strict* literal meaning.
Though all three classifications may apply to an utterance, their content may or may not be distinct. For example, though a sincere utterance of ‘This pen is red’ might have the same first meaning, communicated literal meaning and strict literal meaning, my malaprop ‘All men are cremated the same’ will have as its first meaning that all men are created the same (or, perhaps, that all men are created exactly the same), as its communicated literal meaning that all men are cremated the same and, as its strict literal meaning that all men are cremated exactly the same. In the case of malaprops, slips of the tongue, and the like, first meaning and communicated literal meaning will necessarily diverge (though the communicated literal meaning and strict literal meaning may or may not diverge).

The question remains, how are the standards that determine communicated and strict literal meaning established? There is no easy answer to this question. Given the pervasiveness in everyday discourse of malaprops, ungrammatical utterances, ‘teenager-talk’, slips of the tongue, non-literal language, and the like, it would appear that linguistic standards are not settled according conventions understood as statistical normalities. But how else might they be settled? Here I think we might take a cue from Hilary Putnam’s notion of the division of linguistic labour. According to Putnam, every linguistic community has some terms whose associated ‘criteria’ are known only to a subset of members of that linguistic community, and whose remaining members depend

---

24 Nor could they be settled by conventions understood as pre-arranged agreements, as in the case when the members of a baseball team agree in advance that a certain hand gesture means steal. Members of linguistic communities do not gather together to decide in advance of communicative encounters what the words and phrases of their language will mean—indeed, it would be impossible to accomplish this without there being a system of communication already in place.
on, and defer authority to, these experts ("The Meaning of 'Meaning'") 227f.). Thus although I might have in my vocabulary the word 'gold', and I may use it to communicate successfully about various gold things, I defer to others the authority of determining what exactly qualifies as gold. Indeed, among the members of a linguistic community some individuals or groups generally are considered more authoritative than others as to what certain words and phrases of the language mean. These individuals or groups include experts in specialised fields—for example, physicists serve as the arbiters of meaning for technical terms unique to physics, geologists for the terms unique to geology, astrologers for the terms unique to astrology, and so forth.

But what of words and phrases with wider currency? Though Putnam intends his notion of the division of linguistic labour to apply only to "some terms" of a language, one might ask whether in fact it applies across the board. Indeed, for some languages linguistic standards are determined by a recognized authoritative body, such as the Académie Française in France, the Academia della Crusca in Italy and analogous institutions in Spain and Russia. However, for many languages, including English, no such body exists. Rather, in many of these languages the role of arbiter often is played by the dictionary.25 Indeed, on the whole, speakers of English appeal to, and are willing to be corrected by, a good dictionary if and when they are unsure of the standard (i.e., proper) use of a word or phrase. For example, in a debate with Mrs. Malaprop about the literal meaning of her utterance, one can imagine drawing her attention to the

25 Though to be sure there exist dialects for which there exists neither an accepted body who establishes standards nor a dictionary.
meanings of her words as given in an appropriate—i.e., contemporary and respected—English dictionary. Nevertheless, because for any given language there may exist competing dictionaries, and at points these dictionaries will disagree, and because dictionaries are always, strictly speaking, out of date, they are imperfect arbiters of literal meaning. That is, although there can be little doubt that dictionaries (in English and in other languages) play a central role in establishing standards of proper usage, we need to treat our understanding of their role with care.26

Reconciliation

We are now in a position to reconcile the two seemingly opposed observations made in the Introduction. Recall that the first observation was that there exists no sufficiently determinate set of sufficiently subtle criteria to support a principled distinction between literal and non-literal language. Recall also that the second observation was that we need something like the literal/non-literal distinction to explain (in part) how figurative language works. Because community-based linguistic standards are dynamic and not without controversy, determining the bounds of literal meaning for all words and phrases of a language is impracticable. This coincides with the first observation. Borderline cases of literal and non-literal language are borderline just because the standards that determine their proper (literal) meaning are not yet settled. Indeed, sometimes (often?) we will be unable to decide the literal/non-literal status of a particular utterance. Nevertheless the notion of first meaning allows us to make a principled distinction

26 See the Appendix for a fuller discussion of dictionary meaning.
between what a speaker says and what he implicates or uses his utterance to achieve.

This coincides with the second observation. For any utterance, what it means, if anything, is specified by the speaker's first intention, and is distinct from what the speaker means to accomplish by issuing that utterance (for example, to make a metaphor). Thus the two observations are reconciled by acknowledging different roles for different notions of meaning: first meaning is a semantic notion and applies to individual speakers and their particular utterances; literal meaning is a pragmatic notion that captures the role of community-wide linguistic standards.

Thus whether an utterance qualifies as literal or as non-literal depends in part on which notion of meaning, first meaning or community-standard based literal meaning, is taken as foundational. Hence in assessing the status of Mrs. Malaprop's utterance 'This is a nice derangement of epitaphs', for example, first we must decide against which standard of meaning we want to measure it. By her own linguistic standards her utterance means that the demonstrated item is a nice arrangement of epithets (an utterance which she might employ ironically, or metaphorically, or the like). Strictly speaking, by the standards of our linguistic community, her utterance is nonsensical.
Conclusion

A philosophically sound account of linguistic meaning depends on keeping separate the meanings of the words and sentences of a language and the uses to which those words and sentences are put by the speakers of that language. Put another way, a philosophically sound account of linguistic meaning depends on sustaining the distinction between the literal meanings of utterances and their non-literal uses. Nevertheless, as we have found, arriving at a suitable account of literal meaning is not without its difficulties. If we suppose that a sentence means literally what it means according to some external standard such as Fregean *senses* or shared linguistic conventions, then we cannot account for our ability to interpret utterances which are wholly unfamiliar to us. If we suppose that a sentence means literally what a speaker intends it to be taken to mean (with appropriate constraints), then we cannot account for our ability to interpret utterances for which the speaker’s intention is unknown or absent. In the pages above I argue that these and other important issues pertaining to literal meaning can be resolved if we adopt an account that acknowledges that there are three distinct standards of meaning: *first meaning*, *communicated literal meaning* and *strict literal meaning*. The first meaning of an utterance is specified by the speaker’s first intention—that is, the intention that her utterance be understood as having certain truth conditions—provided that she equips her audience with enough clues so that she might reasonably expect that a radical interpreter would interpret it as she intends. It is this notion of meaning which is foundational. Further, however, an utterance may have a *strict literal meaning* and/or a *communicated
literal meaning, both of which depend for their content on the standards of the linguistic community to which the speaker belongs.

Equipped with this distinction we are better positioned to explain the nature of the distinction between literal and non-literal language. In particular, whether an utterance is employed literally or non-literally depends on which of the above notions of meaning we take as our starting point. Thus the same utterance may, according to the standards of our linguistic community, be non-literal, though, according to its first meaning, it is literal. Moreover, equipped with this distinction we are better positioned to reconcile the seemingly contradictory observations made at the outset of our investigation. It is true that there exists no sufficiently determinate set of sufficiently subtle criteria to support a principled distinction between literal and non-literal language, provided that we employ as our starting point either the notion of communicated literal meaning or the notion of strict literal meaning. This is because the communicated or strict literal meaning of an utterance depends for its content on community-based linguistic standards, standards that are dynamic and sometimes controversial. Nevertheless, we can make a principled distinction between what someone says and what she implicates—that is, between the meaning of her utterance and its non-literal effects—provided that we employ as our starting point the first meaning of her utterance—that is, the meaning of her utterance according to her own linguistic standards.
Appendix
Dictionary Meaning

Dictionaries do several things very well. They set standards of proper spelling and pronunciation. They track the evolution of the lexicon. They demarcate the bounds of verbal language. And, to be sure, they play a central role in determining standards of linguistic use.1 Interestingly, however, in no major dictionary of the English language is it claimed explicitly that its objective is to provide, by means of definitions, the literal meanings of the words and phrases contained therein.2 One might retort that although it is not acknowledged explicitly, nevertheless it is implied that it is in fact literal uses, or literal meanings, that are the subject matter of dictionaries. There are, however, reasons to resist this conclusion. Consider, for example, that no major English dictionary defines ‘literal meaning’ as anything like “the corresponding dictionary definition.” Rather,

1 Monolingual dictionaries as we know them—that is, as alphabetically organised records of the correct spelling, pronunciation, grammatical status, meaning and etymology of the vocabulary of a language—are a relatively recent phenomenon. They arose out of the desire to establish an authoritative voice to settle differences over proper spelling, pronunciation and interpretation. The first attempt to map the English lexicon as a whole was Dr. Samuel Johnson’s two volume Dictionary of the English Language (1755). Johnson’s entries were based on examples taken from the works of best writers of the era, including Dryden, Milton and Addison, in an effort “to preserve the purity... of our English idiom” (qtd. in Cowie, “Language as words: lexicography” 674-75). Johnson’s dictionary was greeted enthusiastically and, for roughly seventy years, stood without serious competition.

typically they define 'literal' as "primitive" or, negatively, as "not figurative or metaphorical." Why would lexicographers avoid making the claim that their definitions provide literal meanings? Unbounded by such a narrowly defined objective, dictionaries are free to (and do) include definitions of dying metaphors, common slang and vulgar expressions, and the like, while avoiding the debate whether these are in fact cases of literal language. The dictionary on my desk, for example, includes definitions of ‘louse’ as "contemptible person," ‘sheep’ as "bashful or docile person," and ‘acrobat’ as "one who changes position nimbly in argument."

Aside from the obvious hesitation among lexicographers to define their task as that of providing literal meanings, there are other reasons to treat carefully the identification of literal meaning and dictionary definition. To begin, there is the banal but important truth that for many languages there are several dictionaries, and at points these dictionaries disagree. This concern involves not only competing definitions of words and phrases that are simply difficult to define—for example, 'love' and 'good'—but also expressions for which the correct use is in dispute. Moreover, because languages evolve, strictly speaking, dictionaries are always out of date. New words enter the

---

3 In rare cases we find more mysterious definitions, such as in The Universal Dictionary of the English Language (1932), which defines "literal meaning" as "words in their narrowest or baldest sense"; or in A Universal and Critical Dictionary of the English Language, which defines "literal" as, surprisingly, "primitive or literal meaning."

4 Some dictionaries explicitly acknowledge that certain definitions of correct usage are in dispute. The Oxford English Dictionary, for example, reserves the designation D for entries of this sort.
language, existing words acquire new uses, old uses become outdated, and so forth.5

Though it may be agreed that dictionaries cannot provide a complete and timely picture of the literal meanings of the words and phrases of a language as a whole, what of their success with individual entries? Might not one argue that in many cases the definition for an individual entry could serve as its literal meaning?6 Were a dictionary definition of a word or phrase to serve as its literal meaning, we would expect that the definition and the word or phrase it defines could be substituted for one another in all literal contexts without changing what is said. So, for example, if the definition of ‘bachelor’ is “unmarried man,” then in all literal contexts, we should expect that one may be substituted for the other without changing the meaning of the utterance. Indeed, in a sentence such as “Kal is a bachelor” we might substitute ‘unmarried man’ for ‘bachelor’ and arrive at a sentence that means the same as the original. But consider the sentence “A bachelor is an unmarried man.” Substituting ‘unmarried man’ for ‘bachelor’ in this sentence yields “An unmarried man is an unmarried man,” which conveys something altogether different. Or, to borrow an example from Baker and Hacker, although “time-table” or “tabular statement” might be offered as the definition of “schedule,” consider

---

5 Lexicographers know this. Consider this passage from the Historical Introduction to the Oxford English Dictionary:

There are many claimants to admission into the recognized vocabulary (where some of them will certainly someday be received), that are already current coin with some speakers and writers, and not yet ‘good English’, or not English at all, to others. There are many words of which it is doubtful whether they are still to be considered as part of the living language; they are alive to some speakers, and dead to others. (Oxford English Dictionary xxviii)

6 Of course, for any given word or phrase, there are likely to be several dictionary definitions on offer. Generally, however, context provides the necessary clues as to which definition applies.
the inapplicability of their substitution in the sentence "The trains are running on schedule" (Language, Sense and Nonsense 215).

There is the further point that many words and phrases are so basic and so widely understood that their definitions are invariably less clear than they are. Although it is a longstanding principle of lexicography that a definition should employ terms that, in Samuel Johnson's words, are "less abstruse than that which is to be explained," this is not always easy (qtd. in Longman Dictionary of the English Language xi). Consider, for example, Johnson's own definition of 'network' as "any thing reticulated or decussated, at equal distances, with interstices between the intersections" (qtd. in Longman Dictionary of the English Language xi). Or consider the Concise Oxford Dictionary's definition of 'sky' as "(the vault of) heaven, weather or climate evidenced by it."

Further, of particular interest to our investigation, dictionaries can play but a limited role in understanding the distinction between strict and communicated literal meaning. Indeed, dictionaries themselves do not draw this distinction explicitly. Moreover, in cases when strict and communicated literal meaning diverge, although sometimes a dictionary will lead us both to the strict literal meaning and to the communicated literal meaning of the sentence, in other cases it will lead us only to the strict literal meaning of the sentence. Consider again the sentence 'The room was silent'. According to my dictionary 'silent' is defined both as "quiet" and as "noiseless," thus leading me to choose between the reading that the people/things in the room were quiet and that the room was without any noise whatever. But recall the sentence 'Nobody knows the trouble I've seen'. The dictionary on my desk defines 'nobody' as "not any
person; not anybody; no one" which leads us only to the *strict literal meaning* of the sentence—i.e., that nobody, not even the speaker, knows the troubles the speaker has encountered. Thus in determining the *communicated literal meaning* of this sentence—i.e., that nobody *but the speaker* knows the troubles the speaker has encountered—my dictionary is of little help.
References


