WHAT CAN BE MEANT:

EVALUATING A SKEPTICAL ARGUMENT BY DONALD DAVIDSON

by

G. ALBERTO RUESGA

B. Sc. Physics, MIT
(1981)

Submitted to the Department of Philosophy
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

at the

Massachusetts Institute of Technology

May 1990

© Massachusetts Institute of Technology, 1990. All rights reserved

Signature of Author

Department of Philosophy
May 4, 1990

Certified by

James Higginbotham
Professor of Philosophy
Thesis Supervisor

Certified by

George Boolos
Chairman, Committee on Graduate Students
For Max.
# Table of Contents

Acknowledgements 4  
Abstract 5  
Introduction 6  

Part I: Presentation of Davidson's skeptical argument 12  
  §1: Some preliminaries 14  
  §2: The Three Plausible Principles model 18  
  §3: The Prior Theory/Passing Theory model 27  
  §4: The Ubiquity Hypothesis 35  
  §5: Summary 37  

Part II: Assessing Davidson's argument 40  
  §1: Some notes on convention 45  
  §2: Davidson versus the Chomskyan mechanist 64  
  §3: Other cognitive faculties 79  
  §4: Davidson's Ubiquity Hypothesis 85  
  §5: Rule-governed versus "mechanical" 88  
  §6: Summary 90  

Part III: Some details of the mechanist response to Davidson 94  
  §1: Some preliminaries 95  
  §2: A model for sentence and utterance interpretation 104  
  §3: Summary 122  

Part IV: Class M phenomena 125  
  §1: A brief look at Davidson's list of Class M phenomena 126  
  §2: The challenge of ungrammaticality 140  
    A few preliminaries 141  
    A little historical perspective 153  
    Summary and discussion 168  
    Current theory 170  

Part V: Summary 185  

Notes 191  
Bibliography 199
Acknowledgements

I would like to thank my thesis committee: Irene Heim, for her helpful eleventh hour reading of my dissertation; Sylvain Bromberger, for his kind encouragement and insightful comments; and especially Jim Higginbotham, my thesis supervisor, for his many attempts to impose order on chaos.

I would like to thank the other members of my department for their encouragement and, alas too seldom, their company. I would especially like to thank George Boolos, without whom \( \neg p \).

A special thank you to my housemates, Deki, Kirsten, George, and Elsa, for putting up with me those times when I was less than present in body and spirit.

I would especially like to thank Max, for all his love, back rubs, and well-timed cups of tea.

Finally, I would like to thank God almighty, with whose help I may yet repay my college loans.
WHAT CAN BE MEANT:
EVALUATING A SKEPTICAL ARGUMENT BY DONALD DAVIDSON

by

G. ALBERTO RUESGA

Submitted to the Department of Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy

ABSTRACT

In this paper I defend a standard mechanistic model of our linguistic competence against a skeptical argument by Donald Davidson ("A Nice Derangement of Epitaphs," appearing in Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson). Taking as his point of departure our ability to interpret malapropisms and other misspeakings, Davidson argues that we cannot give an adequate, rule-governed account of our ability to interpret utterances.

I begin by addressing Davidson's views on the role of convention in language.

I then argue that (1) the mechanistic program outlined by Chomsky and others can satisfactorily accommodate the interpretive phenomena which Davidson claims threaten standard descriptions of our linguistic competence; (2) we need not expect that the mechanistic model will necessarily accommodate every observed interpretive phenomenon; a satisfactory explanation of the interpretive ability in question may, for example, fall outside of what Chomsky has called our cognitive capacity; (3) Davidson is mistaken in suggesting that there are good grounds for extending his skepticism about rule-governed accounts of interpretation from malaprops to all utterances of a language.

Finally, I examine our ability to interpret malapropisms, unfamiliar words, and grammatically garbled utterances in the light of current linguistic theorizing.

I conclude that Davidson's skeptical conclusions are not well-motivated.

Thesis Supervisor: Dr. James Higginbotham, Professor of Philosophy
Introduction
Over the years, philosophers have launched a number of skeptical attacks against the possibility of there being an adequate rule-governed account of our ability to produce and understand language.

As against this picture of language use, Quine, for example, has argued that it is unclear what criteria we can use to decide whether or not linguistic behavior is being guided by internal rules. He argues that "Behavior fits a rule whenever it conforms to it; whenever the rule truly describes the behavior. But the behavior is not guided by the rule unless the behaver knows the rule and can state it."\textsuperscript{1} Quine leaves unchallenged the notion of unconscious compliance with a rule, but only "when this is merely a question of fitting"\textsuperscript{2}:

Bodies obey, in this sense, the law of falling bodies, and English speakers obey, in this sense, any and all of the extensionally equivalent systems of grammar that demarcate the right totality of well-formed English sentences. These are acceptably clear dispositions on the part of bodies and English speakers. The sticking point is this Chomskian midpoint between rules as merely fitting, on the one hand, and rules as real and overt guides on the other; Chomsky's intermediate notion of rules as heeded inarticulately.\textsuperscript{3}

This "inarticulate" heeding of rules, Quine says, is a notion "deserving of close methodological attention."\textsuperscript{4}

Kripke's celebrated analysis of Wittgenstein's "private language" argument\textsuperscript{5} has yielded yet another assault on rule-governed explanations of language use. Kripke has argued that Wittgenstein's skeptical argument about rule-following can be naturally extended to the notion of
rule-following endorsed, for example, by the Chomskyan linguist. Thus he writes:

...if statements attributing rule-following are neither to be regarded as stating facts, nor to be thought of as explaining our behavior..., it would seem that the use of the ideas of rules and of competence in linguistics needs serious reconsideration, even if these notions are not rendered 'meaningless'. ...These questions would arise even if... we deal with rules, like addition, that are stated explicitly. These rules we think of ourselves as grasping consciously; in the absence of Wittgenstein's sceptical arguments, we would see no problem in the assumption that each particular answer we produce is justified by our 'grasp' of the rules. The problems are compounded if, as in linguistics, the rules are thought of as tacit, to be reconstructed by the scientist and inferred as an explanation of behavior.

Thus, unlike Quine, the Wittgensteinian skeptic calls into question the notion of "implicit and unconscious conformity to a rule" whether or not this conformity is merely a matter of "fitting."

Other philosophers, whom I will not discuss here, have attacked a number of subsidiary assumptions made by the rule-attributing language theorist.

In 1986, Davidson published a paper in which he offered what was essentially a new kind of skeptical argument against the enterprise. Citing such phenomena as our ability to recover the meaning of a malapropism, Davidson argued that a rule-governed account of interpretation was inadequate to the task of accounting for many of the
ways that we recover meanings from utterances. He argued that such phenomena as our ability to interpret malaprops and other slips of the tongue, our ability to interpret grammatically garbled utterances, and our ability to interpret new words, threatened "standard descriptions of linguistic competence..."10

In this paper, I will defend a standard mechanistic model of our linguistic competence against Davidson's skeptical claims. I will, in the process, explore the extent to which our interpretive abilities might "reach beyond" what mechanist accounts of interpretation can deliver. The "mechanistic account" which I have in mind is one that seeks to explain our linguistic capacities in terms of a shared knowledge—on the part of both speaker and hearer—of the rules and principles that govern a language. The mechanist attributes to each competent language-user the knowledge of some language $L$, and attempts to characterize this knowledge by constructing a theory of the language-user's grammatical competence. According to the mechanist, our language faculty can be thought of as... kind of cognitive "machine" which, on the production side, translates an intention to express a thought with content $C$ into a phone sequence (or other symbolic string) with meaning $C$, and which, on the interpretive side, translates incoming strings of formatives into representations of meaning. This machine is assumed to be for the most part cognitively impenetrable—to use an expression coined by Pylyshyn. The mechanistic account which I have in mind supplements the proposed production/interpretation machine with a number of pragmatic rules that govern how the meaning of an utterance varies as a function of its literal
meaning and its context of utterance. Strictly speaking, I should call this the "mechanistic/pragmatic" account of language use, but I will opt for the shorter moniker "mechanistic account" when there is no danger of confusion, and I will refer to the proponent of this model simply as the "mechanist."

I present Davidson's argument in Part I.

In Part II, I assess Davidson's challenge and argue that his skepticism is not well-motivated. After briefly considering Davidson's argument against the role of convention in language (§1), I argue that (1) the mechanistic program outlined by Chomsky and others can satisfactorily accommodate the phenomena which Davidson claims threaten standard descriptions of our linguistic competence; (2) we need not expect that the mechanistic model will necessarily accommodate every observed interpretive phenomenon; a satisfactory explanation of the interpretive ability in question may, for example, fall outside of what Chomsky has called our cognitive capacity; (3) Davidson is mistaken in suggesting that there are good grounds for extending his skepticism about rule-governed accounts of interpretation from malaprops to all utterances of a language.

In Part III, I explore some of the details of my response to Davidson, and the assumptions that underlie this response. I take a closer look at the model of sentence and utterance interpretation which I believe answers Davidson's demand for "a machine which, when fed an arbitrary
utterance (and certain parameters provided by the circumstances of the
utterance), produces an interpretation. 11

In Part IV I examine Davidson's list of abilities which he claims
threaten standard descriptions of our linguistic competence. I pay
particularly close attention to our ability to recover the meanings of
grammatically garbled utterances. I end by concluding that these abilities
constitute no real challenge to the mechanist's enterprise.
Part I

Presentation of Davidson's skeptical argument
Introduction

In a much discussed paper titled "A Nice Derangement of Epitaphs," Davidson comes to some rather heady conclusions:

I conclude that there is no such thing as a language, not if a language is anything like what many philosophers and linguists have supposed. There is therefore no such thing to be learned, mastered, or born with. We must give up the idea of a clearly defined shared structure which language-users acquire and then apply to cases. And we should try again to say how convention in any important sense is involved in language; or, as I think, we should give up the attempt to illuminate how we communicate by appeal to conventions.

I will not here attempt to reconcile these conclusions with views that we might attribute to Davidson on the basis of his other works on language. Nor will I attempt to reconstruct how Davidson's ideas have changed or evolved in the time since his paper was first published. I am not so much interested in Davidson's current theories on language, as I am in the kind of skeptical argument that he outlines, and the linguistic data that he adduces in its defense. I do not believe that Davidson's argument is very good; I hope to show, however, that his argument fails in some rather interesting ways. In challenging Davidson, the philosopher of language is compelled to sharpen his account of a number of linguistic notions that he might otherwise take for granted—among them, the distinctions between core and periphery, competence and performance, the notion of a linguistic rule, and the very concept of a language.

The premise that lies at the heart of Davidson's skeptical
argument is this: There can be no adequate rule-governed description of our ability to interpret utterances. Davidson essentially wants to argue that our interpretive abilities far outstrip anything that theory can deliver. The exact senses of these propositions will become clearer in the course of my discussion.

§1 Some preliminaries

As the passage quoted above suggests, Davidson has a specific model of linguistic performance (and competence) in mind. This model will emerge in the course of my exposition, and I will extract its salient features from the context of Davidson's argument rather than from any assumptions that we might be tempted to make about his own theory preferences.

Davidson begins by drawing our attention to a language-user's native ability to interpret utterances that contain one or more malapropisms. (This ability is "native" insofar as it is acquired without explicit training—although in keeping with standard usage we should say that it is the capacity for acquiring this ability that is native, rather than the ability itself.) Imagine for a moment that Mrs. Malaprop\(^3\) and I are walking beside the soon-to-be-dismantled Berlin Wall, admiring the fanciful graffiti. Feeling herself overcome by a strange curatorial mood, Mrs. Malaprop points to an artful section of the wall and exclaims, "What a
nice derangement of epitaphs!" Before she has a chance to correct herself — if, indeed, she has noticed her own error — I correctly recover her intended meaning (nice arrangement of epithets). According to Davidson, my ability to do this falls into a very peculiar category of linguistic phenomena, a category that may also include "our ability to perceive a well-formed sentence when the actual utterance was incomplete or grammatically garbled, our ability to interpret words we have never heard before, to correct slips of the tongue, or to cope with new idiolects." These phenomena, as we shall see, are what motivate Davidson's skeptical conclusions about language. According to Davidson, these special linguistic abilities "threaten standard descriptions of linguistic competence ... ," a claim that I will assess in some detail in Part II.

Davidson makes it clear from the beginning of his paper that he does not intend to "obliterate or even blur the distinction between speaker's meaning and literal meaning." He notes, however, that the term "literal meaning" "is too encrusted with philosophical and other extras to do much work." Here I take Davidson to be referring to the question of whether or not we can coherently talk about the meaning of a sentence in abstraction from any occasion of its utterance, a question to which I will return. Davidson skirts this issue by positing what he calls first meaning. According to him, the concept of first meaning "applies to words and sentences as uttered by a particular speaker on a particular occasion," with the added proviso that the occasion, speaker, and audience be "normal" or "standard" in some unspecified way. Given that these normality conditions hold, "the first meaning of an utterance will be what
should be found by consulting a dictionary based on actual usage..."10
First meanings are obviously intended to serve as palatable substitutes for the disreputable literal meanings — palatable to those theorists who, for example, countenance only speaker meanings.

We may, of course, question the propriety of this move. We are told that in order to recover the first meaning of a sentence \(s\), we must imagine the meaning that \(s\) would have if uttered by a standard speaker to a standard audience on some standard occasion. But how are we to decide what constitutes a standard or normal context of utterance? Suppose that we were to try using Davidson's prescription to recover the first meaning of the following sentence:

[1] The ham sandwich barked at my wristwatch.

Is it possible to specify a "standard" context of utterance for sentence [1]? If not, do we then conclude that [1] has no first meaning or, casting caution to the wind, no literal meaning? I take it that a standard context of utterance for sentence [1] would be one in which the speaker can utter [1] and mean it \textit{literally}. A plausible candidate for a standard context of utterance for sentence [1] might be something like the following:

[2] \begin{itemize}
  \item Occasion: A possible world in which an animate ham sandwich — that thinks he's a dog — barks at my counterpart's wristwatch.
  \item Speaker: My possible world counterpart
  \item Audience: My counterpart's audience
\end{itemize}
Here we have a context in which we might be tempted to say that the speaker utters sentence [1] and means it *literally*—insofar as we can trust our intuitions about possible worlds and possible world counterparts. It appears, however, that we have already appealed to our intuition of sentence [1]'s literal meaning in reconstructing its standard context of utterance. Suppose that we were to try "impoverishing" the context given in [2] as follows:

[3] Occasion: Any occasion will do, as long as a ham sandwich and a wristwatch are present.

Speaker: The owner of the wristwatch

Audience: Anybody

This "standard" context has the advantage of being much more general than the context given in [2]: it does little more than assemble the cast of characters suggested by the noun phrases in sentence [1]. But now it is unclear what Davidson's elaborate prescription buys you, above and beyond the requirement that we take sentence [1] *literally*. Davidson's prescription may of course simply be intended as a means of *excluding* a number of "non-standard" occasions of utterance—occasions, for example, in which we utter sentence [1] and mean it metaphorically, or ironically, or sarcastically, etc. But even here it is doubtful that we can escape acknowledging the literal meaning of sentence [1], for it appears that the criterion of membership in our list of non-standard occasions of utterance is that the occasion be one in which we utter sentence [1] and mean it *non-literally*. The moral of this excursus: literal meanings may be suspect theoretical constructions, but first meanings create more problems than
they solve. In Part III, §2 I will present what I believe to be a non-question-begging account of first or literal meaning, one which has the added advantage of preserving Davidson's intentions.

§2 The Three Plausible Principles model

Having thus tabled our concerns about first meanings, we turn now to the question of what role they play in Davidson's skeptical argument. Davidson articulates three principles that plausibly govern first meanings in language. Together these principles suggest a kind of model for interpretation — more specifically, a model of how a competent language-user successfully interprets the utterances of other language-users. In order to restrict our attention to first meaning, as Davidson suggests, we should imagine that the speaker and his interpreter exchange words in as neutral, or "standard," a context as possible (whatever that means). Davidson's three principles are as follows:

1. First meaning is systematic. A competent speaker or interpreter is able to interpret utterances, his own or those of others, on the basis of the semantic properties of the parts, or words, in the utterance, and the structure of the utterance. For this to be possible, there must be systematic relations between the meanings of utterances.

2. First meanings are shared. For speaker and interpreter to communicate successfully and regularly, they must share a method of interpretation of the sort described in (1).

3. First meanings are governed by learned conventions or regularities. The systematic knowledge or competence of the
speaker or interpreter is learned in advance of occasions of interpretation and is conventional in character.  

In the discussion that follows, I will refer to the model of interpretation suggested by these principles as the Three Plausible Principles (TPP) model.

Principle (1) is, I believe, fairly straightforward: it is little more than the claim that the interpreter has a system for interpreting the utterances of other language-users. Davidson suggests that we think of this system as a "machine which, when fed an arbitrary utterance (and certain parameters provided by the circumstances of the utterance), produces an interpretation." The "parameters" to which Davidson alludes may include those properties of the context of utterance that help "fix" the indexical elements of the sentence used to make the utterance. They may also include those context parameters that help the interpreter determine the referents of the names and definite descriptions used by the speaker, etc.

Principle (3) requires some explanation, for the notion of convention figures quite prominently in it. Davidson's interest in the role of convention in language is underscored by another passage in the Epitaphs paper:

According to [the account of linguistic competence given by principles (1) - (2)], each interpreter (and this includes speakers, since speakers must be interpreters) comes to a successful linguistic exchange prepared with a 'theory' which constitutes his basic linguistic competence, and which he
shares with those with whom he communicates. Because each party has such a shared theory and knows that others share his theory, and knows that others know he knows, (etc.), some would say that the knowledge or the abilities that constitute the theory may be called conventions.\footnote{13}

It is not entirely clear to me that the shared and symmetric knowledge alluded to in this passage in any way constitutes a convention. We can, for example, be reasonably certain that others share our "theory" of vision (e.g. that what looks square to me from this vantage point will also look square to someone else who occupies my vantage point), and others can know that we know this, etc. Nevertheless we would not say that the special abilities that constitute vision are in any way conventional. Something extra is needed if we are to separate the notion of a convention from that of a felicity condition, or from a very natural assumption about other people's abilities — linguistic and otherwise. I will consider Davidson's argument against the role of convention in language in Part II, §1.

And now we close the circle: Davidson does not believe that principles (1) and (2) are incompatible with the existence of malapropisms,\footnote{14} although these principles survive only "when understood in rather unusual ways."\footnote{15} Our troubles begin, he claims, when we attempt to combine these two principles with principle (3).\footnote{16} According to Davidson, Mrs. Malaprop is a kind of linguistic criminal who has gotten away with her crime:

... the interpreter comes to the occasion of utterance armed with a theory that tells him (or so he believes) what an
arbitrary utterance of the speaker means. The speaker then says something with the intention that it will be interpreted in a certain way, and the expectation that it will be so interpreted. In fact this way is not provided for by the interpreter's theory. But the speaker is nevertheless understood; the interpreter adjusts his theory so that it yields the speaker's intended interpretation. The speaker has 'gotten away with it'. The speaker may or may not ... know that he has got away with anything; the interpreter may or may not know that the speaker intended to get away with anything. What is common to the cases is that the speaker expects to be, and is, interpreted as the speaker intended although the interpreter did not have a correct theory in advance.17

It is this last assertion that provides us with the key to this passage, and indeed to Davidson's argument as a whole. If in fact the speaker has gotten away with her linguistic crime, it is only because she was aided and abetted by her interpreter. How was it possible, Davidson asks, for Mrs. Malaprop's interpreter to understand that by uttering "a nice derangement of epitaphs" she in fact meant "a nice arrangement of epithets"? According to Davidson, this interpretation "is not provided for by the interpreter's theory."18 Principle (3) of the TPP model cannot stand side by side with principles (1) and (2) because it asserts that the interpreter's linguistic competence is prepared in advance of the occasion of interpretation, and is conventional in nature. Here is what I take Davidson to mean. Although he acknowledges that speaker and hearer do in fact come to agree on their interpretation of the malaprop, so that the interpretation is in some sense "shared," Davidson does not believe that the hearer comes to agree with the speaker by appealing to, or making use of, any interpretive competence that he shares with the speaker, and which can be adequately modeled as some set of rules governing the
interpretation of particular utterances. There is, furthermore, a sense in
which the very utterance of the malaprop ruptures a convention that some
theorists believe is essential to any adequate description of
communication by language—or so Davidson believes. (I will address the
role of convention in language in Part II, §1.)

Principle (3) does not appear to jive with the fact that a
competent language-user can successfully interpret malaprops,
grammatically garbled sentences, words he has never heard before, slips
of the tongue, and new idiolects. These are not, according to Davidson,
abilities that have their point of origin in a fixed, shared "interpreting
machine" that we carry around with us and deploy as the need arises.
These are, rather, complex abilities that depend in good measure on our
own "wit, luck, and wisdom." Being prepared in advance to interpret a
malaprop would be like preparing an answer to a question that we have not
yet been asked.

The successful interpretation of the malaprop, Davidson would
like to argue, poses a special problem for the rule-attributing theorist
that is quite different from the problem posed by the interpretation of
other non-literal forms of speech, like irony, for example. The
interpretation of ironic utterances is usually explained by the
rule-attributing theorist as a two stage process: (1) the interpreter's
linguistic competence (modeled as a set of rules and principles) delivers
to him a sharply restricted range of possible meanings for the utterance;
(2) the interpreter's general intelligence, together with his knowledge of
the maxims and principles governing the use of nonliteral speech, allow him to infer that the speaker was speaking ironically and thereby to choose the speaker's intended meaning from the utterance's range of possible meanings. Davidson wants to argue that there is no rule-governed competence which we can attribute to the interpreter on the basis of which he can generate a range of possible meanings to the malaprop. In other words, Davidson believes that there is no first stage of a two stage interpretive process possible for the malapropism. I should also note in passing that unlike the case of speaking ironically, or metaphorically, etc., the malapropper intends that his utterance be taken literally. I will explore the special significance of this fact in Part II, §1.

We might also note that at several points in the Epitaphs paper Davidson tips his hat to Paul Grice. He writes, for example:

... [Paul Grice] has shown why it is essential to distinguish between the literal meaning (perhaps what I am calling first meaning) of words and what is often implied (or implicated) by someone who uses those words. He has explored the general principles behind our ability to figure out such implicatures, and these principles must, of course, be known to speakers who expect to be taken up on them. Whether knowledge of these principles ought to be included in the description of linguistic competence may not have to be settled ...21

And in the very next passage Davidson says, "I dip into these matters only to distinguish them from the problem raised by malapropisms and the like. The problems touched on [in the passage above] all concern the ability to interpret words and constructions of the kind covered by our conditions (1) - (3)."22 I will not here explore the question of how Grice's program
might fit into or connect with the TPP model of sentence-interpretation. As Davidson sees it, the conversational implicatures that Grice has written about are governed by "general principles" — leaving open the question of whether or not these principles should be included in a description of the speaker's linguistic competence. Malapropisms, on the other hand, are intrinsically "wild," fettered neither by the rules of a compositional semantics, nor the rules and heuristics that govern the interpretation of nonliteral or indirect speech. Malaprops are conceived in linguistic sin, and the fact that they can be interpreted presents a special problem for the theorist.

The special problem presented by malapropisms is highlighted by a passage in "Communication and Convention" (Davidson, 1984), where the issue is not so much our ability to interpret malaprops, as it is our ability to "shift ground appropriately" in the process of interpretation:

... formal methods are at their best applied to syntax; here at least there is good reason to expect the same model to fit a number of speakers fairly well. And there is no clear reason why each hypothesized method of interpretation should not be a formal semantics for what we may in a loose sense call a language. *What we cannot expect, however, is that we can formalize the considerations that lead us to adjust our theory to fit the inflow of new information.* No doubt we normally count the ability to 'shift ground appropriately' as part of what we call 'knowing the language'. But *in this sense, there is no saying what someone must know who knows the language,* for intuition, luck, skill must play as essential a role here as in devising a new theory in any field; and taste and sympathy a larger role.23
Davidson echoes this same theme in the Epitaphs paper when he asserts that the theory used by the hearer to interpret the malapropping speaker is "derived by wit, luck, and wisdom." He is skeptical about the possibility of "regularizing" this process of interpretation; likewise, in the above passage, he is skeptical about the possibility of giving a formal account of our ability to "shift ground appropriately" in order to accommodate the inflow of new linguistic information. In Mrs. Malaprop's case, it is not her utterance that is "new," but rather the connection between her utterance and its intended meaning. According to Davidson, a language-user's ability to recover this connection cannot be explained by what he calls "formal methods." This connection is established by a combination of luck and pluck, and formal methods — whichever these might be — will fare about as well in this domain as they now do in helping us to explain matters of personal taste.

We should note, before proceeding any further, that Davidson equivocates between first meaning and utterer meaning, at least in the malaprop case. He often speaks as if the meaning recovered by the interpreter of the malaprop were in fact the utterance's first meaning. (Hence the stipulation that principles (1)-(3) govern first meanings in language.) This cannot be right, for, as we have seen, Davidson intends his notion of first meaning to correspond to the linguist's notion of literal meaning; and what the interpreter of the malaprop recovers, if he is indeed successful, is the speaker's intended meaning, which, as a matter of fact is not the literal meaning of the sentence that the speaker actually succeeds in producing. I will not pursue this matter any further because I
do not believe that it affects the substance of Davidson's argument.

Summarizing the argument thus far: Davidson proposes three principles that govern first meanings in language and which constitute or suggest a model of our linguistic competence (what I have elsewhere called the TPP model). The first and second of these principles are compatible with the existence of malaprops, in a sense that I will explain in Part II, §1; the third, however, is not. The competent language-user's ability to interpret malapropisms cannot be formally modeled as a set of rules or principles which he acquires in advance of the occasion of interpretation. Therefore the Three Plausible Principles model of our linguistic competence — the model widely held by linguists and philosophers — is in error.

Notice that Davidson is not so much challenging the details of a well-articulated theory of linguistic competence, as he is questioning the assumptions that underlie a very general framework of linguistic inquiry. Conspicuously absent from our discussion, however, is the model or framework that Davidson would have replace the TPP model discussed above.

In the following section I will continue to develop Davidson's argument by describing his own picture of what transpires between speaker and interpreter.
§3 The Prior Theory/Passing Theory model

Davidson presents what he calls a "highly simplified and idealized proposal" about what goes on between speaker and interpreter:

An interpreter has, at any moment of a speech transaction, what I persist in calling a theory. (I call it a theory, as remarked before, only because a description of the interpreter's competence requires a recursive account.) I assume that the interpreter's theory has been adjusted to the evidence so far available to him: knowledge of the character, dress, role, sex, of the speaker, and whatever else has been gained by observing the speaker's behavior, linguistic or otherwise. As the speaker speaks his piece the interpreter alters his theory, entering hypotheses about new names, altering the interpretation of familiar predicates, and revising past interpretations of particular utterances in the light of new evidence.

Davidson distinguishes what he calls the interpreter's prior theory from his passing theory:

For the hearer, the prior theory expresses how he is prepared in advance to interpret an utterance of the speaker, while the passing theory is how he does interpret the utterance. For the speaker, the prior theory is what he believes the interpreter's prior theory to be, while his passing theory is the theory he intends the interpreter to use.

I shall, following Davidson, assume that there is no harm in speaking as if the interpreter were using the theory that we use to describe his competence. It should be noted that at each instant during the exchange between speaker and interpreter, the hearer possesses a prior theory that
expresses how he is disposed to interpret the speaker. The hearer's prior theory may of course be updated during the exchange, as new information about the speaker's idiolect becomes available, but the hearer's prior theory is never replaced by his passing theory. To make the distinction between prior and passing theory clear, on the part of both hearer and speaker, let us consider the following two cases:

Case A: "Derangement" means arrangement and "epitaph" means epithet in Mrs. Malaprop's idiolect. She says, "What a nice derangement of epitaphs," and her interpreter (correctly) understands her to mean \[[[\text{What a nice arrangement of epithets}]]\]. Suppose further that she does not know the words *arrangement* and *epithet*, so that she is in no way disposed to say "What a nice arrangement of epithets" and mean anything by it.

Case B: "Derangement" means derangement and "epitaph" means epitaph in Mrs. Malaprop's idiolect. She is perfectly disposed to say "What a nice arrangement of epithets" and mean \[[[\text{What a nice arrangement of epithets}]]\], but because she's had too many martoonis, her tongue slips, and out come the words "What a nice derangement of epitaphs." Again her interpreter correctly understands her to mean \[[[\text{What a nice arrangement of epithets}]]\].

To simplify the exposition, let us suppose that the interpreter is a competent English speaker who does not suspect Mrs. Malaprop of being a habitual malapropster. Let \(s\) be the sentence "What a nice derangement of epitaphs."

In Case A as in Case B, the hearer is ex hypothesi predisposed to interpret \(s\) in the "standard" way. Therefore, in his prior theory for Mrs. Malaprop, \(s\) means \[[[\text{What a nice derangement of epitaphs}]]\]. At the instant
of interpretation, however, the hearer’s passing theory tells him that in uttering Mrs. Malaprop really means [[What a nice arrangement of epithets]]. What happens to the hearer’s prior theory after Mrs. Malaprop makes her blunder? Should his prior theory incorporate the information given to him by his passing theory? This depends in large measure on what the hearer believes the nature of Mrs. Malaprop’s mistake to be. If, as we have described it in Case A, Mrs. Malaprop is quite confused about the “real” meanings of the words derangement and epitaph, the hearer should update his prior theory in such a way as to take account of this fact—especially in the light of evidence that Mrs. Malaprop misuses these words systematically. If, on the other hand, Mrs. Malaprop shares the hearer’s idiolect (as in Case B), he does best to leave his prior theory for her unaffected, except perhaps for the parenthetical addition of the statement that she is under some circumstances liable to pronounce “arrangement” as “derangement,” and “epithet” as “epitaph.”

The speaker’s passing theory presents us with a different challenge. According to Davidson, the speaker’s passing theory is “the theory he intends the interpreter to use.”30 But this definition is ambiguous in at least two ways: Is the speaker’s passing theory the theory he intends the interpreter to use in general, or on just that particular occasion of utterance? This problem does not surface for Case A, in which Mrs. Malaprop clearly intends her interpreter to use a theory wherein s means [[What a nice arrangement of epithets]], both on the occasion of uttering s and in general. But how do we characterize the speaker’s passing theory in Case B? If Mrs. Malaprop catches her error, it
is likely that she will correct herself, or smile coyly, or otherwise indicate to her hearer that the form of her utterance was not what she intended—however clear its intended meaning may have been. But this tells us nothing about her passing theory. It appears that in Case B there are two candidates for Mrs. Malaprop's passing theory. On the occasion of uttering s, Mrs. Malaprop intends that her interpreter use a theory wherein s means [[What a nice arrangement of epithets]]; in general, however, she intends that her interpreter use a theory wherein s means [[What a nice derangement of epitaphs]]. Each of these two intentions determines a different passing theory for Mrs. Malaprop in accordance with Davidson's definition. How are we to decide between the two? The solution to this problem lies in another passage where Davidson writes:

The passing theory is where, accident aside, agreement is greatest. As speaker and interpreter talk, their prior theories become more alike; so do their passing theories.31

In Case B, the interpreter's prior theory is one in which s means [[What a nice derangement of epitaphs]], and here Mrs. Malaprop and her interpreter are in full agreement. The interpreter's passing theory, on the other hand, has it that s means [[What a nice arrangement of epithets]]. If, as the above passage suggests, agreement is greatest in the passing theory, then we should expect the two passing theories—Mrs. Malaprop's and her interpreter's—to coincide. I will assume, therefore, that Davidson would have Mrs. Malaprop's passing theory record only her short term intentions for her interpreter's theory.

Davidson nowhere suggests that his talk of prior and passing
theories is anything more than a façon de parler. He does not, for example, predict that the elucidation of prior and passing theories will constitute the cutting edge of future linguistic theorizing. The picture that Davidson sketches of what transpires between speaker and interpreter is not an idea whose time has come, but rather a foil to the "standard" model of our linguistic competence as given by principles (1)-(2):

I think that the distinction between the prior and passing theory, if taken seriously, undermines this commonly accepted account of linguistic competence and communication. Here is why. What must be shared for communication to succeed is the passing theory. For the passing theory is the one the interpreter actually uses to interpret an utterance, and it is the theory the speaker intends the interpreter to use. ...

But the passing theory cannot in general correspond to an interpreter's linguistic competence. Not only does it have its changing list of proper names and gerrymandered vocabulary, but it includes every successful—i.e. correctly interpreted—use of any other word or phrase, no matter how far out of the ordinary. Every deviation from ordinary usage, as long as it is agreed on for the moment (knowingly deviant, or not, on one, or both, sides), is in the passing theory as a feature of what the words mean on that occasion. Such meanings, transient though they may be, are literal; they are what I have called first meanings. 32

How, exactly, does the distinction between prior and passing theory undermine the account of linguistic competence given by principles (1) and (2)? The passing theory, Davidson seems to be saying, yields meanings for utterances that no compositional semantics could ever give you; neither could these meanings be arrived at by the rules or heuristics of some pragmatic theory. My passing theory for Mrs. Malaprop will assign a meaning to every peculiar construction which she has uttered and which
I have had the occasion to interpret successfully. With its "changing list of proper names and gerrymandered vocabulary," and its other deviations from standard usage, the passing theory will be, by its very nature, unsystematic. It does not appear that a competent speaker or interpreter unravels the intended meaning of a malaprop, for example, by appealing to the semantic properties of its parts and the structure of the utterance. Our theory of pragmatics is no help to us here either—or so Davidson appears to claim. There is no system, Davidson is saying, that will allow us to shift ground appropriately when confronted with radically nonstandard utterances. So much for principle (1). And if principle (1) falls, so does principle (2): for if there is no real "system" or "method" of interpretation, then there is no such system or method for speaker and interpreter to share.

I have been purposely vague about the nature of the prior and passing theories in part because it is difficult to nail Davidson down on this question. He appears to equivocate between two very different models of the passing theory. In one model, the interpreter's passing theory is like his prior theory plus a number of "accretions" that correspond to the idiosyncrasies of the speaker's idiolect. Thus it has its "changing list of proper names and gerrymandered vocabulary," and also includes "every successful ... use of any word or phrase, no matter how far out of the ordinary." So, for example, we have a model of the interpreter's passing theory that looks something like the following:

\[4a\] Core (or prior?) +

1. "Derangement" means arrangement theory
2. "Epitaph" means epithet
3. etc.

In contrast to this model, we have one in which "knowing a passing theory is only knowing how to interpret a particular utterance on a particular occasion"; it is a theory whose expected field of application is "vanishingly small." This suggests rather a model like

[4b] "A nice derangement of epitaphs" means a nice arrangement of epithets,

a small piece of propositional knowledge construed separately from the interpreter's more complete prior theory. Given this model, it is clear why Davidson says that a passing theory "is not a theory of what anyone (except perhaps a philosopher) would call an actual natural language." For the rest of this discussion, I will assume that by "passing theory" Davidson is referring to the latter, more impoverished model, although, as we shall see, this choice in no way affects the substance of his argument.

The role of the prior theory and the passing theory in Davidson's argument should now be apparent. Let us, for the sake of brevity, call this account of our linguistic competence the Passing Theory (PT) model. The PT model is intended by Davidson to serve as a kind of theory-neutral description of what transpires between speaker and hearer. The expression "theory neutral description" is, of course, almost a contradiction in terms, but by this I mean a description that makes very minimal assumptions about the mechanisms that underlie the phenomena in question. In this case we have a description of our linguistic competence that does not invoke the notion of an "interpreting machine"
shared by both speaker and hearer, and acquired by each prior to the occasion of utterance. To be sure, Davidson persists in talking about shared "theories." But I attribute this persistence to his inability to find a way around our need to give a recursive account of the interpreter's competence. "We cannot describe what an interpreter can do except by appeal to a recursive theory of a certain sort," Davidson says, "It does not add anything to this thesis to say that if the theory does correctly describe the competence of an interpreter, some mechanism in the interpreter must correspond to the theory." A recursive model of the interpreter's boundless capacity to interpret novel utterances is apparently something that no serious description of his linguistic competence could do without. Nevertheless, according to Davidson, what is essential for successful interpretation is that the passing theory be shared, and this is little more than the requirement that the speaker and hearer agree on the intended meaning of the utterance—a requirement not fraught with too many theoretical commitments.

Here, then, is Davidson's argumentative strategy: by "weakening" his premise (i.e. the PT model), he hopes to make his skeptical conclusions correspondingly stronger. Furthermore, by showing that his skeptical conclusions follow from some very minimal assumptions about the nature of our linguistic competence, he effectively bypasses the methodological squabbles endemic to linguists and philosophers.

There is, of course, more than one premise to his argument. We can agree with Davidson that in some sense (to be further explicated) all
that is needed for successful communication is that speaker and hearer agree from time to time on a passing theory. But we certainly need not agree that "there are no rules for arriving at passing theories, no rules in any strict sense, as opposed to rough maxims and methodological generalities." 41

§4 The Ubiquity Hypothesis

Davidson’s argument would not be so bothersome if only malaprops were at issue. But he wants to cast his net more widely than this:

Malapropisms fall into a different category, one that may include such things as our ability to perceive a well-formed sentence when the actual utterance was incomplete or grammatically garbled, our ability to interpret words we have never heard before, to correct slips of the tongue, or to cope with new idiolects. 42

I will consider this peculiar category of linguistic phenomena more carefully in Part IV. According to Davidson, these phenomena also "threaten standard descriptions of linguistic competence." 43 Not only do we interpret malaprops on a fairly regular basis, but we also recover the intended meanings of grammatically garbled utterances, utterances that contains words and names that are unfamiliar to us, and utterances that contain slips of the tongue. In our normal linguistic intercourse with others we are constantly running up against speakers with idiolects
substantially different from our own. These and other similar phenomena permeate language to such an extent that Davidson concludes, "[T]here is no such thing as how we expect, in the abstract, to be interpreted." The phenomenon of a speaker’s making an utterance which is not interpreted by the hearer in the "standard" way, but which is nevertheless understood by the hearer as the speaker intended, is a phenomenon which Davidson believes is "ubiquitous." According to Davidson:

What two people need, if they are to understand one another through speech, is the ability to converge on passing theories from utterance to utterance. Their starting points, however far back we want to take them, will usually be very different—as different as the ways in which they acquired their linguistic skills. So also, then, will the strategies and stratagems that bring about convergence differ.

There is then, as Davidson claims, "no learnable common core of consistent behavior," and we therefore engage in a kind of radical homophonic interpretation all the time. I will call this further claim of Davidson’s the "Ubiquity Hypothesis," and I will postpone a discussion of it until Part II. Simply put, Davidson's claim is that the fancy guesswork we must use to recover the intended meaning of the malaprop is the kind of guesswork that we are compelled to use in all cases of utterance interpretation.
§5 Summary

Even the most sympathetic reader would be startled by the strength of Davidson’s conclusions:

The problem we have been grappling with depends on the assumption that communication by speech requires that speaker and interpreter have learned or somehow acquired a common method or theory of interpretation—as being able to operate on the basis of shared conventions, rules, or regularities. The problem arose when we realized that no method or theory fills this bill. ... I conclude that there is no such thing as a language, not if a language is anything like what many philosophers and linguists have supposed. 48

In the chapter that follows, I will argue that Davidson’s skepticism about “a clearly defined shared structure which language-users acquire and then apply to cases” 49 is unwarranted. His mistake, I hope to show, is essentially one of underestimating the robustness of rule-governed accounts of interpretation, and of misconstruing linguistic’s self-conception.

I should mention in passing that not everybody agrees with my reconstruction of Davidson’s argument. In a recent paper, Alex George (George, 1990) offers the following analysis:

Reflection on communication involving ‘deviant’ speech reveals that neither the theory used to interpret another’s utterances, nor the theory employed by a speaker to assign interpretations to his or her own words, plausibly offers characterizations of a language, in the ordinary sense. This is allegedly so because the wide range of communicational phenomena in question
shows that our idiolects (what these theories are theories of) so vary with the present moment, audience and speaker that no two people are likely to share an idiolect and even that at no two times is an individual likely to have the same idiolect. Thus those linguists and philosophers are mistaken who took themselves to be characterizing some shared and stable body of information sufficient unto communication.50

George's line is to argue that Davidson is attempting to support the thesis of "radical semantic instability"—the major support for this thesis being "the view that intentions can determine meaning." Contra Davidson, George argues that Davidson's examples (the malaprop case and others) are not instances of a speaker's idiolect undergoing modification in the course of an exchange, but rather of changes in our own beliefs about the speaker's idiolect.

I believe, however, that George has missed the point somewhat of Davidson's argument. Nowhere does Davidson argue that the speaker's production of the malaprop gives evidence that there has been a change in his idiolect. I believe that Davidson is more than willing to admit that in these cases and others the competence of both speaker and hearer remain unchanged.

George is correct when he attributes to Davidson the view that theorists are "mistaken" when they take themselves to be characterizing "some shared and stable body of information sufficient unto interpretation"—but Davidson is nowhere concerned with the issue of stability. Furthermore, on my reading of Davidson's argument, he is also willing to admit that speaker and hearer share a kind of competence.
Recall, for example, Davidson's words to the effect that principle (2) of the TPP model—the requirement that first meanings be shared—survives "when understood in [a] rather unusual [way]." Here, then, is what I take Davidson to mean by this claim: What is shared by speaker and hearer is the ability to converge on the correct passing theory from time to time; there is, however, no body of rules or principles that can adequately characterize this ability.
Part II
Assessing Davidson's argument
Introduction

Davidson's argument is in many ways very unnerving. He tells us that

... [a] passing theory really is like a theory at least in this, that it is derived by luck, and wisdom from a private vocabulary and grammar, knowledge of the ways people get their point across, and rules of thumb for figuring out what deviations from the dictionary are most likely. ¹

The picture that emerges is one that shifts the emphasis from shared "interpreting machines" to rough and ready maxims, from mechanisms to luck and pluck, from rules to vaguely articulated personal strategies that help us find a trajectory through language space. The phenomenology of language therefore presents the interpreter with a kind of scientific challenge. The interpreter is compelled to devise a "theory" in which certain sounds or signs made by the speaker are first recognized as linguistic, and then integrated into a framework in which it becomes possible to "make sense" of the speaker's behavior—linguistic and otherwise. In "Communication and Convention" (Davidson, 1984b), Davidson tells us that there is no saying what someone must know who knows a language, "for intuition, luck, and skill must play as essential role here as in devising a new theory in any field ... ."² Yet one wonders how this can be true. For the ability to devise good theories in science is a specialized skill acquired only after many long years of training; it is a rare gift among scientists. Yet even the most feebleminded language-user can devise a correct passing theory from time to time.
Davidson has two big targets in the Epitaphs paper: (1) the notion that convention plays some essential role in language; and (2) the notion of a shared linguistic competence. (In §1 I suggest how these two subjects might be related.)

Davidson wants to argue that convention plays no essential role in natural language. He may, for other purposes wish to generalize this to the thesis that convention plays no essential role in communication *simpliciter*, or that it plays no essential role in communication by means of a language in some more general sense of the term (where this latter notion of a language is defined appropriately), but this does not appear to be the thesis that he is defending in the Epitaphs paper.

Here is the general shape of his argument. He asks us to consider the following phenomenon: \(A\) utters malaprop \(m\) intending thereby to mean that \(p\); his hearer \(B\) comes to believe (correctly) that by uttering \(m\), \(A\) meant that \(p\). Davidson now argues that there is no preexisting convention that "links" \(m\) — or the use of \(m\)—to meaning that \(p\), and that therefore \(B\)'s coming to believe that \(A\) meant that \(p\) was accomplished without any essential appeal to convention. His conclusion?: Convention plays no essential role in communication by means of a language.

I examine this argument in §1 and I suggest that Davidson's skepticism about the role of convention in language does not appear to be well-motivated. I hope to show that the argument of the Epitaphs paper,
even if sound, fails to engage the thesis of the convention-in-language theorist (henceforth, the "conventionalist").

Davidson's second target is the notion of a shared, rule-governed interpretive competence on the part of both speaker and hearer. Davidson argues that the speaker's utterance of the malaprop catches the hearer's competence "off guard," and that the hearer is unable, in any rule-governed way, to forge the necessary link between the sounds made by the speaker and their intended meaning.

Davidson's argument is essentially three-pronged. He argues: (1) There is no such thing as a malaprop-interpreting competence; and because there is no such competence, it certainly cannot be shared; (2) Even if we could adequately describe a rule-governed process by which B interprets the malaprop (which we cannot), it would be of such an idiosyncratic nature that it would be unlikely to be shared by A; (3) This despair over the possibility of any shared malaprop-interpreting competence is extended—via the Ubiquity Hypothesis—to the possibility of any shared language-interpreting competence (where the implied notion of a language is extensional). This last move is essential because we might, for example, object that malaprops and the other interpretive phenomena that Davidson alludes to are very rare or very peculiar and therefore lie outside the proper margins of that phenomenon we call communication by language. To this he responds by saying that "such things happen all the time; in fact, if the conditions are generalized in a natural way, the phenomenon is ubiquitous." What I have called the "Ubiquity Hypothesis"
is essentially Davidson's response to an objection of this kind.

The picture of utterance interpretation that I would like to defend against Davidson's skeptical claims is the following. Speaker A makes an utterance $U$ and interpreter B recovers the speaker's intended meaning in two stages: (1) B's (rule-governed) competence generates for $U$ a range of its possible meanings; (2) B's knowledge of the maxims and principles governing communication by language (the subject of a well-developed theory of pragmatics) together with his general intelligence, allow him to choose correctly one of the range of possible meanings for $U$, in accordance with the speaker's intentions. Davidson is arguing that there is no competence that we can attribute to B on the basis of which B can generate the essential range of meanings for $U$. He does not, in other words, believe that stage (1) is possible for the malaprop—nor, in fact, does he believe that it is possible for utterances in general (the Ubiquity Hypothesis).

George$^4$ has highlighted an interesting distinction between the rules that help us to determine what a speaker meant, given that the speaker was speaking (roughly) English, and rules that help us to determine which rules the speaker was trying to conform himself to (i.e. rules for identifying the speaker's idiolect). I am assuming that Davidson's skepticism is directed against rules of the former kind. The possibility that the idiolects of speaker and hearer may be quite different initially only adds to the difficulty of accounting for the hearer's interpretive competence via the attribution of rules known to both
speaker and hearer in advance of the occasions of utterance.

In this chapter I will argue that there is good evidence that there does in fact exist a shared malaprop-interpreting competence, and that this competence shares many features with the competence that we use to interpret utterances in general.

In §2, I will counterpoise Davidson's notion of our linguistic competence with that of the mechanistic program championed by Noam Chomsky and others of the generative grammar school, and I will defend the Chomskyan framework of inquiry against Davidson's skeptical claims. In §3, I will argue that this framework can accommodate interpretive phenomena which do not strictly fall within its theoretical purview. There is, in other words, a window of some value into which non-mechanistic accounts of interpretation can fall.

In §4, I will argue that the Ubiquity Hypothesis (Part I, §4) is false.

§1 Some notes on convention

Davidson concludes the Epitaphs paper by writing:

In linguistic communication nothing corresponds to a linguistic competence as often described: that is, as
Principles (1) and (2) survive when understood in rather unusual ways, but principle (3) cannot stand. ... We must give up the idea of a clearly defined structure which language-users acquire and then apply to cases. And we should try again to say how convention in any important sense is involved in language; or, as I think, we should give up the attempt to illuminate how we communicate by appeal to conventions.  

Principle (3), as the reader will recall, is the requirement that first meanings be governed by learned conventions or regularities. I will return to the question of how principles (1) and (2) survive, but my present concern will be to explore how the argument of the Epitaphs paper addresses the role of convention in language. In particular, I would like to explore how Davidson's argument in the Epitaphs paper might settle the ongoing debate between Davidson and those philosophers who claim that conventions are somehow essential to the existence of communication by language.  

It is a rather unfortunate feature of Davidson's argument that he mentions only one plausible candidate for a convention governing the use of language:  

Because each party has... a shared theory and knows that others share his theory, and knows that others know he knows, (etc.), some would say that the knowledge or the abilities that constitute the theory may be called conventions. 

I have already argued (Part I, §2) that this shared and symmetric knowledge does not in and of itself constitute a convention; it may be one of a number of criteria constituting a given convention governing the
behavior of speaker and hearer, but as it stands it is incomplete. If this is the only notion of convention that Davidson is willing to countenance, then we can see how he is drawn to the conclusion that convention plays no essential role in linguistic communication. For if, as Davidson argues, there simply is no shared theory, then each party to the communicative exchange cannot know that the others share his theory, nor can the others know that he knows, etc. There may, of course, be independent grounds for arguing that the shared and symmetric knowledge alluded to does not play an essential role in linguistic communication, but I will not explore this possibility here.

I believe that it is a safe assumption, however, that Davidson has a more general conception of convention in mind. The problem for the interpreter of Davidson's thought, and indeed for the proponents of convention-in-language theories, is to specify the notion of convention that might be relevant to understanding the "social aspects" of linguistic meaning.

In his paper "Communication and Convention," Davidson hints at what this essential notion of convention in language might be. There he tells us that "It is mainly in making the connection, or connections, between linguistic meaning and human attitudes and acts described in non-linguistic terms that convention is asked to do its work." Here is what I take Davidson to mean. Linguistic meaning often appears to us in two different guises. On the one hand, we have the representations of sentence and word meaning posited by the linguist. The linguist's theory
of these representations is ultimately a kind of psychological theory: it is a theory about what psychological states we must attribute to a human being (or some other appropriately constituted entity) in virtue of which he is a competent language-user. The linguist thus presents a theory, for example, of how "smokeless ashtray" means smokeless ashtray for a normally constituted English speaker, as though the business of assigning meanings to expressions were independent of assumptions about the behavior—past, present, and future—of other people. On the other hand, it appears that my taking the words "smokeless ashtray" to mean smokeless ashtray somehow reflects my conformity to the behavior—linguistic and otherwise—of other people in general, and other language-users in particular. Here we are cleaving closely to the Wittgensteinian intuition that the meaning of a word is determined by its use in a language, or, more broadly, its use in a community of language-users.

How, then, do we represent the connection between these two aspects of linguistic meaning: between meaning as constituted by the rules of a language, and meaning as constituted by certain kinds of intentional activity? The convention-in-language theorist believes that the connection is mediated by the notion of a convention, in a sense that I will make clearer below.

When we attempt to zero in on the relevant notion of a convention we have an embarrassment of riches. There appear to be at first blush any number of conventions at work in language. Few people will deny, for example, that the connection between the sound pattern or shape of given
word and its meaning is in some sense conventional.\textsuperscript{9} We see no reason why "apple" could not have meant orange, and vice versa. When we were learning our first language we tried to make our linguistic behavior conform to the linguistic behavior of others, to adhere to the preexisting conventions of language use. We would not, for example, have been accepted into the community of language-users if we had systematically used the word "yes" to mean no.

I do not believe, however, that Davidson has any real issue with this notion of convention. Beyond the arbitrary, or, if you wish, conventional connection between the sound patterns of words and their meaning, Davidson asks, is there some notion of convention that is necessary to the existence of communication by language? Is there some essential regularity that communication by language could not possibly do without? I put the question this way following Davidson's suggestion that "If there is to be a convention in Lewis's sense (or in any sense, I would say), then something must be seen to repeat or recur over time."\textsuperscript{10} Here we must always keep in mind the distinction between those cases in which the existence of a convention explains the observed regularity, and those cases in which the regularity simply constitutes—as it were by default—a given convention. An example of the former would be a game of chess, and an example of the latter would be the dance of the bumble bee.

What we are after, then, is some notion of convention—or its equivalent—that is essential to any adequate description of linguistic phenomena. I stress these words because it is certainly possible to
disguise the role of some candidate convention by rewriting our
description of the phenomena in such a way that we make no explicit use
of the word "convention."

The linguistic phenomenon that most interests Davidson is, of
course, communication. But even here we must tread carefully. For
Davidson sometimes argues as if the notion of convention were
unessential to an adequate description of existing linguistic
practice—communication between the speakers of a given natural
language, say. At other times he argues as if the notion of convention
were unessential to an adequate description of communication by language
simpliciter.

Knowledge of the conventions of language is thus a practical
crutch to interpretation, a crutch we cannot in practice afford
to do without—but 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.¹¹

We can, I suppose, imagine a kind of "proto-language" with no discernible
or established syntax; where the referents of various sounds are
established by pointing; and where, by means of an artful combination of
sound-making and miming, we can successfully "communicate" with our
fellow proto-language-users. We can furthermore assume that the
propositions so communicated are just as often false as they are true.
Suppose that a fellow proto-language-user succeeds in expressing the
thought, "Your cave is on fire." If, fearing the worst, I run to my cave in
time to put out the fire and salvage my belongings, has my informant not
succeeded in communicating with me "linguistically"? Here we have a plausible "language" without any of the contentious conventions that Davidson so much objects to, and in which verbal or linguistic communication is still possible. But to proceed in this way is not, I believe, to engage the arguments of Lewis and others who contend that some notion of convention is essential to an adequate description of the way we in fact use language to communicate with one another.

I have assumed that the structure of Davidson's argument in the Epitaphs paper is essentially as follows: A's production of the malaprop constitutes a rupture or breaking of some convention C; B nevertheless successfully interprets A; therefore, Davidson argues, convention C is not strictly relevant to the phenomenon of interpretation. From this point on Davidson can argue that those philosophers and linguists are mistaken who take convention C to be inextricably wound up with the possibility of communication: if this is what a language is, then there is no such thing as a language.

The most obvious problem with this reconstruction of Davidson's argument is identifying which convention or conventions Davidson has in mind. What will satisfy our demand for a convention C that has the following properties: (1) convention C is broken by A's utterance of the malaprop (or, less plausibly, by B's successful interpretation of the utterance); (2) the convention in question connects linguistic meaning with human attitudes and acts described in non-linguistic terms; and (3) linguists and philosophers believe that convention C is inextricably wound
up with the possibility of communication?

Davidson wants to say to say that it is somehow unconventional for the speaker to express the thought


by using the expression


One interesting feature of [6] is that the speaker, in uttering it, does not intend the hearer to take his meaning to be the literal meaning of [6]. We see something like this "nonliteral strategy" at work in other ways in language. The speaker may, for example, intend to express the thought, "George is anything but a nice fellow" by uttering the sentence, "George is a fine fellow" with an ironic tone of voice. It is generally supposed that there exist pragmatic rules or principles that govern the recovery of nonliteral meanings from utterances such as these. We might, on this basis, assume that when Davidson denies the role of convention in language, he means to deny the existence of any pragmatic rules or principles that might yield the speaker's intended meaning [5] from the speaker's utterance of [6]. I believe, however, that this would be a misinterpretation of Davidson's argument. He is clearly after bigger game.

The speaker's utterance of [6] is set apart from such "standard" cases of nonliteral speech as irony, metaphor, hyperbole, etc. in another most peculiar way: unlike these other cases, here the speaker makes an
utterance and intends that he be interpreted \textit{literally}. Of course the speaker does not intend that the sentence he actually succeeds in producing be interpreted literally, but rather that he be \textit{taken} by his hearer to be speaking literally.\textsuperscript{14} It is here, I believe, that the rupture of the purported convention lies.

But what can this convention be? The rupture of the convention consists in the fact that the speaker utters [6] intending that the hearer should take him to be speaking literally, while at the same time intending that the hearer should take him to mean [5]. The convention in question apparently governs what we can and cannot intend to mean literally by using certain expressions of the language. Pursuing this line of inquiry, we might suppose that one of the theorems of this convention of literal meaning would be something like the following:

[7] \textbf{Theorem M:} One cannot utter [6] unless one intends to mean \texttt{[What a nice derangement of epithets]} thereby,

and that it is \textit{this} principle that the speaker violates when he utters [6] with the intention that his utterance be taken to mean \texttt{[What a nice arrangement of epithets]}. Theorem M suggests that our convention-in-language theory might have something like the following structure: On the interpretation side, the hearer has before him an infinitely long list that enumerates all of the expressions of his language. Next to each expression of the language there is another list that gives that expression's range of possible meanings:

[8] \begin{tabular}{ll}
Expression & Meaning \\
\end{tabular}
\[ x_1 \quad p_{11}, p_{12}, p_{13}, \ldots \\
 x_2 \quad p_{21}, p_{22}, p_{23}, \ldots \\
 x_3 \quad p_{31}, p_{32}, p_{33}, \ldots \\
 \text{etc.} \]

The range of possible meanings is, of course, determined by the given convention. On the production side, the speaker has a similar list that he can likewise refer to, only in this case, the roles of expression and meaning are reversed:

[9] \[ \begin{align*}
\text{Meaning:} & \\
p_1 & \quad x_{11}, x_{12}, x_{13}, \ldots \\
p_2 & \quad x_{21}, x_{22}, x_{23}, \ldots \\
p_3 & \quad x_{31}, x_{32}, x_{33}, \ldots \\
\text{etc.} & 
\end{align*} \]

We now come to a lacuna in Davidson's argument, for although our conventionalist theory specifies the range of meanings for any expression of the language, there is nothing to imply that the rupture of this convention will ipso facto disrupt communication.

We might attempt to fill this gap in Davidson's argument as follows. Imagine a simple language where the range of meanings of an expression is established by explicit convention, on the model of [8]. Consider, for example, the Revolutionary War convention, "One, if by land; two, if by sea." Paul Revere and his "signaler" agree beforehand that if the signaler hangs one lantern in the North Church tower, Revere is to take this to mean that the British troops will come by land; and if he hangs two
lanterns, that they will come by sea. An interesting feature of this convention\textsuperscript{15} is that it allows for the possibility of Gricean implicature. Thus if the signaler hangs one lantern and one candle in the tower, Paul Revere might take this to imply, for example, that most of the British troops will come by land, but that a small fraction (represented by the candle) will come by sea.\textsuperscript{16} What this convention does not appear to make allowance for, however, is the possibility of malapropism: it is simply not a feature of this kind of convention that the signaler can hang one lantern in the tower and be taken to mean what he would otherwise mean by hanging two (and vice versa). We might therefore take Davidson to argue as follows: Whatever kinds of conventions there are that philosophers generally suppose govern language, \textit{this} cannot be one of them. If language were governed by a convention of this kind, then it would not be possible to utter a malaprop and be understood. We can therefore make room for the phenomenon of malapropism in language only by dropping the idea that the sound/meaning link is conventional.

I find this kind of argument less than convincing for several reasons. First, I disagree that the convention established between Paul Revere and his signaler does not make room for the phenomenon of interpretable malapropism. If, for example, Paul Revere was not miles away from the coast atop his horse, but was instead allowed to see for himself the comings and goings of the ships in Boston Harbor, and if he could on this basis conclude that there were no British troops coming by sea, then he would have good grounds for inferring that by hanging out two lanterns his signaler had "misspoken." What is missing from this account
of the communicative exchange between "speaker" and "hearer" is a clear counterpart of the notion of utterance context. Paul Revere cannot interpret his signaler to be misspeaking if the only knowledge or competence that he can apply to the communicative exchange is his knowledge of the conventionally established meanings of one or two lanterns in the North Church tower. But neither can I successfully interpret Mrs. Malaprop's utterance if I chance upon it written on some small piece of paper in the middle of an open field; for I have not thereby been given the means to recognize it as being a malaprop.

Likewise the conventionalist might respond that the convention in question governs "standard" or "core" cases of interpretation, and that competent speakers of the language use other means to arrive at the intended meanings of malaprops. The issue here, of course, is how much the conventionalist weakens his theory by granting the possibility of these other methods of interpretation. Davidson may wish to respond, for example, by asking why these "other means" cannot govern the interpretation of utterances generally: How can we, in one breath, claim that conventions are essential to an adequate description of communication by language, and in the next say that there are cases where the notion of a convention does not apply? But contrary to this objection, I do not see how the conventionalist weakens his theory by making allowances for non-standard or "unconventional" cases of communication by language, unless, of course, he wishes to be dogmatic about the descriptive adequacy of his theory.
In the following section I will argue that there is good evidence that something like a malaprop-interpreting competence exists for speakers of a natural language, and that this competence appears to share many features with the competence that linguists normally attribute to language-users. The competence of which I speak is, of course, grounded in some grammar $\Gamma$ that is shared by both speaker and hearer. The role of this competence is essentially that of "mapping" the utterance [6] to the target utterance "What a nice arrangement of epithets," so that [6] receives the interpretation that the hearer would, under normal conditions, assign to the target utterance. There is, I believe, good reason for assuming that the conventionalist must also avail himself of the linguist's notion of a shared competence or grammar. Stephen Schiffer suggests something along these lines in his book Remnants of Meaning. Acting as an apostle for what he calls "Intention-Based Semantics," or more succinctly, "IBS" theories of meaning, Schiffer argues that Gricean accounts of expression meaning must ultimately be wedded to what he calls "psycholinguistic" theories of expression meaning:

The IBS theorist must stand ready to explain—or at any rate to throw light on—how speakers of $L$ can correlate the sentences of $L$ with their meanings in $L$, and it is here that we should expect the notion of a grammar of $L$ to play its explanatory role. If $L$ is used in [a population] $G$, then some grammar $\Gamma$ of $L$ explains the ability of members of $G$ to correlate the sentences of $L$ with their meanings in $L$; this is what makes $\Gamma$ stand in the distinguished relation to $G$ that I, on behalf of the IBS theorist am trying to explicate.

The role of the grammar $\Gamma$ is essentially that of showing how, using
"finite means," the language-user can correlate the infinite set of possible expressions in [8] with their meanings. What I am suggesting in this paper is that it is this same grammar I that "correlates" Mrs. Malaprop's utterance of "What a nice derangement of epitaphs" to her target utterance "What a nice arrangement of epithets." Davidson's denial of the existence not only of a malapop-interpreting competence, but also of a general linguistic competence may therefore be seen as a way of denying the legitimacy of this kind of move on the part of the IBS theorist.

Where does this leave Davidson's argument? If contrary to what I have urged, we take Davidson to be arguing against the thesis that no communication—linguistic or otherwise—is possible without essential appeal to convention, then Davidson is arguing against a straw man. I have found no evidence that this is a widely held belief among the "linguists and philosophers" that Davidson so often alludes to. I have also suggested a way that we might supplement the account of expression meaning given by the IBS theorist so as to make the interpretation of Mrs. Malaprop's utterance dependent on the conventional link between the expression "What a nice arrangement of epithets" and its possible range of meanings.

Someone might object that on my view, communication is a matter of faith and luck: faith, because the speaker must simply assume that the hearer will possess the right interpretive "equipment"; and luck, because fortunately for the speaker, his assumption will often turn out to be correct. I might myself be accused of rejecting the role of convention in language—stripping convention of its "social" or "cultural" aspects by
claiming, for example, that the essential convention consists in nothing more than the sharing of some mechanism. But I am making no such claim. I do not believe that communication by language is all a matter of language-users colliding in the void; on the contrary, I firmly endorse Lewis's view that "It is a platitude—something only a philosopher would dream of denying—that there are conventions of language, although we do not find it easy to say what those conventions are." 19 Even the most dyed-in-wool mechanist must, on pain of negligence, demand some account of how external conditions affect the maturation and ultimate character of the language faculty. 20 The salient question for my view is how the notion of a convention might illuminate the process or processes by which the representations of meaning posited by the linguist acquire their contents. To say that this process must make essential appeal to interactions with one's community of language-users is one thing; to say that this process involves the irreducible notion of a convention is quite another.

Consider the following scenario: Person A produces an utterance (in context C) and person B intuits (or understands) the meaning of this utterance. If, in making this utterance, person A produces a token of sentence S, we say that person B intuits the meaning of sentence S (as spoken by A in context C). Given the mechanistic account of interpretation that I will defend in this chapter, I might tentatively venture the following "definition" of what it means to "intuit" the meaning of S:

[10] Person B intuited the meaning of S (as uttered by A in context C)
    =_df_ a representation R of the meaning of S (as uttered by A in context
occurred in person B.

We note in passing that the usual competence/performance distinction may play a role here as elsewhere. For example, person B might have intuited the meaning of S as per our definition, and yet not behaved as if he had so intuited it, etc. We should, of course, distinguish intuited the meaning of a sentence S from simply knowing the meaning of a sentence S. I assume that we can be said to know the meaning of a sentence in the same sense that we know that it is a sentence of our language. The biggest problem with our definition is that it still contains the unanalyzed term "the meaning of S (as uttered by A in context C)," so that we are still tempted to ask, What is this meaning? We can attempt to remedy things by replacing the definiens with the following:

[11] a representation R with property P occurred in person B,

where we say that the representation in question was induced in some purely mechanistic fashion by the utterance of S (by A in C). We can assume, perhaps, that the property in question will be specified by our theory of semantics, and that it is in virtue of having this property that R is a representation of the meaning of S rather than of some other sentence S*.

This is fine as far as it goes, but it doesn't go very far. We are left wondering, among other things, what has happened to our commonsense notion of sentence or word meaning. And we still have not explained what it is that hinges on our modeling the representation of S as having some property P rather than some other property P*. Here is where
the connection with the world, or with the society of language-users, becomes manifest.

We might, for example, assume that the representation $R$ of sentence $S$ has certain "causal powers" in virtue of its having property $P$ (if it had a different property, call it $P^*$, it would have different causal powers). Consider, for example, the sentence

[12] Please bear with me, and at the count of three raise your right hands in the air.

I have found that by uttering sentence [12] under the right circumstances, people's right hands have consistently popped into the air after I counted to three. We can explain this fact by assuming that a representation $R$ of the meaning of sentence [12] "occurs" (as per our definition [10]) in each of my hearers; that this representation $R$ has some theory-specified property $P$; and that it is in virtue of having this property that $R$ can "cause" the observed behavior.

Representations of sentence meaning need not always cause or induce overt behavior, of course. If someone utters the sentence "Dan Quayle is a troglodyte" within hearing, the occurrence of a representation of the meaning of this sentence in my brain may simply have a role in bringing me to believe that Dan Quayle is a troglodyte, and this new belief of mine may never manifest itself as part of my repertoire of external behaviors. Which behaviors, beliefs, or other representations the occurrence of some $R$ induces in me depends in great measure on my
cognitive state, my environment, and other factors. The so-called causal powers of some representation may simply be dissipated, or stored, etc.

All this talk of "causal powers" is one way of getting at what we mean when we speak of the "content" of a particular representation of meaning. On the account I am sketching here, to give the content of a representation of meaning is to specify the complex of that representation's connections—actual or potential—with the world. These connections are mediated by the internal and external behavior of the person harboring these representations. If I say to my colleague, "My father is an effete man," and my colleague responds, "Oh he's a snob, is he?", I can surmise that my colleague's representation of the meaning of the word effete does not have the property that will allow his use of the word, or his reactions to the use of this word, to covary in the appropriate way with the behavior—linguistic and otherwise—of those people who are said to know the meaning of the word. I assume that my colleague's inner representation of the meaning of the word effete had some role in determining that he would respond to me by saying, "Oh he's a snob, is he?" rather than by saying, "Oh he's worn out, is he?" On the linguistic end of things, the semanticist might say that my colleague's representation of the word has the features [-value] (to say that someone is effete is to say something bad about the person), and [+status], whereas my representation of the word has only the former feature. 21

We have thus moved out of the cloistered domain of linguistic representations and into the world. But where are the conventions which
are to act as the glue binding the linguistic and social aspects of meaning together? I find it hard to say what these conventions might be, and for this very reason, I find it impossible to rule out a priori the possibility that some notion of convention will play an essential role in any adequate description of meaning in language. I believe that the essential convention or conventions must ultimately be grounded in the notion of a shared grammar, in the sense that any adequate description of these purported conventions must make essential appeal to our shared linguistic competence. But to be a realist about underlying mechanisms is not ipso facto to be a reductionist about convention.

At the beginning of this section I pointed out the fact that Davidson concludes the Epitaphs paper by saying that "Principles (1) and (2) [of the TPP model] survive when understood in rather unusual ways, but principle (3) cannot stand..." \(^{22}\) Here is what I think Davidson means by this. Principle (3) obviously cannot stand because it is the requirement that first meanings be governed by learned conventions or regularities. But how do principles (1) and (2) manage to survive? Principle (1) survives because, according to Davidson, "We cannot describe what an interpreter can do except by appeal to a recursive theory of a certain sort," and, furthermore, "It does not add anything to the thesis to say that if the theory does correctly describe the competence of the interpreter, some mechanism in the interpreter must correspond to the theory."\(^{23}\) What Davidson is suggesting, in other words, is that he does not propose to be a realist about recursive theories of meaning. Principle (2) survives because Davidson does not deny that first meanings continue to be shared:
speaker and interpreter do, after all, come to share their passing theories. Nothing more, I believe, is implied by Davidson's claim that principles (1) and (2) survive.

In the sections that follow, I will defend the mechanist's account of our shared linguistic competence against Davidson's sceptical claims.

§2 Davidson versus the Chomskyan mechanist

Davidson's arguing that there is no such thing as a shared interpretive competence may appear rather anomalous to some readers. It is. How could the man who championed a Tarski-style theory of truth for a theory of meaning in natural language claim that there is no such competence? Another commentator on Davidson's Epitaphs paper, Ian Hacking, has also noted this element of retraction.24 Hacking wonders what happens to true-in-L in the "no-language story," and he rationalizes Davidson's apparent break with the past in the following way:

The interpreter can be modeled as interacting with his speaker by a Tarski-style theory, or so we were told. But now we have the picture of the interpreter going through a whole series of constantly modified Ls very quickly. Interestingly, Davidson has from the start always had a slight tendency to elide the L in T-sentences. ...I am inclined to say that the Ls (often so happily elided) are not languages at all, in any common sense of the world [sic.], and one could still keep the Ls even if one said that there is no such thing as a language.25
For reasons that may already be apparent (cf. §1), I believe that Davidson is more than willing to drop his countenancing of Ls altogether, or at least to adopt a non-realist perspective toward them. When Davidson claims that malaprops and other such phenomena threaten standard descriptions of linguistic competence, he acknowledges that these may include descriptions for which he himself is responsible.26

Neither do I believe that this retraction has no precedent. The careful reader of Davidson would have already noted a kind of "ambivalence" toward descriptive theories of meaning in language. In his introduction to "Radical Interpretation,"27 for example, he writes:

Kurt utters the words 'Es regnet' and under the right conditions we know that he has said that it is raining. Having identified his utterance as intentional and linguistic, we are able to go on and interpret his words: we can say what his words, on that occasion, meant. What could we know that would enable us to do this? How could we come to know it? The first of these questions is not the same what we do know that enables us to interpret the words of others. For there may easily be something we could know and don't, knowledge of which would suffice for interpretation, while on the other hand it is not altogether obvious that there is anything we actually know which plays an essential role in interpretation.28

The problem, as Davidson sees it, is that "interpreting an agent's intentions, his beliefs and his words are parts of a single project, no part of which can be assumed to be complete before the rest is."29 He therefore despairs of our ability to say anything interesting about the phenomenon of meaning in language independently of our being able to say something significant about an agent's intentions and beliefs. In the
Epitaphs paper, Davidson simply takes his qualms about a theory of meaning one step further.

In this section I will argue that such phenomena as our ability to interpret malaprops do not motivate the kind of skepticism that Davidson now appears to endorse. The kind of linguistic competence described by Chomsky and other linguists of the generative grammar school appears to accommodate many of the phenomena which motivate Davidson's skepticism.

Chomsky's notion of linguistic competence is perhaps the centerpiece of his linguistic theory. According to Chomsky, to give an account of a speaker's competence is simply to give an account of his knowledge of the language—of his "linguistic capacity." He distinguishes at least two kinds of competence: grammatical competence and pragmatic competence,\textsuperscript{30} a distinction which I will assume is familiar to most readers. This rough taxonomy will certainly not be of much use to the theorist unless he can further specify the kind of knowledge that constitutes the speaker's "knowledge of the language"; how the speaker accesses this knowledge; what constitutes proper grounds for attributing this knowledge to the speaker; and a myriad of other questions that I will not address here. As current theory has it, a speaker manifests his linguistic competence when, for example, he makes grammaticality judgments, or when he successfully orders a pizza, or when he writes a history of the French Revolution. The linguist's theory of the speaker's competence is a pliant repository of everything that the linguist can
meaningfully say about the speaker's knowledge of his language within a given scientific framework.

Apart from serving as that vague signifier of our linguistic knowledge, the notion of competence also plays a slightly more polemical role as foil to the notion of performance. Thus the speaker's linguistic errors are attributed to a lapse in performance, rather than to an imperfect knowledge of his own language. The notion of performance does not stop there, however. A well-articulated theory of performance will explain, among other things, how the speaker uses his knowledge of the language to produce utterances and construct interpretations in concrete situations. Chomsky makes it clear, for example, that a generative grammar is not in itself a model for a speaker or hearer:

When we say that a sentence has a certain derivation with respect to a particular generative grammar, we say nothing about how the speaker or hearer might proceed, in some practical or efficient way to construct a derivation. These questions belong to the theory of language use—the theory of performance.31

We can expect, therefore, that a model of the speaker's performance will incorporate inter alia a model of speech perception, as well as a model of speech production. Although many of these assumptions have been hotly debated, I will, for the sake of discussion, assume that they are essentially correct.

In very broad outline, Chomsky proposes to develop his core theory of syntax—what he calls Universal Grammar, or UG—in abstraction
from any concrete occasion of utterance. He gains access to this core grammatical competence by studying speaker judgments of well-formedness, judgments of co-reference and anaphora, judgments of sentence meaning and constituent structure, etc. The interpreter's ability to recover the intended meaning of a malaprop does not ex hypothesi present any evidence that might bear on the structure of this core grammatical competence. In careful stages, the theorist builds outward from the core. He uses his theory of grammatical competence to build his theory of pragmatic competence. At this stage of theory construction, he might proceed by attempting to isolate the relevant parameters of an utterance context, and by observing how utterance meaning changes when these candidate parameters are changed in a controlled way. Taken together, the linguist's theories of our grammatical and pragmatic competence constitute what I will call an account of our overall basic linguistic competence.32

In the Epitaphs paper, however, Davidson does not appear to share Chomsky's optimism or enthusiasm for this kind of research program. He believes that the language theorist is caught between the horns of a dilemma: if he does not give a direct account of our ability to interpret malaprops and the like, then his theory of our linguistic competence is incomplete; if, on the other hand, he does try to explain our ability to interpret malaprops, he is no longer studying language as such, for the kind of "competence" that he is attempting to characterize is then no longer exclusively linguistic. This special competence is not different in kind from the competence that one generally needs to find one's way
around in the world. It includes a measure of "wit, luck, and wisdom," together with a "knowledge of the ways people get their point across, and rules of thumb for figuring out what deviations from the dictionary are most likely." Davidson despairs of the linguist's ability to give an account of a competence that he believes is not governed by rules or conventions that are known to both speaker and hearer in advance of the occasions of interpretation.

Let us attempt to draw the battle lines a little more carefully. We can best tease out the differences between these two schools of thought by looking at an example. For the sake of concreteness, let us suppose that the speaker utters:

[13] Familiarity breeds attempt

and the hearer (correctly) interprets him to mean:


Davidson finds it easy enough to explain the hearer's ability to interpret the speaker in the way that the speaker intends:

... the hearer realizes 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.
This reconstruction of how the interpreter recovers the speaker's intended meaning is certainly one that the Chomskyan linguist might endorse. The account is also somewhat schematic, for it tells us nothing about what constitutes a "standard" interpretation. Nor does it tell us how the similarity in sound between the malaprop and the intended utterance tips the speaker off as to the malaprop's correct interpretation.

Let us pause to consider how we might go about filling in some of the missing pieces. We assume that the ideal interpreter possesses a basic, or "core" linguistic competence: a general knowledge of the ways in which words and phrases can be combined to form sentences, and a knowledge of what these sentences mean. We agree with Davidson that upon hearing the speaker utter sentence [13], the hearer gives the sentence its "standard," or literal interpretation. As Davidson points out, the hearer very quickly realizes, or comes to believe, that the speaker's words cannot be taken literally in the given context.

It is at this point that our account becomes rather more anecdotal. The hearer's realization that the speaker could not have meant his words literally is a kind of "error message" that sets off the search for an alternative interpretive strategy. In the absence of any evidence that the speaker was speaking humorously or ironically or sarcastically, etc., the interpreter might simply venture the hypothesis that the speaker misspoke—that he did not succeed in saying what he in fact wanted to say. This being the case, there begins on the hearer's part a search through the space of sentence [13]'s phonological neighbors, until at last the hearer
chances upon the speaker's intended utterance, "Familiarity breeds contempt," and its intended meaning (given in [14]).

All that I have said about "error messages" and "searches through the space of phonological neighbors" is, of course, highly conjectural. In the absence of any evidence to the contrary, however, there is no reason to believe that something like this might not be the case. The account which Davidson himself has given, and which I have embellished, certainly has some initial plausibility. If we furthermore reflect on those interpretive occasions when we did not instantaneously recover the speaker's intended utterance, but rather lingered over his chosen words for a time before the penny dropped, we might be more inclined to endorse the account given here. I recall, for example, once having heard a friend complain that after leaving his parasite-ridden dog in the house for a few days, he was compelled to "irrigate the whole place." I remember puzzling over what my friend could have meant by his words. Keying on the water theme suggested by the word irrigate, I first assumed that the dog's parasites had destroyed the control of its bladder, and that my friend was therefore obliged to wash the floors of his house. This chain of reasoning could not, however, mitigate the oddity of the word "irrigate" in the given context. I then began running through a list of possible replacement words that ended with the morpheme -ate—e.g. instigate, imitate, etc.—until I finally chanced on the intended word "fumigate."

Davidson is not concerned, however, with the empirical plausibility of our account of malaprop interpretation. For Davidson, even
my very embellished account of the process may in some sense correspond to the truth of the matter. What does concern him is the theorist's ability to give a principled account of the process that we have just described. The search through a space of phonological neighbors is apparently not a procedure that Davidson believes is governed by rules and conventions shared by the speaker and hearer, and learned in advance of the occasion of interpretation. Neither is perhaps the hearer's choice of an "alternative interpretive strategy" when confronted by the absurdity of the malaprop.

But Davidson's skepticism appears to be unwarranted. The analysis of speech error data has revealed many surprising regularities in the way that we produce slips of the tongue. (I am here taking the malaprop to be a very special subclass of this larger category.) For example, Fromkin cites a study by Boomer & Laver who conclude that "The origin syllable and the target syllable of a slip are metrically similar, in that both are salient (stressed) or both are weak (unstressed) ... ."37 Fromkin also notes that "when vowels or syllables or parts of syllables or whole words are substituted or transposed, there is no change in the stress pattern or contour of the sentence."38 In other research, Nooteboom has found that "a mistakenly selected word always or nearly always belongs to the same word class as the intended word ... ."39 Other speech error studies, that I will not discuss here, have uncovered many striking regularities that involve smaller units of articulation, like segments and clusters.40

The implications of this research for our model of malaprop
interpretation are clear: if the speech errors that we produce are not entirely random, but are rather tightly constrained by a number of general principles that research might uncover, then our search through the space of candidate replacements might be enormously simplified. This research, generally conducted by assuming an underlying generative model of grammar, suggests that the mechanistic description of our linguistic competence may well be able to accommodate the ability that Davidson finds to be the cause of so much skepticism. What I am here suggesting, in other words, is that the hearer recovers the speaker's intended meaning by first recovering the speaker's target sentence ("Familiarity breeds contempt") from the sentence the speaker utters ("Familiarity breeds attempt") by a kind of "mechanical empathy," and that it is the shared knowledge of the rules of their language, together with a knowledge of the ways in which these rules are likely to be broken, that makes this empathy possible. Nothing in all of this rules out the possibility that other kinds of knowledge or competence may be involved in the process—a possibility which I explore in §3.

We might ask to what extent the Chomskyan framework we are discussing here corresponds to the TPP and PT models discussed earlier. The TPP model of our linguistic competence posits three principles that govern first meanings in language. Briefly put, these are the requirements that first meanings be (1) systematic, (2) shared, and (3) prepared. As for requirement (1), part of the burden of UG is to show how the "semantic" and other idiosyncratic properties of words conspire with sentence structure to determine sentence meaning. This relation between form and
meaning is systematic, and is mediated by a number of general rules and principles. Literal meanings are further "processed" to yield utterance meanings. This "interpreting machine" is acquired in advance of the occasions of utterance and, as required by principle (3), is conventional in character (when this last proviso is understood in a highly attenuated sense; cf. §1). As for requirement (2), in the framework under discussion it is an idealization to say that both speaker and hearer will share this basic grammatical (and pragmatic) competence. In a heterogeneous speech community this condition may never obtain: there may always be large or small differences between the idiolects of speaker and hearer. It is a sufficient condition for communication that speaker and hearer share the same "interpreting machine," but it is by no means a necessary condition. We can, however, make sense of condition (2) in the Chomskyan framework by asserting that members of the same "language community" (as identified by certain gross features of speech, for example) tend to have grammatical competencies that look pretty much the same, when this "sameness" is measured by some theory-internal metric. Thus, for example, any random subset of native speakers of American English will possess grammars that share many identifying, or "key," features. On the view that I am developing here, speaker and interpreter share a kind of second-order ability—"second-order" in the sense that for any one language-user, many different instantiations of his syntax and lexicon will constitute the same competence with respect to some set of malaprops or other misspeakings. Suppose, for example, that A knows the meaning of the word "palimpsest," whereas B does not. If this is the only difference between the languages attributable to A and B, we can expect
that they will coincide in their interpretation of the malaprop who utters, "Ida's sick, so we sent her a card expressing our symphonies," as the interpretation of this utterance does not involve the meaning of the word "palimpsest" in any direct way. In this case I would say that with respect to this malaprop, and a host of other misspeakings that we might enumerate, A and B essentially share the same malaprop-interpreting competence. It is this possibility that Davidson wishes to guard against when he argues that there is "no learnable common core of consistent behavior." 42

It is not difficult to find the correspondences between the PT model and the Chomskyan framework that I sketched above. Consider first the prior theory. In the PT model, the hearer's prior theory describes how he is prepared in advance to interpret the speaker, whereas the speaker's prior theory describes what he (the speaker) believes the hearer's prior theory to be. In the Chomskyan framework, the hearer's prior theory corresponds to the hearer's basic linguistic competence supplemented whatever knowledge the hearer may have about the speaker's linguistic idiosyncrasies (also leaving some room for the intervention of non-linguistic faculties of mind in the interpretive process). The speaker's prior theory is not really all that interesting, for in effect it expresses the speaker's "best guess" at what the hearer's basic linguistic competence might be. This guess is conditioned by a number of extra-theoretic factors—among them, the country or language community that the speaker believes himself to be in, the speaker's prior knowledge of the hearer's linguistic behavior, etc. It is, of course, typical for the
speaker simply to assume that the hearer's prior theory is exactly like his own. The dogged persistence of this assumption can be seen in those rude American tourists who believe that by raising their voices and making pointed gestures they can get their monolingual French-speaking waiters to understand English.

In the PT model, the hearer's passing theory is how the hearer interprets the speaker's utterance. In the case of Mrs. Malaprop, the hearer's passing theory consists of the following bit of propositional knowledge and little more:

[15] "What a nice derangement of epitaphs" means [[What a nice arrangement of epithets]].

Here we must sharply distinguish between what the hearer interprets Mrs. Malaprop to mean (this is given on the right hand side of [15]), and how the hearer arrives at this interpretation. The latter is given by the linguist's theory of the hearer's basic linguistic competence, supplemented by whatever special knowledge of the speaker the linguist can attribute to the hearer.

On the speaker's side, the passing theory is how he (the speaker) wishes the hearer to interpret his utterance. For reasons discussed above, the speaker's passing theory is of no great theoretical interest.

We are now in a position to interpret Davidson's claims about prior and passing theories in a general Chomskyan framework. Davidson
asserts, for example, that as speaker and hearer talk, "their prior theories become more alike; so do their passing theories." We must here proceed carefully, however, for a little reflection will show that each of both speaker and hearer has two prior theories (and two passing theories). The hearer, for example, has a prior theory that describes how he is prepared in advance to interpret the speaker; but the former hearer may in turn become the new speaker, and this new speaker will have a prior theory that describes how he believes the new hearer will interpret his utterances. I will assume, therefore, that when Davidson uses the plural form, "prior theories," as he does here, he is referring to the prior theories of hearer and speaker on the occasion of a particular utterance.

How comes it, then, that prior theories become more alike? The answer to this question is quite straightforward: as hearer and speaker exchange words, they learn more and more about each other's idiolects. The hearer's prior theory comes to reflect not only his own basic linguistic competence (which we assume remains constant throughout the exchange), but also his knowledge of the speaker's special speech patterns and lexical idiosyncrasies. The speaker also grows more confident that he will be interpreted as he intends; or, if not, he adjusts his speech to fail in line more closely with the hearer's expectations.

A similar story goes for the passing theory: the more that speaker and hearer know about each other's idiolects, the more likely it is that their passing theories will coincide.
This is little more than what Davidson has already told us. The problems begin, however, when we make further inquiries about the interpreter's passing theory. "... [T]he passing theory," Davidson tells us, "cannot in general correspond to the interpreter's linguistic competence." On one rather narrow reading of this passage, the claim is obviously true. A passing theory like that given in [15] cannot, of course, correspond to anybody's basic linguistic competence. As Davidson himself tells us, "its expected field of application is vanishingly small." He persists, however, in calling it a "theory" for the following reasons:

... when a word or phrase temporarily or locally takes over the role of some other word or phrase (as treated in a prior theory, perhaps), the entire burden of that role, with all its implications for logical relations to other words, phrases, and sentences, must be carried along by the passing theory. Someone who grasps the fact that Mrs. Malaprop means 'epithet' when she says 'epitaph' must give 'epithet' all the powers 'epitaph' [sic.] has for many other people. Only a full recursive theory can do justice to these powers.

We can agree with Davidson that only a full recursive theory can do justice to these powers, but we certainly need not agree that we must invoke these powers in order to adequately describe the interpreter's passing theory. As already noted, the interpreter need not "update" his prior theory in order to accommodate the new influx of linguistic information from Mrs. Malaprop. He need only record the fact that on that particular occasion of utterance Mrs. Malaprop meant arrangement by "derangement" and epithet by "epitaph." Exactly what role this kind of explicit knowledge plays in utterance interpretation is difficult to say.
However we decide this question, we are left with Davidson's assertion that "there are no rules for arriving at passing theories, no rules in any strict sense, as opposed to rough maxims and methodological generalities." And this, as we have seen, is where the mechanist and Davidson have chosen to disagree.

§3 Other cognitive faculties

At this stage in the argument, we can reply to Davidson in one of several ways. We can, for example, simply demur on the question. In the absence of any well-articulated and widely accepted theories of grammatical or pragmatic competence, we are simply tilting at windmills. It is still too early to tell whether or not the mechanist's theory of our basic linguistic competence will have anything meaningful to say about our ability to interpret malaprops. There is, as I indicated above, some suggestive research on the question, but nothing that is conclusive. Linguists continue to pitch battles for the allegiance of future theorists; psycholinguists are forced to make some rather provisional assumptions about the syntactic and computational models that they use to interpret their results; all is flux.

Underlying all this talk of future theorizing, however, is the assumption that our ability to interpret malaprops falls within what Chomsky has called our "cognitive capacity":

79
An intellectually significant science, an intelligible explanatory theory, can be developed by humans in case something close to the true theory in a certain domain happens to fall within human "science-forming" capacities. ... Thinking of humans as biological organisms in the natural world, it is only a lucky accident if their cognitive capacity happens to be well-matched to scientific truth in some area. It should come as no surprise, then, that there are so few sciences, and that so much of human inquiry fails to attain any intellectual depth. Investigation of human cognitive capacity might give us some insight into the class of humanly accessible sciences, possibly a small subset of those potential sciences that deal with matters concerning which we hope (vainly) to attain some insight and understanding. 48

If Davidson's analysis of our malaprop-interpreting ability is correct, if, in fact, much of what we do in this domain is "derived by wit, luck, and wisdom," then we might have grounds for pessimism. It would then appear that any explanation of our malaprop-interpreting abilities would have to include some account of how wit, luck, and wisdom intermingle—and in what measure—to yield the intended interpretations. I am here speaking figuratively of that mysterious process wherein past experience, beliefs and desires, sudden flashes of inspiration, and other factors that we cannot even name, cooperate and contend to determine our behavior—linguistic and otherwise. But although I am myself skeptical about the possibility, why is Davidson so certain that wit, intuition, and skill—the factors that he claims are necessary for utterance interpretation—cannot be given an adequate rule-governed description?

Whether or not we accept Chomsky's assessment of our
biologically limited cognitive capacity, we might worry over the question of how we, as theorists, can give a rule-governed account of an ability that takes us so far into the storms of human intentionality. Can there be a rule-governed account of our ability to recover the intended meaning of a malaprop? Need our account be rule-governed in order to achieve some measure of explanatory and descriptive adequacy? There is, I believe, no a priori answer to these questions, for the sciences—old and new—continue to make conceptual inroads to previously recondite domains. A scientific study of the nature of human cognitive capacity might, as Chomsky suggests, "give us some insight into the class of humanly accessible sciences,"50 but we are from the stage when we can talk meaningfully about the a priori limitations of any inquiry into this domain.

Davidson may be right when he argues that much of what the interpreter can do "ought not to count as part of his linguistic competence."51 But then there is nothing to rule out the possibility that other non-linguistic faculties of mind may have a role in the process of utterance interpretation and production. There are several senses in which this observation is trivially true. We hear the spoken word, or see it written on the page, or feel it spelled in our hand—but certainly none of these faculties is exclusively, or even primarily, linguistic. We use language to recount our memories, express our thoughts, and signal our intentions; but again those faculties of mind that make it possible for us to harbor memories, thoughts, and intentions are not, exclusively, linguistic faculties. As an example of a non-linguistic faculty of mind intervening in the process of interpretation, consider my leftist friend
who turns to me and says, "George Bush is a fine president." Even if the conventional signals of ironic speech are missing—the sarcastic tone of voice, the rolled eyes, the pause for effect—I nevertheless take him to mean the exact opposite of what he has said. What has intervened in this case is my knowledge of my companion's strong political likes and dislikes. In order to interpret him correctly, I had to exercise a non-linguistic as well as a linguistic faculty of mind.

Chomsky has written quite extensively about these "interactions" of the language faculty with other non-linguistic mental faculties. In his *Reflections on Language* (Chomsky, 1975), for example, he writes:

> The place of the language faculty within cognitive capacity is a matter for discovery, not stipulation. The same is true of the place of grammar within the system of acquired cognitive structures. My own, quite tentative, belief is that there is an autonomous system of formal grammar, determined in principle by the language faculty and its component UG. This formal grammar generates abstract structures that are associated with "logical forms" ... by further principles of the grammar. But beyond this, it may well be impossible to distinguish sharply between linguistic and nonlinguistic components of knowledge and belief. Thus an actual language may result only from the interaction of several mental faculties, one being the faculty of language. There may be no concrete specimens of which we can say, These are solely the product of the language faculty; and no specific acts that result solely from the exercise of linguistic functions. 52

The extent to which the language faculty interacts with other systems of knowledge and belief is, of course, a matter for empirical discovery. There is no saying a priori what kinds of competencies are needed for the
interpretation of utterances. It is an idealization to assume that it is possible to "isolate" the language faculty from the other faculties with which it is intertwined—but it is an idealization whose theoretical utility has been tested and proven by many linguists and philosophers.

We can summarize and sharpen our response to Davidson as follows. Modern linguistic theory, as least one very influential school of it, assumes that a language-user has a certain core grammatical competence that is characterized by the theory which the linguist calls UG. Chomsky writes:

... it is hardly to be expected that what are called "languages" or "dialects" or even "idiolects" will conform precisely or perhaps even very closely to the systems determined by fixing the parameters of UG. This could only happen under idealized conditions that are never realized in fact in the real world of heterogeneous speech communities. Furthermore, each actual "language" will incorporate a periphery of borrowings, historical residues, inventions, and so on, which we can hardly expect to—and indeed would not want to—incorporate within a principled theory of UG.53

The mechanist assumes that the interpretation of actual utterances is not achieved by this core grammatical competence alone, but is rather aided and abetted by the interpreter's pragmatic competence. To put the whole thing rather crudely, grammatical competence gives us sentence meaning, whereas pragmatic competence gives us utterance meaning. The assumption underlying this distinction is that the sentence used to make an utterance has a well-defined meaning which I shall, following standard practice, call its literal meaning. Our pragmatic competence allows us to
"build on" or "fill out" the literal meaning by using material taken from the context of utterance. Building in this way, we thus recover the speaker's intended meaning. I shall discuss these and other underlying assumptions in Part III. There is room in this modular picture of the language faculty for the intervention of our other cognitive faculties. We can expect the language faculty qua biological system to interact with other such systems, in part as a matter of utility, in part as a matter of historical or evolutionary accident, and in part as a result of forces that we cannot even begin to name or understand.

Against the background of this theory or, more precisely, this framework of inquiry, we can ask a number of meaningful questions about the phenomena which Davidson claims "threaten standard descriptions of linguistic competence ... ."54 The manner in which we interpret malaprops, for example, may or may not be affected by our basic linguistic competence—it is still too early to tell—although, as we have seen, there have been a number of suggestive experiments in this domain. As far as I know, no one has yet proposed a detailed model of how grammatical and pragmatic competence interact in the interpretation of malaprops. What I have called "alternative interpretive strategies" may be involved, but whether or not these strategies are governed by shared rules and conventions known to both speaker and hearer in advance is still an open question, as is the question of the extent to which these strategies should count as part of our basic linguistic competence.

I believe, however, that the very fact that within the mechanist
framework we can raise meaningful questions about the phenomena alluded to by Davidson, suggests that this framework or model may not be far wrong. It is now only a matter of working out the all-important details.

§4 Davidson's Ubiquity Hypothesis

Davidson's Ubiquity Hypothesis (Part I, §4) is a little difficult to assess. Even on Davidson's own account, it appears to be obviously false. In his own outline of the process whereby the hearer recovers the speaker's intended meaning from the malaprop (§2, above), Davidson makes essential reference to the hearer's "standard interpretation" of the utterance in question. It is the ludicrousness of this standard interpretation (in the given context) that alerts the hearer to the fact that something has gone wrong with the speaker's utterance. In this case can we not say, contra Davidson, that there was such a thing as how the speaker could expect, in the abstract, to be interpreted? If, for example, the speaker could assume that both he and the hearer share a "standard" knowledge of English in which "derangement" means derangement and "epitaph" means epitaph, could not the speaker, if he carefully uttered the words "What a nice derangement of epitaphs" in the proper context, expect to be interpreted in the "standard way"?

Davidson might counter that this hardly constitutes a case in which the speaker could expect to be interpreted in a certain way in the
abstract. The description we have given of the context in which the speaker could expect to have his words interpreted literally is thick with felicity conditions: the speaker and hearer must share a standard dialect; the speaker must not misspeak his words; and the context of utterance must be appropriate.

And yet it appears to me that these felicity conditions are far from rare, and are in fact met much of the time. Of course these conditions are not met all the time: there certainly is no such thing as how we can expect, in the abstract, to be interpreted, if the only language-users we are likely to encounter are monolingual speakers of Yoruba. How, then, do we interpret Davidson's Ubiquity Hypothesis?

Davidson discusses the Ubiquity Hypothesis in the following passage:

The less we know about the speaker, assuming we know he belongs to our language community, the more nearly our prior theory will simply be the theory we expect someone who hears our unguarded speech to use. If we ask for a cup of coffee, direct a taxi driver, or order a crate of lemons, we may know so little about our intended interpreter that we can do no better than to assume that he will interpret our speech along what we take to be standard lines. But all this is relative. In fact we always have the interpreter in mind; there is no such thing as how we expect, in the abstract, to be interpreted. We inhibit our higher vocabulary, or encourage it, depending on the most general considerations ...

Thus Davidson agrees that we often assume that our speech will be interpreted along "standard lines," but he implies that because we always
have our interpreter in mind, "there is no such thing as how we expect, in the abstract, to be interpreted." Much depends, of course, on the extent to which we incorporate our knowledge of the hearer's linguistic idiosyncrasies into our prior theory for the hearer. Oftentimes, however, it appears that we have no such prior knowledge; and that we make few, if any, adjustments in those cases where we have ample knowledge of our hearers. If, under fairly standard circumstances, and under little more than the assumption that my hearer belongs to the same language community, I order a crate of lemons by using the words, "I would like a crate of lemons," then it is true that I can do little better than expect that my hearer will interpret me to mean that I would like a crate of lemons. I simply assume, in this case, that my utterance will be interpreted along standard lines—however groundless my assumption may later turn out to be. Here is a case in which, as Davidson suggests, I clearly had my interpreter in mind. Why is it not, however, also a case in which I expected, in the abstract, to be interpreted in a certain way? I may, as Davidson says, have inhibited my higher vocabulary by choosing to say, "I would like a crate of lemons," rather than, "I would like a crate of the yellow, egg-shaped fruit of the spiny evergreen *Citrus limonia.*" But how would this consideration play against the fact that I had in this case a clear foreknowledge of how I would be interpreted?

Perhaps Davidson is arguing that I am not *licensed* to expect my words to be interpreted in a certain way unless I have good grounds for believing that my hearer shares my language, or, perhaps, my idiolect. The question now becomes: What constitutes good grounds for this belief?
This epistemological question, however, has no real bearing on the ability of a speaker and hearer to communicate successfully. The mechanist asserts that speaker and hearer interpret each other successfully because they share a knowledge of the same language, or, at the very least, because they share a kind of second-order interpretive competence with respect to a wide range of possible utterances (see §2, above). At the beginning of a conversational exchange, the speaker may assume—rightly or wrongly—that he will be interpreted in accordance with the rules of the posited common language. Communication succeeds if they share a certain competence, and fails if they do not—however well-motivated or unmotivated the speaker's assumptions about the hearer's competence may have been.

§5 Rule-governed versus "mechanical"

Davidson sometimes speaks as if the issue at hand were the theorist's inability to provide a strictly mechanical algorithm for recovering the intended meaning of a malaprop. In the same passage where he asserts that passing theories are derived by a combination of wit, luck, and wisdom, he also makes the following claim:

There is no more chance of regularizing, or teaching, this process than there is of regularizing or teaching the process of creating new theories to cope with new data in any field—for that is what this process involves.
Here is what I think Davidson is after. He wants the theorist to (1) provide an effective method for deciding whether a given utterance is, or is not, a malaprop, and (2) give an algorithm that constitutes a model of all the steps that the interpreter takes to recover the meaning of the malaprop, with the further proviso that none of these steps should appeal to abilities or judgments which are not themselves given the same kind of mechanical description. I will call such an account a mechanical description of our ability to interpret malaprops. I believe, however, that Davidson's requirement is much too strong. It does not follow that because a cognitive process is rule-governed, it can also be given a mechanical description, and our mechanist/pragmatist, despite his name, is not committed to such a view. Consider, for example, a candidate "rule" for the recognition of ironic speech:

[16] The speaker is speaking *ironically* if the utterance made by the speaker, when interpreted literally, does not jive with what the hearer knows about the speaker's previous statements and/or his past or present behavior and/or his beliefs, desires, etc., and the hearer detects a note of irony in the speaker's voice.

I have stipulated that no mechanical account of a cognitive process $C$ occurring in person $P$ can make explicit appeal to some judgment (conscious or otherwise) on the part of $P$ unless that judgment also be given a mechanical description. Assuming that the kind of judgments appealed to in [16] cannot be given a mechanical description, and assuming further that these judgments are essential elements in the recognition of ironic speech, it then follows that it is not possible to given a mechanical account of how we recognize ironic speech. Nevertheless we would be
flying in the face of intuition to say that the recognition of ironic speech was not somehow governed by rules or conventions. There is, of course, no a priori reason why we cannot achieve some measure of descriptive and explanatory adequacy in these domains without invoking these "mechanical" descriptions.

§6 Summary

It would be unfair to Davidson to say that he was not anticipating the kind of response that I am formulating here. Many of the ideas that I have reviewed were and are in common currency, and Davidson's own references to the work of Chomsky and other linguists indicate that he was au courant of modern trends in linguistic theorizing. What, then, can account for Davidson's profound skepticism? The answer is, I believe, the very peculiar nature of the phenomena that he chose to study. Malaprops, misspeaking, and the like—what I will henceforth call Class M phenomena—were to Davidson the unruly offspring of our unguarded tongues, oftentimes appearing suddenly in mid-sentence or mid-speech, leaving the interpreter no time to "accommodate" his linguistic competence to them, and yet finding themselves nonetheless successfully interpreted. To another theorist, with perhaps stronger allegiances to the Chomskyan framework that I have sketched above, the fact that we can spontaneously interpret malapropisms together with the fact that this ability appears to be universal, suggests that it might be possible to
explain the phenomenon in part by appealing to our basic linguistic competence. I have argued that there is good evidence for believing that such is the case.

Davidson, on the other hand, imagines that language can and does break out of whatever bounds theory might impose on it. Not for him the view in which meanings march in step with syntactic forms and the laws of pragmatics. Furthermore, linguistic behavior is not as far different in kind from other human behavior as some theorists now imagine. Hence his assertion that by accepting his skeptical conclusions, "we have erased the boundary between knowing a language and knowing our way around the world generally."58

Davidson’s explicit attempt to ward off a Chomsky-style response to his skeptical argument is seen in the following passage:

Perhaps it will be said that what is essential to the mastery of a language is not knowledge of any particular vocabulary, or even detailed grammar, much less knowledge of what any speaker is apt to success in making his words and sentences mean. What is essential is a basic framework of categories and rules, a sense of the way English (or any) grammar may be constructed, plus a skeleton list of interpreted words for fitting into the basic framework. If I put all this vaguely, it is only because I want to consider a large number of actual or possible proposals in one fell swoop; for I think they all fail to resolve our problem.59

The "problem" that Davidson refers to is, of course, our inability to satisfy the demand for "a description of an ability that speaker and hearer
share and that is adequate to interpretation. The problem with "general frameworks," according to him, is that any such framework, "by virtue of the features that make it general, will be itself be insufficient for interpreting particular utterances." He explains that our general framework or theory "can't be all that is needed since it fails to provide the interpretation of particular words and sentences as uttered by a particular speaker." This is, of course, a very odd claim for Davidson to make, for Chomsky would be the first to admit that UG is not a theory of utterance meaning, nor does it purport to be such a theory. Linguists of the generative grammar school would agree with Davidson that a language-user's knowledge of grammar is only one element of what is needed for sentence interpretation. Therefore Davidson's worry cannot simply be a worry over the distinction between sentence meaning and utterer meaning.

At the beginning of the Epitaphs paper the issue for Davidson appears to be the rather singular nature of much of what we come to interpret successfully. Starting with our ability to interpret malaprops, grammatically garbled sentences, new words and idiolects, slips of the tongue, etc., Davidson finds cause to doubt the standard account of linguistic competence given by linguists and philosophers. Very quickly, however, what was once seen as a fairly peculiar set of abilities, is now seen to permeate all of language, so that Davidson ends by concluding—incorrectly, I believe—that "there is no such thing as how we expect, in the abstract to be interpreted"
Contra Davidson I have argued, among others things, that (1) there is good reason to believe that the range of possible meanings of an utterance—malaprops included—is sharply restricted by the rules of the language; (2) Davidson has not shown that malaprops and the like are any more a problem for competence theories than "standard" demonstrative utterances. Even if we accept Davidson's claim that passing theories are derived in large measure by wit, luck, and wisdom, he has given us no reason to believe that these capacities cannot themselves be formally modelled.

It appears to me that Davidson has essentially misconstrued linguistic's self-conception, at least insofar as he has misjudged its awareness of its own present limits. He has, as it were, proclaimed nescience the victor before the final count was out—in fact, before the mechanist could even be said to have been knocked onto the canvass. I do not believe, therefore, that Davidson accomplished what he set out to accomplish, namely: to show, on more or less a priori grounds, that rule-governed accounts of interpretation cannot keep up with our ability to recover meanings from utterances, and thus arguing, to successfully motivate a profound skepticism for the current direction of linguistic theorizing.
Part III

Some details of the mechanist response to Davidson
Introduction

As we have seen, Davidson has tried to motivate his skeptical conclusions by pointing to such phenomena as our ability to recover the intended meaning of a malaprop, and, on that basis, arguing that the "standard" model of our linguistic competence cannot explain our broadly ranging interpretive abilities.

After a few preliminaries (§1), I will in §2 describe a little more carefully the standard interpretive model that I believe is the proper foil to Davidson's skeptical argument. It is this model that I will use to continue exploring the question of how much in the interpretive domain lies outside the reach of theory.

§1 Some preliminaries

The model of language use that emerges from the Epitaphs paper is one in which both speaker and hearer essentially "make up" a language as they go along, exploiting whatever strategies they can to understand other speakers and to make themselves understood. Davidson admits that shared grammars and the like may be "key ingredients" in what is needed for interpretation, but he plays down their role, and assumes instead that much of what transpires between speaker and hearer lies outside the pale of theory.
On the mechanist’s account, speaker and hearer are not entirely free to make up their language as they go along. Speaker and hearer can, of course, construct an artificial language—like Esperanto—and agree to use it in their conversational exchanges, but this kind of explicit treatment of language is not what our mechanist has in mind when he asserts that our use of language is constrained by our linguistic competence. What he does have in mind is this: the speaker cannot, for example, say

[17] John believes that Mary likes himself

and expect that his hearer will take him to mean that John believes that Mary likes him (John). According to the mechanist, this interpretation is simply not provided for by their shared grammar (assuming that speaker and hearer are speaking what we may loosely refer to as "English"). Speaker and hearer may, of course, conspire to use [17] in this non-standard way, but even here it is unlikely that their new method for interpreting [17] will "percolate" to other sentences that contain similar constructions, for example:

[18] John believes that Sam likes himself.

The basic linguistic competence shared by speaker and hearer does more than simply constrain the kinds of sentences that they can produce; it is in fact what makes it possible for them to interpret each other in the spontaneous—oftentimes unerring—manner that they do.
Thus the theorist by no means plays down the role that shared grammars may have in the phenomenology of utterance interpretation. And he leaves open the question of how much of what transpires between speaker and hearer lies beyond the grasp of theory. It is in fact sound methodology to assume that his theory can "tell all," and then to test the limits of his theory on the basis of that assumption.

In the next chapter I will explore the extent to which a mechanistic theory cannot account for our ability to recover meanings from utterances. As it stands, this question is hopelessly vague, for I have left unanalyzed the notion of "recovering a meaning"; I have given no account of what I wish to include or exclude from the idea of a language; and I have said precious little about the "theory" that I would like to endorse, or about what might constitute an acceptable "account" of the phenomena in question.

Let me first, by way of circumlocution, begin to address the question of what it means to "recover a meaning."

What we might broadly call our "interpretive powers" are not confined to language alone. Imagine, if you will, a panel of musical cogniscienti who are invited to sit and "interpret" a Haydn symphony. These experts are of course familiar with the "language" of music. Where the untutored ear hears only an admixture of sound and silence, they discern a remarkable amount of structure. Within the overall construction of the piece, they notice, for example, that the first movement is written
in sonata form. They also notice that the composer begins developing his main theme in the A section of this movement. This theme has a complex structure all its own and includes a motive marked mainly by its rhythm, etc. Some of our experts might even be able to recognize individual chords and notes within the piece. When it comes to interpreting the whole of this wonderfully elaborate structure, or any of its parts, we are, however, likely to find wide disagreement among our panel of experts. What to one panelist sounds "airy" and "delicate" to another sounds "anemic" and "thin." One expert concludes that the symphony as a whole explores the struggles of the Viennese working class, whereas another believes that it expresses Haydn's contempt for Italian opera. It is, I suppose, an interesting fact about our cognitive makeup that we can extract these rather abstract "meanings" from a medium that is almost purely form. (There might also be some good reason why interpretations of music are so seldom expressed as complete predications, given that it is often the mark of a good piece of music that it "makes a complete statement.") But the lack of agreement among our panelists, and our inability to say how these various judgments have their point of origin in the music—to say, in effect, how symphonic structure maps onto symphonic meaning, leads us to despair of any theory in this domain. De gustibus non est disputandum.

Contrast the above scenario with one in which an expert panel of language-users—rhetoricians, poets, philosophers, and linguists—is invited to sit and interpret the discourse of a speaker. As was the case with the Haydn symphony, there is a remarkable amount of structure discernible in the speaker's discourse. There are, for example, such minutiae as the
phonetic features of his words, the contours and rhythms of his utterances, the lengths of his pauses—all of which flavor his discourse. At the other end of the spectrum, we find the more extended features of his speech: the division of text into such gross units as "paragraph" and "syllogism." It is also quite possible that in this new scenario we will find some of the disagreements that plagued our former panel of experts: no two panelists will agree, for example, on how to interpret the speaker's discourse as a whole. Is the speaker presenting an argument to the effect that missiles are peacekeepers, or is he speaking with profound irony?

At the level of sentences and sentence structure, however, we find quite a different story. There appears to be a remarkable congruence of opinion when we ask our panelists to interpret individual sentences, when construed both inside and outside the context of the speaker's discourse. All of our panelists agree, for example, that in the sentence

[19] Who did these scoundrels expect to attack them?

these scoundrels and them can be coreferential, whereas in the sentence

[20] These scoundrels expect to attack them

they cannot. And so on, for a myriad of other sentences comprising the speaker's text and a myriad of other judgments of interpretation. From a scientific point of view, this universal agreement on judgments of utterance meaning is a remarkable fact—one that cries for explanation.
Consider now how Davidson proposes to account for this phenomenon. Using the language of the PT model, his claim is that each member of our panel of experts comes to share a passing theory with the speaker. According to Davidson, each panelist comes to the interpretive session equipped with a number of "rough maxims and methodological generalities"; each derives his passing theory "by wit, luck, and wisdom from a private vocabulary and grammar, knowledge of the ways people get their point across, and rules of thumb for figuring out what deviations from the dictionary are most likely." Given the highly personal nature of their interpretive arsenals, however, we might ask how our panel members come to achieve such wide agreement over the meanings of particular sentences. If each panel member uses a different road map, then how do they all end up at the same place?

By putting the question in this way, we have, in a sense, upped the ante for Davidson. No longer is it a matter of two language-users—speaker and hearer—in private cabal, attentively observing and recording each other's speech habits and idiosyncrasies, quietly yearning to break each other's linguistic codes. Now we have the spectacle of one speaker addressing an entirely assembly of hearers, with each hearer unerringly zeroing in on the speaker's intended meaning and, most importantly, with all the hearers agreeing amongst themselves as to how the speaker should be interpreted. This added dimension of intersubjective agreement amongst many hearers makes Davidson's talk of private interpretive strategies appear suspect.
If each member of our language panel goes about interpreting the speaker in his own idiosyncratic and unsystematic way, then how come they to share so many judgments of utterance meaning? We cannot explain this agreement by appealing to the fact that they are all aiming at the same target—namely, what the speaker means by uttering S—for none of the hearers knows in advance of the utterance what the speaker intends to mean by making his utterance. Put another way, the hearers will not know what target they are aiming at until after they have hit it. The purpose of each hearer's interpretive strategy is certainly to recover the intended meaning of the speaker's utterance by whatever means he can, but this observation buys us nothing. Each hearer is undeniably "aiming at" and hitting the same target—i.e. each hearer comes to agree that by uttering s the speaker meant m. The question is how so many different interpretive strategies could so accurately converge on the same meaning for a particular utterance.

I have already argued, contra Davidson, that it is possible to give an account of a shared "interpreting machine" that is adequate to the interpretation of particular utterances, along the lines that Chomsky and others have sketched. I will therefore assume that we can attribute some such basic linguistic competence to each language-user, and that this basic competence furthermore comprises the "core" of the language-user's "interpretive mechanism." I will give some of the details of this "mechanism" below. Against the background of these assumptions, we can now attempt to sharpen our original question: Are there cases of an interpreter "getting meanings" from language that theory cannot account
If we are talking about one interpreter in isolation from all the others, then the answer to this question is almost certainly "yes": we cannot expect our theory of linguistic competence to anticipate the idiosyncrasies of each particular language-user. Whatever theory we devise will have to be general enough to account for the competence of each member of our language panel, for example, and yet also be flexible enough to accommodate their individual idiolects. Here I leave open the exact nature of this "accommodation." When I refer to the interpreter's "getting a meaning" from an utterance, I therefore intend those meanings that we might expect our interpreter as a member of a larger speech community to recover—meanings that a significant fraction of our speech community might agree upon.

Assuming, then, that we are talking about a "normally constituted" language-user, in a homogeneous speech community, what can we say about the limits of his linguistic competence? Do recovered meanings far outstrip theory, as Davidson suggests? We can begin to answer these questions by considering a little more carefully the kind of theory that our skeptic might expect recovered meanings to outstrip. I have sketched one such "theory," or model, in figure A.
Figure A. A model for sentence and utterance interpretation.
§2 A model for sentence and utterance interpretation

The model shown in figure A is a distillation of some leading ideas in linguistics and the philosophy of language, together with some brash interpolations about what occurs in the process of utterance interpretation. It is, in short, a conceptual Frankenstein's monster, one whose details I will not too jealously defend. My purpose in presenting this model is to make explicit the kinds of processes that I assume take place when we interpret the speech or text of another. The purpose of this section is not to undertake a detailed, critical analysis of the assumptions underlying the mechanistic picture of interpretation, but rather to sketch some of its salient features and to touch on some of the problems associated with models of this kind. I do hope in the process to motivate a particular reading of the notion of literal meaning, one which I believe is motivated by current work in linguistic theory. I ask the reader to postpone asking too many probing questions until I have had a chance to establish the commitments of this model.

Here are its gross features. Figure A shows the "path" followed by a phone sequence or written string string S (as produced by P in context C) through the linguistic modules of an interpreter. This path is indicated by the solid arrows. Initially S is "scanned" by the parser and assigned a constituent structure. The solid line connecting the parser to the grammar is put there to suggest the idea that in assigning a constituent structure to S, the parser—in one way or another—accesses the information represented in the grammar. I will explore this connection in
more detail in Part IV. The output of the parser, a representation of the constituent structure of $S$, is further processed (hence the "Processor" in figure A), and the output of the entire module that I have labeled the SID (for Sentence Interpretation Device), is what linguists of the GB school call the LF-representation of $S$.

The LF-representation of $S$ is, in effect, a representation of what I take to be the "literal meaning" of $S$—putting aside Davidson's qualms about this notion. I have construed it here as the output of the so-called "LF" level of syntactic structure, this being the level that serves as the interface between the formal grammar and a representation of the meaning of $S$. According to Lasnik (Lasnik, 1988), "LF represents the contribution of the grammar to the meaning of particular sentences." I will return to these assumptions, but let us now continue to follow $S$ through our enigmatic interpreting machine.

The LF-representation of $S$ now enters waters that I believe are fairly well unchartered by theory. What I have sketched to the right of the LF-representation of $S$ in figure A is something very much like those maps of ancient cartographers who knew only the rough outlines of the continents, and who, through simple ignorance, altogether excluded some very large land masses and sea passages.

Here the LF-representation of $S$ is assigned a semantic representation by the Standard Semantic Unit (or SSU, for short). For the purposes of my discussion it is not entirely essential that we assume that
any such level of representation exists. It is enough for my purposes that we have in our model something the corresponds roughly to the literal meaning of $S$ (i.e. its LF-representation) and something that corresponds to its utterance meaning (i.e. the output of the entire system—what $P$ meant by producing $S$ in context $C$). An intermediary level of semantic representation may, however, be motivated on certain grounds that I will discuss shortly.

What I have provisionally labeled the Pragmatics Unit in our model is really a catch-all for rule systems, heuristics, principles of conversational cooperation—whatever they might be—that help the interpreter do some or all of the following: (1) disambiguate ambiguous sentences; (2) identify the speaker's intended referents; (3) figure out conversational implicatures; (4) interpret ironic, sarcastic, figurative, and other non-literal speech; etc. Connected to the module that I have called the UID (for Utterance Interpretation Device), but lying outside of it, are the other cognitive faculties that we might exploit to recover the utterer's intended meaning—among them, our common sense. I place these cognitive faculties outside the UID only because I wish to suggest their independence from the language faculty. I expect that these faculties play key roles in other domains that are not essentially linguistic (e.g. scientific or mathematical reasoning, etc.).

Finally, out of this interpretation machine comes what our interpreter takes to be the utterance meaning of $S$—in a sense that I will make clear below. There I will argue that the LF-representation of $S$
delivered by the SID is essentially \textit{schematic} and must be "filled in" by the Pragmatics Unit and its associated cognitive faculties.

This, in very broad outline, is how our interpretation machine works. I will now spend some time "fine-tuning" the machine: providing more details of its operation and fielding anticipated objections.

Consider first the Sentence Interpretation Device. My model of the SID follows from two central assumptions: (1) the theory of grammar endorsed by Chomsky and others of the GB School is essentially correct; and (2) it is still too early to determine the exact nature of the relationship between the parser and the grammar. According to Chomsky, a grammar can be modeled as a number of interdependent levels of syntactic representation, as shown in [21]:

\begin{center}
\begin{tikzpicture}
  \node (d) {D-structure}
    child {node (s) {S-structure}
      child {node (pf) {Phonetic Form}}
      child {node (lf) {Logical Form}};
  }
\end{tikzpicture}
\end{center}

I assume that the oversimplified model I have presented here is familiar to most readers. Very briefly, the rules of the "base" of the grammar, which itself consists of a categorial component and a lexicon, generate D-Structures that are related to S-Structures by the transformational component (or, equivalently, by the rule Move-\(\alpha\)). S-Structure is further related to the so-called "interpretive" components, Phonetic Form (PF) and
Logical Form (LF). According to Chomsky:

... rules of the LF component ... convert S-structures to representations in LF, where scope and other properties are represented. PF and LF constitute the "interface" between language and other cognitive systems, yielding direct representations of sound, on the one hand and meaning on the other as language and other systems interact, including perceptual and production systems, conceptual systems and pragmatic systems.5

Given these and other similar comments in the literature about the LF component's constituting a kind of "interface" between the formal grammar and representations of meaning, I have taken the liberty of assuming, in the model given in figure A, that the syntactic domain (the SID of figure A) connects with the semantic domain (the SSU and the Pragmatics Unit) via the LF-representation of the given phone sequence or word string. This picture of the SID is, I believe, entirely natural, and it makes some rather minimal assumptions about the mechanisms involved.

Chomsky has stressed, on a number of occasions, that a grammar of the kind outlined in [21] is not to be taken as a model for a speaker or hearer. This is not to say, however, that the grammar will have little or no role in a reasonable model of language use. Indeed, if it is to have any explanatory value vis-à-vis the phenomenology of language, the grammar must play a central role in an interpretive model like that given in figure A. As Chomsky puts it:

No doubt, a reasonable model of language use will incorporate, as a basic component, the generative grammar that expresses the speaker-hearer's knowledge of the
language; but this generative grammar does not, in itself, prescribe the character or functioning of a perceptual model or a model of speech production.⁶

Thus we cannot simply ask how a correctly specified grammar will assign a structural description to an incoming phone sequence or word string \( S \); but we must ask rather how the parser together with the grammar will cooperate in such a way as to yield (eventually) the appropriate LF-representation of \( S \). The reader will notice that in figure A I leave the connection between the grammar and the sentence recognition device, or parser, rather ill-defined, suggesting it only by a single line connecting the two modules. I will return to some nagging questions about the parser in Part IV.

I urge the reader not to make too much of the role of the "Processor" in figure A. I include it there only because I need something that will map the structural representation of \( S \) (yielded by the parser) to the LF-representation of \( S \) (the output of the SID).

So much then for the SID. Its primary characteristic, as we have seen, is its ability to use grammatical information to generate a representation of the literal meaning of the phone sequence or word string in question. (Henceforth I will refer to phone sequences and word strings simply as "strings.") It is common among language theorists to assume that one can make a clear distinction between sentence meaning and utterance meaning, both at the level of common parlance and at the level of linguistic theory. In the model under discussion, I simply equate the
literal meaning of S with what is represented by the LF-representation of S. This, I believe, is a natural move, given what linguists have told us about the nature of LF.

If indeed the LF-representation of a string S represents something like what theorists have called the literal meaning of S, then how is it that we can access the literal meaning of S? The problem is this. The LF-representation of S appears as an intermediate step in the derivation of the utterance meaning of S. When we as interpreters hear a phone sequence or read a word string, it is always in the shadow of some utterance context. That is to say, we are never presented with the literal meaning of a given string directly, but rather with its utterance meaning. If by happenstance we were to find S scribbled on a small piece of paper in the middle of an open field, we would not be able to help guessing at the intentions of its author, or at the context in which it was written, and we would therefore interpret S against the background of this assumed context. Even the most sterile laboratory of the most cautious psycholinguist would be to us rich in suggestive context parameters. Put another way, it looks as if we can access the LF-representation of a string only after it has been doctored by the Pragmatics Unit.

On the other hand, it appears that we do in fact have access to the literal meanings of sentences, whatever their context of utterance might be. For I take it that part of what it means to know the literal meaning of a sentence S is knowing what context parameters are relevant to the interpretation of S, and I assume that we demonstrate our knowledge of
these relevant context parameters when, for example, we reason counterfactually about the meaning that \( S \) would have when uttered by different speakers on different occasions of utterance.

I suppose that in principle there is no bar to our being able to access what is in effect an intermediate step in the mental "derivation" of utterer meaning—although we cannot yet say what this means. Let \( r_{LF} \) be a theory-internal name denoting the representation of the literal meaning of some sentence \( s \) as uttered by \( p \) in context \( c \), and let \( r_U \) be the name for the representation of the meaning of the speaker's utterance. Let us suppose that when we assert that person \( p \) "knows" or "intuits" the meaning of an utterance, we are essentially saying that \( p \) stands in a well-defined (computational) relationship \( R \) to a representation of that meaning. We represent \( p \)'s state of knowledge by saying, in the language of our theory, that \( pRr_{LF} \). Can we also assert that at some time before, or even after, the mental derivation of \( r_U \) that \( pRr_{LF} \)? Do we cognize the representation of a literal meaning in the same way that we cognize the representation of an utterance meaning? These, of course, are matters for research to decide.

Perhaps even more vexing is the question of how we can claim to know the literal meaning of something that is essentially schematic. Consider, for example, the following sentence in abstraction from any occasion of utterance:

[22] Ramona's pet chicken likes home fries.
There are many things that we do not know about this sentence. To begin with, we do not know if anyone has ever actually uttered this sentence in some concrete situation. We can, of course, imagine sentence [22] being uttered in various contexts; and we can have some foreknowledge of what this sentence would mean in these contexts. But how do we answer the person who asks us what sentence [22] means literally? The disquotational theory of meaning is useless to us here, for we cannot very well say that the sentence "Ramona's pet chicken likes home fries" means that Ramona's pet chicken likes home fries. Saying this would leave our interlocuter with the impression that we were being evasive. To begin with, who on earth is Ramona, and which of her (possibly many) pet chickens are we referring to? Does this chicken like to eat the home fries, or does it rather feel an unnameable attraction to them? It is for this and other reasons that I say that the literal meaning of a sentence is essentially schematic. In the case of sentence [22], we need a theory of reference that shows us how to "locate" the referent of the name Ramona and of the noun phrase Ramona's pet chicken. We need contextual information to decide which of the many senses of the verb like the speaker intends. Nevertheless, in spite of our fragmentary knowledge, we do not hesitate to say that we know the literal meaning of sentence [22]. How can this be?

Consider an analogy. How is it that we might claim to know the meaning of an $n$-place function, $f(x_1, x_2, x_3, ..., x_n)$? To know the meaning of this function is to know how to compute different values of $f$ for different
values of its independent variables. We can do a lot of interesting mathematics with \( f \) without ever knowing the value of a single variable. Likewise we might suppose that to know the literal meaning of a sentence \( S \) is to know how to compute the meaning that \( S \) would have in context \( C \) with context parameters \( (C_1, C_2, C_3, \ldots, C_r) \). We need not know the values of these context parameters; we need only know how utterance meaning varies with \( (C_1, C_2, C_3, \ldots, C_r) \).

Returning now to the interpretation machine of figure A, we can inquire a little more carefully after the nature of the module that I have called the "Pragmatics Unit." Consider first the Standard Semantic Unit, or SSU. Why, for example, have an SSU in the first place? We have already said that the LF-representation of \( S \) represents the contribution of the formal grammar to the meaning of \( S \). Why not simply assume that this representation suffices for a determination of the utterance meaning of that string? Why posit an intermediary level of semantic representation? In answer to these questions, I ask the reader to consider again a sentence like [1], repeated here:

[23] The ham sandwich barked at my wristwatch.

A typical speaker judgment concerning this sentence might read as follows: the sentence is syntactically unimpeachable; semantically, however, it is a little "odd." The semantic oddity of this sentence is, of course, attributable to the fact that ham sandwiches are not animate, and are certainly not the kinds of things that would behave like dogs. We can, however, slice things a little differently and say that sentence [23] is
semantically impeccable—it does, after all, appear to express a meaningful proposition about a rather bizarre ham sandwich. We can imagine ourselves having uttered sentence [23] in a dream filled with playful lunch items. Upon waking, we might have concluded that it was not our language that had changed in the dreamscape, but rather what we had accepted to be our model of the world. On this construal, we would essentially be positing three possible levels of representation for sentence [23]: (1) the syntactic level, where we judge that [23] is a well-formed sentence of English; (2) the semantic level, where we judge that [23] expresses a meaningful proposition; and (3) the level of cognitive coherence (for lack of a better name), where we determine that [23] does not mesh with our beliefs about the world. As to judgments of well-formedness, our grammar can give us those, and judgments of cognitive coherence might be forthcoming from our Pragmatics Unit. But how do we come to conclude that [23] is semantically unassailable? It is here that we begin to feel the temptation for including something like a Standard Semantic Unit in our model of sentence and utterance interpretation. We would certainly like our SSU to account for simple judgments of semantic well-formedness, or coherence. But we might also like it to make a principled distinction between sentences like [23], whose oddity might disappear in the appropriate context, and sentences like

[24] My stepfather is taller than himself

whose oddity no context can erase. As I said earlier, it is not essential to my argument that there be anything like the SSU in our interpretation machine; and I do not mean to suggest anything by making the SSU a
subcomponent of the Pragmatics Unit rather than an independent component of the UID altogether. I therefore leave any decision concerning its status to the reader.\textsuperscript{7}

As mentioned earlier, I have held the Pragmatics Unit responsible for almost all of what goes on at the non-syntactic end of language. Other than the exigencies of exposition, there really is no good reason for doing this. The Pragmatics Unit, as we have seen, subsumes our ability to disambiguate sentences, identify speaker referents, interpret non-literal forms of speech (irony, sarcasm, etc.), recognize conversational implicatures, and perform a host of other functions that involve, among other things, correctly identifying the intentions of the speaker. We certainly cannot expect such an odd assortment of abilities to be subsumed under one theory. By including them under one rubric, I do not mean to suggest that they are in any way consubstantial. There is, however, one feature that they all share: they are all abilities that help us determine the intended meaning of the speaker by drawing on features of the utterance context.

This raises its own set of questions. We assume that the process whereby we integrate contextual information with syntactic information to construct utterance meaning is rule-governed. Can these rules be properly studied within a science of language? or do these rules reflect abilities that are not exclusively linguistic and are better studied in the context of some other domain of human behavior? If we wish to cleave as closely as possible to our "core" grammatical competence, we might, for
example, ask how this competence "informs" or "constrains" our pragmatic competence. What clues can our grammatical competence provide as to the structure of our pragmatic competence? On this matter Higginbotham writes:

Perhaps nothing at all that people say has its meaning wholly independently of context. It does not follow that semantic theory has little to say, or that it is in any way intrinsically incomplete. On the contrary, it is only through the context-independent features of meaning that we know what aspects of a context to take as relevant in the first place to the interpretation of the utterances around us.\(^8\)

The suggestion here is that our theory of grammatical competence will itself be able to suggest which context parameters are relevant to the interpretation of particular utterances. But things might also happen the other way around: a well-articulated theory of pragmatics might show us how certain aspects of utterance meaning can be extracted from the context proper, thus relieving somewhat the burden of the formal grammarian. Much of what we call pragmatics might very well walk lockstep with our theory of grammatical competence. Or, as some theorists have suggested, certain subfields of pragmatics might be more naturally incorporated into some broader cognitive theory—a theory of convention-governed behavior, say, or a theory of names.

Closely allied with this question of boundary drawing is the question of where utterance-meaning begins and where it ends. Where it begins is fairly clear: a theory of pragmatics must pick up where a theory of grammar leaves off, assuming, of course, that the latter can in large
measure be developed independently of the former. Where utterance meaning ends, however, is a little harder to say. Part of the reason for this is the ambiguity in the way we have chosen to circumscribe pragmatics's domain: we say that pragmatics is essentially concerned with the contributions of context to utterance meaning. But our notion of context may be too all-inclusive. Consider, for example, the husband who asks his wife, in a very pointed way, the whereabouts of his "power tie." We can imagine his wife turning to their befuddled child and explaining, "This means that Daddy's meeting with a client today." Of course literally his request means nothing of the sort, and the question we must ask is whether this implied meaning should count as part of the speaker's utterance meaning.

Here we have strayed rather far from the picture I outlined earlier where literal meanings are modeled as n-place functions waiting to have their argument places filled. In that picture, all we need do is throw a few context parameters into our meaning function, and out comes the recovered utterance meaning. This may be a useful picture for imagining what transpires between the speaker and his hearer, but the model becomes somewhat strained in the case of our suburbanite in search of his power tie: the connection between the implied meaning of his utterance and its literal meaning is much too oblique. Recall that the "meaning function" under discussion is essentially a representation of the utterance's literal meaning, this representation being derived from the structure of the utterance together with the meanings of the individual words used in the utterance. In this case, the purported "meaning
function" would have to be very strange indeed to yield the utterance meaning recovered by his wife (essentially, "I am meeting with a client today"). It might, of course, be possible to draw a line between "proper" utterance meanings (the possible values of the proposed meaning functions), and speaker implicatures that bear only a distant relationship to the utterance's literal meaning. We are not yet at the stage, however, where we can adequately address this question.

What, then, is the utterance meaning of a phone sequence or word string $S$ as produced by $P$ in context $C$? On my account, it is, or rather, can be, many things to many people, depending in large measure on their ability to pick up contextual clues; their knowledge of the beliefs, desires, etc. of the speaker; their own willingness or desire to understand the full import of the speaker's words, or draw inferences from them, etc. The speaker, on the other hand, may fully intend that his hearer interpret him, for example, to be speaking ironically, and he may in fact provide all the requisite signals of ironic speech, and yet all these clues as to his intended meaning may fall on deaf ears. This observation need not, of course, affect our theory of the relevant context parameters vis-à-vis the identification of ironic speech. It might, however, move us to update inter alia our theory of the hearer's knowledge of these parameters or, perhaps, our theory of the hearer's cognitive state.

In short, the Pragmatics Unit as I have sketched it is a hodgepodge of interpretive strategies, some perhaps fitting neatly into their own domains of inquiry, other fading imperceptibly into the
maelstrom of our other cognitive faculties. The interpretation machine that I have presented in figure A is, on this account, much too "modular." It is unlikely that the neatly packaged "cognitive faculties" and other modules that I have sketched there correspond in any interesting way to what we find in nature, unless, of course, the phrenologists have been right all along. There is no reason why we should not expect the various modules sketched in figure A to affect each other in unforeseen ways. There is no bar to the existence of essentially "pragmatic" principles that intervene, as it were, in the affairs of syntax. Such, for example, is the suggestion made by Chomsky on the question of what conditions determine the choice of "controller" in syntactic representations that involve the empty category PRO.9 Other such "interventions" are certainly possible. The machine outlined in figure A and defended here is a useful idealization that helps us picture the gross features of that very complex and multi-layered phenomenon we call language.

We can fully expect that some phenomena relating to the recovery of speaker meaning will not so neatly fit our "useful idealization." We have assumed that any well-formed sentence of English has a literal meaning that the hearer can use to construct an appropriate speaker or utterance meaning. But are there sentences for which the initial "semantic integration" is itself not possible without key clues provided by the context? In his paper, "Boundaries and Components of Linguistics" (Fillmore, 1984), Charles Fillmore suggests that expressions known as contextuals may belong this category:

If the customer who has ordered a ham sandwich is referred
to by the waitress as the *ham sandwich* ..., if the one tour bus in a multi-bus tour that is scheduled to stop at a pumpkin farm is known briefly as the *pumpkin bus* ..., it might seem reasonable for us to say that these sentences do not *have* literal meanings outside of their context of use.\(^\text{10}\)

Fillmore is drawn to the conclusion that "...we need to bring into our account of language the fact that code-changing as well as code-exploiting creativity is an important part of linguistic competence, and that perhaps *ad hoc* name creating occurs in everyday language on a large scale,"\(^\text{11}\) but I am unconvinced that any such "code changing" occurs in these cases, if by code-changing Fillmore means an adjustment or addition to the rules or maxims that we use to "locate" the referents of noun phrases. The problem, as I see it, is not one of how context might shape the representation of *literal* meaning, but rather one of how the speaker might get the hearer to associate a given noun phrase with some object in his field of cognition. The "*ham sandwich*" and the "*pumpkin bus*" of the above passage may be no more or no less problematic for a theory of reference than the definite description "the man who wrote *Waverley*," and their meaning may be no more or no less schematic. As with the development of syntax in the past ten years, we may only now be discovering the relevant paradigms for an adequate theory of reference, and contextuals may be more properly construed as primary data for this theory. Imagine that the waitress says to her co-worker,

[25] The *ham sandwich* wants another cup of coffee,

and that her co-worker unambiguously understands her to mean that the gentleman who ordered the *ham sandwich* wants another cup of coffee.
Fillmore suggests that [25] has no literal meaning, and that "semantic theory will have to take into its scope the principles by which contexts figure in informal naming."\(^{12}\) This last claim might certainly be true, but I do not believe that we need abandon the assertion that [25] has perfectly well-defined literal meaning. I have suggested that knowing the literal meaning of an utterance is knowing what context parameters are relevant to the interpretation of that utterance, or, what may be equivalent, knowing how to reason counterfactually about the meaning that the utterance would have when uttered by different speakers on different occasions, and Fillmore obviously has a somewhat different view of what the literal meaning of an utterance consists in. On my view, the waitress of our example can utter [25] and get away with it because (1) she knows its literal meaning, and (2) she knows how, in the given context, this meaning will "react" with her hearer's theory of reference to yield her intended referents. Thus in answer to Fillmore I would not say that our waitress has "changed the code," but rather that she has exploited her knowledge of the code in order to refer in a hitherto novel way. By looking at the example this way we are essentially shifting the explanatory burden from our theory of semantics to our theory of reference, but in the process we are helping to preserve the sharp line that theorists have drawn between literal and utterance meaning.
§3 Summary

Having originally set out to explore the extent to which recovered meanings outstrip theory's ability to deliver them, I have to this end made explicit the kind of "meanings" and "theory" that I intend. I have presented a candidate "interpretation machine"—warts and all—one that I believe satisfies Davidson's request for "a machine, which when fed an arbitrary utterance (and certain parameters provided by the circumstances of the utterance), produces an interpretation."\(^{13}\)

Putting aside the possibility of an intermediate level of semantic representation (as given by the Standard Semantic Unit of figure A, for example), there are basically two kinds of meaning that we might expect our theory to have some say about, and I have very broadly called these literal meaning and utterance meaning. To put the matter as neutrally as possible, pragmatic "strategies" go to work on representations of literal meaning to yield, ultimately, representations of utterance meaning. It is of course entirely possible that an utterance meaning may "evolve" in a number of well-defined stages. For example, the first order of business might be the determination of the referents or exact senses of the various formative elements of the utterance, yielding an initial representation of utterance meaning. Next, the "disambiguation unit" might go to work on the representation (or representations), yielding a second, hopefully more accurate, representation of the speaker's intended meaning, etc. Thus the utterance meaning may pass through several stages, arriving finally at a fairly stable representation that we might venture to call the utterance meaning.
meaning. It is also entirely possible that the recovery of utterance meaning is a process that is rather more parallel and rather less serial than the one I have described here. In this case, we might expect the utterance meaning to be delivered to the interpreter without passing through several intermediate levels of representation.

Against the background of this theory of interpretation, one of the central questions of this paper becomes: Can a normally constituted member of a homogeneous speech community recover literal or utterance meanings that our interpretation machine cannot deliver to him? We of course want the meanings recovered by our interpreter to agree with those recovered by other members of our speech community. We are not, in other words, talking about highly individualistic interpretations of speaker meaning, but rather about the kind of widespread agreement on judgments of meaning that we might expect a scientific study of language to explain. By putting the question in this way, I am essentially asking from which quarter will come the challenger to the rather mechanistic account of utterance interpretation that I have sketched above, an account which I believe accurately reflects much current thought on the process of utterance interpretation.

As we have already seen, Davidson has tried to argue, on more or less a priori grounds, that our interpretive abilities are much too ad hoc, much too unconstrained, to walk lockstep with any "theory" that we might devise. The motivation for his skeptical argument, he has told us, derives from a consideration of certain abilities that I have elsewhere called
Class M phenomena, abilities which "threaten standard descriptions of linguistic competence ...". In the chapter that follows, I will explore the ways in which these phenomena might threaten to topple our interpretation machine.

There may, of course, be other phenomena that can bring our neatly circumscribed interpreting machine to grief, and which Davidson has failed to consider. I believe, however, that Davidson's list of contenders provides us with a good launching point for our discussion.
Part IV

Class M phenomena
Introduction

In §1 I will consider Davidson's list of Class M phenomena in the following order: (1) our ability to interpret malaprops and to correct slips of the tongue; (2) our ability to interpret words we have never heard before; (3) our ability to cope with new idiolects; and finally, (4) our ability to interpret grammatically garbled utterances. I will argue that this last ability constitutes the strongest contender to our standard picture of the interpretive process.

In §2 I will take a more detailed look at our ability to interpret the ungrammatical.

§1 A brief look at Davidson's list of Class M phenomena

Malaprops and slips of the tongue fall into a much broader category of linguistic phenomena known as speech errors. Speech errors have been given a lot of attention in the literature, but as far as I know, few, if any, researchers have proposed detailed mechanisms that might explain how competent language-users go about interpreting utterances that harbor them. Very often, speech errors have been seen to provide a kind of indirect evidence which supports the formative units and grammatical rules posited by linguists. According to Fromkin, a prominent speech error researcher, if the formatives and rules of the
linguists are found necessary to explain certain aspects of linguistic
performance, then "there is little reason to doubt their existence as part
of grammars, particularly when their grammatical 'existence' is
necessitated by other linguistic evidence."2 Another researcher, D. B. Fry,
asserts that speech errors "cannot be accounted for except by attributing
to speakers a knowledge of a language system certainly similar to and
even in many details coincident with that constructed by linguists."3

We must be careful how we interpret these claims, however. For
in the absence of any widely attested model of performance, it is unclear
how we might relate the evidence concerning the production of speech
errors to the putative process whereby we interpret the utterances that
harbor them.

Consider first the case of the malaprop. We can break down the
process of malaprop interpretation outlined earlier (Part II, §2) into a
number of steps:

[26] (a) Recover the literal meaning of the sentence containing the
malaprop.
(b) Come to realize that the literal meaning cannot be the speaker's
intended meaning.
(c) Conclude that the speaker misspoke—i.e. that he produced a
malaprop.
(d) Search the space of the phonological neighbors of the sentence
used to make the malaprop.
(e) Continue to do step (d) until you locate a sentence that makes
sense in the given context.

At first glance it appears that much of what we do from step (b) onward
is shrouded in mystery. I assume that there is some process whereby we detect a "mismatch" between a sentence's literal meaning and the context within which it is produced, and I further assume that this process is rule-governed. But how is it that we come to decide what kind of error, if any, the speaker has committed in making the utterance (step (c))? I suppose that our knowledge of the context and a number of pragmatic principles might help us here. We might very quickly come to realize, for example, that none of the usual methods of sentence construal apply in the given context: we find no evidence, for example, that the speaker is speaking ironically, or sarcastically, or figuratively, etc. On the contrary, there might be good positive evidence that the speaker has indeed misspoken—for example, we may know from past experience that he is a habitual malapropiser, or his face may somehow register the fact that he has committed an error, etc.

The fun really begins at step (d). What kind of mechanism or mechanisms can we propose for our "search" through the space of phonological neighbors? The problem with any proposed search mechanism is the potentially large space of candidate replacement sentences. Consider, for example, the malapropism

[27] We sent her a card expressing our symphonies.

We might, on pragmatic or semantic grounds, begin by determining that the word *symphonies* makes little sense in the context "We sent her a card expressing our ...," and we might therefore attempt various replacements of this word:
[28] We sent her a card expressing our syntheses
    ... cynosures
    ... slim ponies
    ... etc.

until we finally arrive at the speaker's intended utterance, "We sent her a card expressing our sympathies." Suppose that we can give a mechanical procedure for searching through the space of [27]'s phonological neighbors—by scanning the lexicon or a subsection of it in some kind of dictionary order, for example. We might nevertheless find that the list of candidate replacements for sentence [27] (like that given in [28]) will grow too large for us to scan in the time that it usually takes us to recover the intended meaning of the malaprop.

This, of course, is not a good reason for abandoning a model that has as much initial plausibility as does [26]. We might, for example, assume that a number of search principles or heuristics operate to constrain the list of candidate sentences that we search through. I have already suggested one such simplifying principle above: we should attempt to replace only those words that appear to be out of place in the sentence that we have decided contains a malaprop. In sentence [27], for example, we do not attempt to find substitutions for the word sent, or the word card, etc. Other possible heuristics might dictate that we search only for replacements in the same word class, or with the same number of syllables, or with the same syllabic stress, etc.—these heuristics being suggested, for example, by our study of speech errors.
I spend as much time as I do dwelling on the case of malaprop interpretation because contrary to what Davidson has suggested, there is no a priori reason for believing that the phenomenon in any way threatens standard descriptions of our linguistic competence. Admittedly, much of what happens in the interpretation of a malaprop appears to happen at the pragmatic end of the meaning spectrum, where our knowledge of the relevant mechanisms is rather sketchy. But this in itself is no cause for pessimism. The very fact that the production of malapropisms and other speech errors appears to be fairly rule-governed suggests that it may also be possible to uncover the principles underlying their interpretation.

Davidson has suggested that our ability to interpret words we have never heard before also threatens standard descriptions of our linguistic competence. This is, in some ways, a very odd claim for Davidson to make, for we can expect that an explanation of this ability will figure prominently in the linguist's description of first language acquisition.

The phenomenon is of course ubiquitous. We will often guess at the meanings of new words that we encounter in speech and in literature, with varying degrees of success. I have found, for example, that people systematically misconstrue the word *effete* to mean something like "elitist" or "snobbish." It is quite possible that these people have attempted to extract the meaning of this word from such phrases as "effete intellectual" or "effete aristocrat," spoken in disparaging tones. In such cases, we would say that context was unable to provide any good
clues as to the meaning of the word, and that the meanings that people
reconstructed for this word were correspondingly far off the mark.

In other cases, context will provide us with the initial semantic
integration of a new word, but nevertheless leave us quite unable to use
the word in any meaningful way. Fillmore has described his own struggles
with the word *steeplechase*:

Some time ago I had occasion to realize that I didn't know the
meaning of the word *steeplechase*. I didn't know whether it was

- the name of a kind of place where a certain kind of horse
  racing takes place;
- the name of a sport (so that I could say *I enjoy
  steeplechase*);
- the name of a particular contest within this sport (so
  that one could say *The final steeplechase of the day is
  about to begin*);
- the name of the kind of horse that is trained to race in
  the way we associate with the concept of
  steeplechase—with hedges, wall, and ditches as
  obstacles.

I think I completely understood the brute and institutional
facts underlying the meaning of the word, but I didn't know
how the word "keyed" into the schema.⁴

I take it as a given that learning the meaning of a new word
essentially involves learning its categorial, selectional, and semantic
features.⁵ Each new context in which we encounter the word—including
any session we might have with the dictionary—helps us to assign new
features to a representation of that word in our mental lexicon. It is also possible that new contextual information will cause us to modify previously assigned features that we now take to be incorrect. All of this, of course, happens without any conscious effort on our part, and it behooves the theorist to give an adequate account of this process.

Although I have not indicated this in figure A, I have assumed that our knowledge of the lexicon is connected in some way with our knowledge of the grammar. This is not to say, of course, that other cognitive modules cannot somehow access this knowledge (I will in fact describe one such possibility below). It is, however, one thing to assert that an adequate description of our linguistic competence must make essential appeal to the notion of a lexicon, and quite another to say how it is that we go about constructing such a mentally represented list of words and their idiosyncratic features. The current division of labor would have the grammarian asserting the former, and the psycholinguist attempting the latter.

If, as Davidson appears to claim, it is true that correctly interpreting the meaning of a new word involves a lot of inspired guesswork, then it is also true that we guess incorrectly much of the time. It may take us a number of exposures to a word, in a number of revealing contexts, before we can begin to place the word in the overall schemum of word meanings—and even then we might get it wrong. The meanings of some words may be so recherché that it would be impractical for us to expect normal contextual exposure to reveal their meanings. If
our ability to interpret new words depends in large measure on wit, luck, and wisdom (Davidson's words), then why hold the linguist accountable for an explanation of this ability? It appears to me that we should then also demand from the linguist an explanation of our ability to write clever limericks, create palindromes, and play Wheel of Fortune.

It may appear to Davidson a small miracle that we can interpret new words at all—never mind the fact that it may be a hit or miss affair. In which case I can only respond to Davidson by agreeing that it is no mean task to give an adequate account of this ability. I would also argue, however, that it is impossible to set any a priori limits to progress in this domain. I must again conclude that Davidson is being overly pessimistic when he suggests that this phenomenon threatens standard models of our linguistic competence.

Turning now to our ability to cope with new idiolects, we first need to agree on what exactly constitutes an idiolect that is essentially different from our own. An idiolect is of course the language spoken by an individual, but it is usually construed as a pattern of language use that is unique in some way among the members of that individual's speech community. But even here a wide range of interpretations is possible. On one reading, no two people ever share what is essentially the same idiolect. Their portable speech interpretation and production machines may be the same as ever you please, but any two people will always differ in their speech habits and other matters of performance. On the other hand, we might stipulate that two people have the same idiolect when, for
example, they share the same grammar, or when they share the same grammar and Pragmatics Unit—leaving theory to decide the metric by which we measure "sameness." On this latter interpretation, if two people share the same idiolect, they will essentially agree on all judgments of grammaticality and utterance meaning. At the other end of the sameness metric, we might find the monolingual speaker of English and the monolingual speaker of Walbiri staring at each other mutely.

When Davidson talks about our ability to cope with new idiolects, I assume that he means a new idiolect of what is essentially our own language—more specifically, an idiolect whose features differ in some interesting way from our own. Here too we have a wide range of possibilities.

I have already discussed our ability to recover the meaning of a new word, and this would cover the case of a language user whose idiolect differed from our own only in matter of vocabulary. But what about the fellow who uses whole constructions that sound to us unfamiliar or odd or downright uninterpretable? Let us consider a few examples.

At one time I had occasion to puzzle over the oddity of the sentence

[29] She defies me like Turk to Christian.

The sentence is perfectly interpretable and it makes complete sense in its own context. What I find most interesting, however, is the charming way
that it breaks parallel structure. Instead of expressing the parallelism with the clunkier "She defies me like Turk defies Christian," the author has opted for a more compact and, I would say, more elegant, expression of his thought. In this case I would say that the difference between our idiolects lay in the domain of performance rather than in the domain of competence. The author of [29], being a much better wordsmith than I, chose to exploit his mastery of the simile and of the preposition to's ability to express—among other things—opposition or confrontation.

I recently had the pleasure of running across the following noun phrase in a book by Lynda Barry:

[30] scary pink fish condiment that the smell would make you start running

Ms. Barry was attempting to capture the very unguarded and oftentimes ungrammatical speech of preadolescents (very successfully, I would say). We have probably all at one time or another heard someone use a construction similar to that in [30]. In a child's mouth, we would simply say that it was a performance error that time would correct. But what if we were to find an adult who used this kind of relative clause construction with some frequency? Would we then say that this person was speaking an idiolect fundamentally different from our own?

This last question is tightly wound up with the question of how we go about attributing one grammar rather than another to a language user. If we have independent reasons for believing that our subject's
grammar "rules out" such constructions as [30], then we say that the production of [30] is essentially a performance error, and we are left with the quite different problem of explaining why our subject uses these constructions with as much frequency as he does. If, on the other hand, we have reason to believe that our subject's grammar licenses such constructions as [30], then, depending on how we choose to slice idiolects, we might conclude that our subject's idiolect is essentially different from our own. To put the matter simply, it appears that most interesting differences of idiolect have a grammatical component. That being the case, let us now pursue the subject of ungrammaticality.

Davidson has included our ability to interpret grammatically garbled utterances in his list of Class M phenomena. An ad hominem argument for this inclusion might go as follows. Suppose that person A utters a sentence which is novel to person B, e.g. Mermaids and aquamen alike care nothing for the opera. As good scientists we record the fact that person B has no problem interpreting this utterance: he recognizes familiar words arranged in a novel but nevertheless appropriate way. As linguists we attempt to explain this phenomenon in part by assuming that A and B share a knowledge of the same language, their knowledge being characterized in part by a finite set of structure-forming rules, and licensing and interpretive principles that we call a grammar. Roughly speaking, we construct a grammar in such a way as to capture the regularities of a language. Thus we say that sentences produced by the speaker A are not sui generis, but rather have their point of origin in a human "language faculty," a faculty that is partly characterized by A's
knowledge of the grammar of his language. Hearer B, sharing a knowledge of these rules, has no problem interpreting sentences produced in accordance with them.

What happens, however, when our speaker produces a sentence that is grammatically garbled, or ungrammatical? Suppose, for example, that person A utters

\[\text{It is known George to be silly}\]

and person B unhesitatingly interprets him to mean that George is known to be silly. Most linguists, I believe, would judge [31] to be ill-formed, or ungrammatical. But if that is the case, then how did person B come to interpret the utterance successfully? The problem is this: sentence [31] breaks the grammatical rules that person B assumes person A uses to produce his utterances. How, then, can person B so unerringly recover the speaker's intended meaning? Is this perhaps a bona fide example of our ability to recover meanings that theory cannot deliver?

Two things must be said here. First, this is not an argument that I endorse. I hope to show in §2 that the argument simply will not wash. Second, I am not asserting that this is an argument that Davidson would himself endorse. I am simply taking the part of that philosopher or linguist who is skeptical about rule-governed accounts of language interpretation, and who would point to such phenomena as our ability to interpret grammatically garbled utterances as evidence for his skeptical position. I will henceforth refer to this person simply as the "skeptic."
There are several reasons why our ability to interpret the ungrammatical might seem a stronger challenger to our mechanistic picture than the other abilities in Davidson's list of Class M phenomena. First, recall that grammars have been the traditional bread and butter of linguists; theories of grammar lie at the very heart of explanations of our linguistic abilities. If it now appears that we can interpret a very broad class of sentences (like [31]) without appealing to shared grammars, then the entire enterprise built on these grammars falls into question.

When we correctly interpret the intended meaning of a malaprop, we begin by recovering the literal meaning of the speaker's utterance. The ludicrousness of this literal meaning in the given context acts as a kind of "error signal" that sets off the search for a phonological neighbor of the sentence produced by the speaker. Notice, however, that we interpret sentences like [31] spontaneously and unerringly, oftentimes not even noticing their ungrammatical status. There is, in this latter case, no "error message" that might let us know that something has gone wrong with the speaker's production mechanism. There is no apparent intermediate step in the interpretation of sentence [31] wherein we acknowledge, for example, that because sentence [31] is ungrammatical it therefore has no literal meaning of its own.

Recall that in order to recover the speaker meanings of malaprops and other slips of the tongue we must make essential appeal to mechanisms that lie outside of the SID (the Sentence Interpretation
We must appeal to these mechanisms in part to recover the sentence that the speaker intended to utter. In §2 I will argue that no such intermediate step of "sentence recovery" is essential to the interpretation of the grammatically garbled utterance. Unlike our ability to interpret slips of the tongue, our ability to interpret ungrammatical utterances does not appear to be much different from our ability to interpret the meanings of their grammatical counterparts. That is to say, when we interpret sentence [31] to mean that George is known to be silly, we are in fact recovering its literal meaning. Furthermore, in the case of tongue slips, new words, and the like, it is too easy to write off the skeptical argument as being much too premature: our knowledge of the relevant mechanisms at the pragmatic end of things is rather sketchy, and we are obliged to play a "wait and see" game with the skeptic. Such, I will argue, is not the case with our ability to interpret the ungrammatical, where we appear to have a fairly good—though admittedly tentative—knowledge of the relevant mechanisms.

In the following sections, I will explore the ways that our ability to interpret grammatically garbled utterances might threaten standard descriptions of our linguistics competence. I will, so to speak, give this phenomenon a full run for the money, attempting to show in the process that the mechanistic paradigm is much more powerful than many people imagine.
§2 The challenge of ungrammaticality

I have previously argued that it is not possible to motivate Davidson's skeptical argument on the basis of our ability to interpret malaprops. In this section I will explore how we might motivate the skeptical argument on the basis of our ability to interpret ungrammatical utterances.

The skeptic would like to argue that this ability threatens standard descriptions of our linguistic competence. Standard models of our linguistic competence are built around the notion of a *shared* sense—on the part of both speaker and hearer—of the ways in which sentences of the language can be constructed and understood. This "shared sense" is codified as a *grammar*, which is, quite simply, the linguist's theory of the language-user's linguistic competence. We assume that speaker and hearer somehow access the information codified in this grammar when they produce and interpret utterances and other linguistic signs. (I will again assume that there is no harm in speaking as if speaker and interpreter were using the theory that the linguist uses to describe their competence.) The skeptic now points to the fact that the hearer can unambiguously interpret the speaker even when the speaker produces an utterance which does not "adhere" to the proposed grammatical rules. The skeptic asks whether or not this obviates the need for shared grammars, since it now appears that speaker and hearer can get along without them. Does not this phenomenon compromise the explanatory value of shared grammars in the interpretation of utterances? Do we have here a good
candidate for a case where the interpreter recovers an utterance meaning that theory cannot deliver to him?

The answer to both of these questions, I will argue, is clearly no. In this chapter I hope to show (1) that the skeptic has misconstrued the notion of a grammar and a grammatical "rule," and (2) that modern grammatical theory has made significant progress toward providing an explanation of this phenomenon.

A few preliminaries

A first step in understanding a language user's native ability to produce and interpret ungrammatical sentences is getting clear on the notion of *ungrammaticality*. How do we give a principled account of which sentences are and which sentences are not to count as ungrammatical or ill-formed? We may decide, for example, to leave this distinction entirely to the judgment of native speakers. Language is, after all, a public entity; everything from the rarefied thoughts of the great philosophers to the most playful banter of the schoolyards finds its expression in a language. Why not simply leave it to the tribunal of English speakers to decide which sentences of the language should be classed "ungrammatical"?

Paul Ziff suggested something along these lines in an early paper
titled "About Ungrammaticalness" (Ziff, 1964a). In this paper, he argues that an intuitive idea of grammaticalness is a necessary precondition of our ability to assess the correctness of a grammar. According to Ziff, no grammar is adequate unless it somehow captures this intuition. His characterization of the ungrammatical is as follows:

... a sentence of a language is ungrammatical if and only if first, native speakers of the language balk when an arbitrary token of the type is uttered, and secondly the sentence type has an associated structural description that is classed as nonaccepted on theoretic grounds.\(^7\)

Ziff's definition goes a long way toward capturing our pre-theoretic notion of the ungrammatical. He capitalizes on our tendency to "balk" at sentences whose structure we deem ill-formed or deviant. But he also notices that native speakers of English have a tendency to balk at sentences like

[32] He had a green idea.\(^8\)

This sentence, according to Ziff, is perfectly grammatical: although the thought expressed by the sentence is "odd," its form is nevertheless unassailable. Correspondingly, it is the task of the linguist constructing a grammar for English to insure that this grammar assigns to [32] an acceptable structural description. If the grammar in question assigns to [32] an acceptable structural description, then by Ziff's criteria the sentence cannot be ruled ungrammatical—a desirable consequence.
Ziff's characterization of the ungrammatical has several undesirable consequences, however. Consider the following "deviant" sentence:

[33] More people have been to France than I have.

It is a remarkable fact that this sentence often passes undetected as a piece of pure nonsense. On the dozen or so occasions that I have deployed sentence [33], I have never encountered anything that I might characterize as "balking" behavior: my listeners consistently assumed that I meant something along the lines of "There are others who have been to France—not just me," or "There are people who have been to France more often than I." By Ziff's criteria, therefore, sentence [33] is simply not ungrammatical—a consequence that I suspect few linguists are willing to accept. This suggests a rather tenuous link between speaker judgments of sentence acceptability and the notion of grammaticality that we would want theory to preserve.

The unreliability of speaker judgments can also be seen with respect to what linguists call "garden path" sentences, such as

[34] The horse raced past the barn fell.

This sentence, and others like it, often elicit "balking" behavior: those hearing sentence [34] for the first time assume that the predication is complete once the word barn is pronounced; the final word fell is judged semantically adventitious, and its presence makes the sentence sound "odd." Speakers who make this judgment fail to construe The horse raced
past the barn as a complex noun phrase (a restrictive relative clause, to use the linguistic jargon), and thereby fail to interpret sentence [34] along the following lines:

[35] [NP The horse which was raced past the barn] [VP fell].

Clearly we cannot expect every speaker of a given language to be equally sensitive to the subtleties of sentence construction and the possible pitfalls of sentence interpretation. Perhaps if we restrict grammaticality judgments to a distinguished panel of language users, we might overcome the problems associated with sentences like [33] and [34], and thereby salvage Ziff's characterization of the ungrammatical. The publishers of dictionaries, it will be observed, have already adopted this policy. It is common practice for lexicographers to commission a "usage panel" consisting of novelists, essayists, poets, journalists, and others who have "demonstrated their sensitiveness to the language and their power to wield it effectively and beautifully."9 This usage panel decides, for example, whether it is or it ain't good English to use the word anxious in the sense of eager.

The notion of ungrammaticality that I would like to develop, however, is rather more theory-internal. This notion depends first and foremost on the proper characterization of a grammar and an assessment of its role in explaining various aspects of our ability to acquire and use language. To put the matter as succinctly as possible, I will simply assume that an ungrammatical sentence is one whose structure violates
one or more rules of a given grammar, where the "grammar" in question is of the kind that might be proposed by a linguist of the generative grammar school. I will put aside for the moment the question of exactly which "rules" I mean, or how the structure in question might "violate" these rules. These matters will emerge in the course of my exposition.

Recall, however, that we are restricting our attention to *interpretable* ungrammatical sentences. Thus our list of candidate ungrammatical sentences might include anything from the mildly unsyntactic

[36] To who are you speaking?

where the interrogative pronoun has been given the wrong case, and

[37] The child seems sleeping,

which is interpreted as a raising structure:

[38] \( NP \text{ seems } [t \text{ Predicate}]; \)

to the chaotic

[39] anyone lived in a pretty how town with up so floating many bells down,\(^{10}\)

where it is more difficult to pinpoint the source of the sentence's deviance. I say that our list "might" include the above sentences because in the absence of any widely attested theory of grammar, we can only guess at what this list might include. We will, however, temporarily
exclude from our list such "word salads" as

[40] is of and lying the the Ronald Reagan,

which are simply uninterpretable.

Also excluded from consideration are such sentences as

[41] Colorless green ideas sleep furiously,

and

[42] John solved the pipe.

These sentences are traditionally judged deviant on the basis of their content rather than on the basis of their form. Which rules, if any, are broken by sentences like these depends very much on the internal organization of one's theory of the language faculty. There may be, for example, a level of semantic representation where [42] is assigned a structure that is ill-formed or unlicensed in some way (e.g. a representation where the semantic features of the constituent the pipe do not match the semantic features of the context John solved ...). We may, on the other hand, find that we should rather attribute the oddity of [42] to a clash between the proposition it expresses and our beliefs about pipes. These same considerations may also apply to analytically false sentences like [24], repeated here:

[43] My stepfather is taller than himself.

I will, for the sake of brevity, call these sentences and others like them
unsemantic, their distinguishing feature being the fact that their
semantic oddity coexists with their syntactic well-formedness. I am not
claiming that the unsemantic sentences form a natural kind; the existence
of any such class of sentences may turn out to be a mere epiphenomenon. I
only wish to point out the fact that there are other kinds of sentential
"deviance" not covered by the rubric ungrammatical.

There may, of course, be yet a third way of accounting for the
oddity of sentences like [41]-[43]: it may, for example, turn out that a
properly specified grammar will account for the oddity of these sentences
on purely syntactic grounds. Chomsky addresses this question of the
proper division of labor between syntax and semantics in section 2.3.1 of
Aspects of the Theory of Syntax. There he presents two contrasting
lists of sentences (here numbered [44] and [45]):

[44]  [i]  The boy may frighten sincerity
        [ii]  sincerity may admire the boy
        [iii]  John amazed the injustice of that decision
        [iv]   the boy elapsed
        [v]    the boy was abundant
        [vi]   the harvest was clever to agree
        [vii]  John is owning a house
        [viii] the dog looks barking
        [ix]   John solved the pipe
        [x]    the book dispersed

[45]  [i]  sincerity may frighten the boy
        [ii]  the boy may admire sincerity
        [iii]  the injustice of that decision amazed John
        [iv]   a week elapsed
        [v]    the harvest was abundant
        [vi]   the boy was clever to agree
The expressions in [45] are grammatically irreproachable; there is nothing unexceptional about their form or diction. But most readers will find some or all of the sentences in [44] "odd" or "deviant." Each of the sentences in [44] can be obtained from the corresponding sentence in [45] by a word substitution, an interchange of grammatical function (e.g. moving an NP from subject to object position, and vice versa), or some other simple operation. It should be noted that at the time of writing *Aspects*, Chomsky considered the sentences of [44] to have a rather "borderline character," in the sense that it was not entirely clear how to explain their aberrant status. The issue for Chomsky was the proper division of labor between what he called the syntactic and the semantic components of the grammar. Some deviant sentences, he noted, presented clear cases of the violation of purely syntactic rules, for example:

[46]  sincerity frighten may boy the;

while other deviant sentences were only semantically incongruous:

[47]  I knew you would come, but I was wrong.

The sentences of [44], however, presented a special problem for the theorist. The linguist acknowledged two important facts about these sentences: (1) they were in some way "aberrant," and (2) they were, to some extent, interpretable. With an eye to the internal organization of his theory, the linguist had to decide whether to extend the methods of
syntactic or semantic analysis to account for these facts:

A priori there is no way to decide whether the burden of presentation should fall on the syntactic or the semantic component of the generative grammar. If the former, we must design the syntactic component so that it does not provide for the sentences of [44] directly, but assigns them Phrase-markers only by virtue of their structural similarities to such perfectly well-formed sentences as those of [45] ... Alternatively, if we conclude that the semantic component should carry the burden of accounting for these facts, we can allow the syntactic component to generate the sentences of [45] as well as those of [44], with no distinction of grammaticalness, but with the lexical items specified in such a way that the rules of the semantic component will determine the incongruity of the sentences of [44] and the manner in which they can be interpreted (if at all).14

Thus even at this early stage in the development of generative grammars, the problem of how to account for our ability to successfully interpret ungrammatical sentences was acknowledged. In a later passage, Chomsky assumed that it fell to the syntactic component of the grammar to explain the aberrant status of sentences like those in [44]. I do not wish to give the reader the impression, however, that I take the sentences of [44] to be perfect paradigms of ungrammaticality. This, I have argued, is a theory-internal matter. I will, in the course of my exposition, come to consider a very broad range of "deviant" sentences, some of which have been classed ungrammatical by those linguists whose theories I here endorse, and some sentences whose statuses have yet to be decided.

The skeptic, as I have described him here, is one who believes that standard grammar-based models of our linguistic competence cannot
account for our ability to interpret utterances that are ungrammatical.

Thus we find philosophers like Mario Bunge, for example, who argue that "since ... an ideal speaker-hearer is supposed to handle only grammatical (well-formed) sentences, grammars are prescriptive or normative, contrary to Chomsky's claim and in agreement with classical linguistics." Furthermore, "The function of a grammar is to generate all and only the infinitely many grammatical (well formed) sentences of the language." The roots of this skeptical view have been traced by Chomsky to the early influence that the study of formal languages had on the development of modern linguistics:

... the study of formal languages was misleading in this regard. When we study, say, the language of arithmetic, we may take it to be a "given" abstract object: an infinite class of sentences in some given notation. Certain expressions in this notation are well-formed sentences, others are not. And of the well-formed sentences, some express arithmetical truths, some do not. A "grammar" for such a system is simply some set of rules that specifies exactly the well-formed sentences. ... It is easy to see how one might take over from the study of formal languages the idea that the "language" is somehow given as a set of sentences or sentence-meaning pairs, while the grammar is some characterization of this infinite set of objects, hence, it might be thought, a construct that may be selected one way or another depending on convenience or other extraneous concerns. The move is understandable, but misguided, and it has engendered much pointless discussion and controversy. 17

The "move" discussed by Chomsky is one in which we equate a language with an "externalized" object—a set of sentence/meaning pairs, say. Chomsky attributes this extensional, or "E-language," view of language to
such philosophers as Quine and Lewis, for whom, he argues, the notion of a grammar is essentially derivative. On their account, "the linguist is free to select the grammar one way or another as long as it correctly identifies the E-language." In "On Semantics," Higginbotham tells us that there is an essential element missing from Lewis's description of a language, where "Sentences or their structural descriptions are paired with meanings, or with ranges of meaning, so that what is excluded from the range of meaning of a sentence comprises those things that it does not mean ... ." Higginbotham writes:

Here ... is what I think is missing from Lewis's picture. In that picture, sentences have various ranges of meaning, and some are meaningless, but nonsentences do not have any meanings. However, the last statement is false to natural languages: nonsentences must have definite meanings, as full-blooded as those of ordinary sentences ...

It is sometimes difficult to resist the neat idealization of a language as an infinite collection of well-formed structures. And the view is not endemic only to philosophers. In their Introduction to the Theory of Grammar, Riemsdijk and Williams, two linguists of the Chomskyan school, have written that "A language ... is a set of objects that we call sentences." I have noticed that in beginning linguistics courses it is common practice to ask students to picture a grammar as a set of rules that generate the well-formed sentences of a language. Non-sentences are left out of the picture, except when they are needed by linguists to motivate some grammatical rule or other. The going assumption seems to be that there are no grammatical principles that are obtainable only by considering the interpretation of ungrammatical
In all fairness it should be said that there is some initial plausibility to the view that there exists a correlation between well-formedness and interpretability. We notice, for example, that (roughly speaking) the farther a sentence moves away from full grammaticality, the more difficult it becomes to interpret. We can imagine, for example, taking a well-formed sentence and, one by one, randomly shifting the position of each of its words. At a certain point in the process, our original sentence will evolve into a word string with little or no discernible constituent structure, and the grammatical functions of its words will be lost. There are, of course, many "axes" along which a sentence can approach the limiting case of complete word salad. Not only can we shift the constituents of sentences around in strange ways, but we can also make the appropriate constituents disagree in person, number, gender, tense, mood, etc.; we can assign the wrong cases to pronouns; we can add extraneous words; etc.

Intuitively speaking, it appears that the more "grammatical structure" a sentence preserves, the more likely it is that it will be interpretable. Previous attempts to explain our ability to interpret ungrammatical sentences exploited this intuition. The idea was in some sense to find the "shortest path" between the ungrammatical sentence and one that was fully grammatical, as I will show in the next section.
A little historical perspective

In this section I will briefly consider a former avatar of the problem of ungrammaticality, one that appeared more or less around the time of the printing of Chomsky's *Aspects of the Theory of Syntax*. I will focus on this period for two reasons: (1) There were several papers published around this time that dealt exclusively with the language-user's ability to recover the meaning of ungrammatical utterances. Since then the subject has all but disappeared from the literature (for reasons that will soon become apparent). (2) Previous attempts to solve this perceived "problem" are instructive vis à vis modern efforts to deal with the same phenomenon. I will show how the theorists' proposed solutions to the problem might well have engendered a certain skepticism toward the whole enterprise.

I will consider the problem of ungrammaticality from the point of view of the generative grammarian of the middle 1960s, and I will refer to this linguist as the "linguist," or "theorist."

I should begin by saying that the general picture of the interpretive process which I have outlined in figure A (p. 103) is one that our linguist might have gladly endorsed. Our linguist, however, would have had a different theory of grammar, and this would have colored not only his perception of the problem of ungrammaticality, but also his approach to a solution of that problem. I will, for the sake of discussion, refer to
this early notion of a grammar as \textit{grammar}_e, where the subscript "e" is appended to suggest the word "early." There are two key differences between grammars\textsubscript{e} and their more modern counterparts (as described by the GB school, for example): (1) Unlike modern theories of grammar, grammars\textsubscript{e} had no LF level of representation to act as an interface between syntax and the "semantic component," or, as Lakoff has put it, "between formal grammar and something in the 'real world'."\textsuperscript{24} (2) Unlike their modern counterparts, grammars\textsubscript{e} were very much in the business of "generating" well-formed structures—in a sense that I will make clear below.

Consider first the organization of one of these early grammars. Chomsky used the word "grammar" to refer very broadly to system of rules having three major components: syntax, phonology, and semantics.\textsuperscript{25} The syntactic component of a grammar\textsubscript{e} consisted of an iterative system of rules with two major subcomponents: (1) a \textit{base} that generated a "highly restricted (perhaps finite) set of basic strings, each with an associated structural description called a base Phrase-marker"\textsuperscript{26}; and (2) a transformational subcomponent, "concerned with generating a sentence, with its surface structure, from its basis."\textsuperscript{27} The base Phrase-markers were taken to be the "elementary units" from which deep structures were constructed.\textsuperscript{28} I assume that this description of a grammar\textsubscript{e} is familiar to most readers.

Linguists of the time assumed that it was the deep structure of a
sentence that determined its semantic interpretation, and its surface structure—as yielded by the transformational component of the syntax—that determined its phonetic interpretation.\textsuperscript{29} I think it is fair to say that subsequent research has shown this model to be essentially incorrect; the model was nevertheless successful in guiding linguistic research for quite some time.

The key to our inquiry lies in the relationship between the deep structures and their associated surface structures—this relationship being mediated, on Chomsky’s account—by the rules of the transformational component. Chomsky writes:

What, then, is the test that determines whether a generalized Phrase-marker is the deep structure of some sentence? The answer is very simple. The transformational rules provide exactly such a test, and there is, in general no simpler test. A generalized Phrase-marker $M_D$ is the deep structure underlying the sentence $S$, with the surface structure $M_S$, just in case the transformational rules generate $M_S$ from $M_D$.\textsuperscript{30}

And there is a further condition on the surface structure $M_S$: it must be well-formed:

A deep structure is a generalized Phrase-marker underlying some well-formed surface structure. Thus the basic notion defined by a transformational grammar is: deep structure $M_D$ underlies well-formed surface structure $M_S$. The notion "deep structure" is itself derivative from this. The transformational rules act as a "filter" that permits only certain generalized Phrase-markers to qualify as deep structures.\textsuperscript{31}
The force of the requirement that the surface structure $M_s$ be well-formed is made clearer in another passage where Chomsky describes a possible sequence of steps that we might use to recover the deep structure, and hence the meaning, of some well-formed sentence of the language:

The base rules and the transformational rules set certain conditions that must be met for a structure to qualify as the deep structure expressing the semantic content of some well-formed sentence. Given a grammar containing a base component and a transformational component, one can develop innumerable procedures for actually constructing deep structures. These will vary in exhaustiveness and efficiency, and in the extent to which they can be adapted to the problems of producing or understanding speech. One such constructive procedure is to run through the base rules (observing order) so as to form a generalize Phrase-marker $M$, and then through the transformational rules (observing order) so as to form a surface structure $M'$ from $M$. If $M'$ is well-formed, then $M$ was a deep structure; otherwise, it was not. All deep structures can be enumerated in this way, just as they can all be enumerated in many other ways, given the grammar. 32

Assuming, as Chomsky does, that an ungrammatical sentence $S$ violates one or more rules of the grammar, it may not be possible, using this same sequence of steps that he describes, to recover its deep structure. When we, for example, "run through" the base and transformational rules of the grammar in our attempt to form the surface structure of $S$, which of these rules do we "relax" and how? It was considerations like these, I believe, that led linguists of the time to take such an active interest in our ability to recover the meanings of ungrammatical sentences.
I will now pause to consider some early attempts to bring ill-formed sentences into the sway of linguistic theory.
Noam Chomsky's "Degrees of Grammaticalness," 1961

In this early paper, Chomsky suggests a way that the current theory might be extended to accommodate the phenomenon of ungrammaticality. His approach is outlined as follows:

Given a grammatically deviant utterance, we attempt to impose an interpretation on it, exploiting whatever features of grammatical structure it preserves and whatever analogies we can construct with perfectly well-formed utterances. We do not, in this way, impose an interpretation on a perfectly grammatical utterance.33

Chomsky does not, however, detail how the language user "exploits" the grammatical structure that an ungrammatical sentence preserves, or how he "constructs" the relevant analogies. According to Chomsky, the first step in the interpretation of a deviant sentence \( S \) is the recognition that \( S \) deviates from one or more of the grammatical rules.34 This recognition on the part of the interpreter initiates a series of steps which includes, among others, the construction of the aforementioned "analogies" between the deviant sentence \( S \) and other well-formed expressions of the language. These well-formed expressions in turn provide the key to the interpretation of the deviant sentence \( S \). But there is no description of how the interpreter actually goes about constructing the crucial analogies. It is possible that a description of this process belongs properly to the theory of performance. The ability to construct analogies—that is, to detect structural similarities between sentences—is not necessarily language specific and may depend on more general cognitive capacities.
In this paper, Ziff is concerned with the problem of explaining our ability to understand syntactically deviant sentences. For Ziff, syntactically deviant sentences of English include those sentences exemplifying problems of co-occurrence, as well as sentences traditionally classed as ungrammatical. Before we can understand a syntactically deviant sentence, Ziff claims, we must apprehend the syntactic structure of that sentence. He writes:

Apprehending the syntactic structure of a syntactically deviant utterance is ... a matter of grasping the simplest relation between the utterance and the set of non-deviant utterances. ... A deviant utterance has that structure that constitutes the terminus of the simplest route from the regular grammar to the utterance in question.

In order to make clear the concept of a "simplest route" let us consider one of Ziff's examples:

Consider the utterance "He expressed a green thought": it might seem as though the utterance were deviant owing simply to the combination of green and thought. That this is not the basic reason is indicated by the fact that the class of elements that can occur without syntactic deviation in the environment "He expressed a green ..." is null, e.g. "He expressed a green tree" is also deviant. Hence, the deviance of "He expressed a green thought" cannot be attributed to thought. It can only be attributed to green. Let $E_j$ be the class of elements that can occur without syntactic deviation in the environment "He
expressed a ... thought": then we can relate the utterance to the regular grammar by invoking the rule $E_j \to \text{green}$.

On the other hand, the utterance "Over there is a green thought" is deviant owing to the occurrence of thought. Let $E_j$ (where $i \neq j$) be the class of elements that can occur in the environment "over there is a green ...": then by invoking the rule $E_j \to \text{thought}$ we can relate the utterance to the regular grammar.38

It is rules like

[48] $E_j \to \text{green}$

and

[49] $E_j \to \text{thought}$

in the above examples that map the simplest route from the regular grammar to the syntactically deviant sentence in question. Ziff also presents a number of other possible rules:

I should like to suggest that there are five basic types of routes from the regular grammar to a syntactically deviant utterance. In consequence, there are five types of rules that may be involved; (where capital letters are variables for word classes and lower case letters are variables for words) these five types of rules may be rendered schematically as follows:

[50] ... $A$ ... $B$ ... $\to$ ... $B$ ... $A$ ...
[51] ... $A$ ... $B$ ... $\to$ ... $A$ ...
[52] ... $A$ ... $\to$ ... $A$ ... $B$ ...
[53] $A \to a$
[54] $(A \to a) \mid B \to b$
The first is the rule of inversion, the second of deletion, the third of addition, the fourth of word-class extension, and the fifth of word-class contraction.\textsuperscript{39}

As an example of the use of inversion (rule \textsuperscript{50}), consider the ill-formed string:

\begin{enumerate}
\item \textbf{[55]} The man the cake ate.
\end{enumerate}

According to Ziff, the simplest route from the grammar \(g\) to this syntactically deviant sentence is given by the inversion rule

\begin{enumerate}
\item \textbf{[56]} \(NP_1 \ V \ NP_2 \rightarrow NP_1 \ NP_2 \ V\).
\end{enumerate}

Under this classification of rules, rules \textsuperscript{48} and \textsuperscript{49}, discussed above, would count as word-class extension rules. There is a great deal of flexibility in how we are allowed to define word-classes. In \textsuperscript{48}, for example, the word-class \(E_i\) is the class of those elements that can occur without syntactic deviation in the environment "He expressed a ... thought"—a relatively small class of words. On the other hand, the word-classes used to express rule \textsuperscript{56} are the very broad lexical categories \(NP\) and \(V\).

According to Ziff, finding a route from the regular grammar \(g\) to the deviant sentence is primarily a matter of isolating the source of the sentence's deviance.\textsuperscript{40} Once the source of the sentence's deviance is identified, we can construct a rule or rules (on the model of rules \textsuperscript{50} to \textsuperscript{54}) that map a grammatical string to the deviant sentence in question.

161
Although Ziff does not state this explicitly, understanding the deviant sentence is a matter of understanding the grammatical string to which it is mapped. More precisely, the *meaning* of the deviant sentence is exactly the meaning of the grammatical string to which it corresponds.

Ziff's task is to account for our ability to understand or interpret syntactically deviant sentences. What then is the status of those syntactically deviant sentences that we cannot straightforwardly interpret? Consider, for example, the "ill-formed" sentence

\[57\] My the believes in a just god,

where the anomalous element *the* is being used, not mentioned. There exists a very simple route from the set of fully grammatical sentences to sentence [57]. Consider the rule

\[58\] \( B \rightarrow \text{the}, \)

where \( B \) is the class of all words that can occur in the context "My ... believes in a just god" without syntactic deviation. In this case, we have followed Ziff's prescription very closely: (1) we have isolated the source of deviance for sentence [57], namely, the occurrence of the word *the* in the given context; and (2) we have given a simple rule, [58], that relates the set of fully grammatical sentences to the deviant sentence [57]. The reader will note, however, that we are no closer to understanding [57] than we were before. I will grant Ziff a pinch of salt and assume, for the sake of discussion, that his theory is meant to apply only to the class of *interpretable* deviant sentences.
Now consider the interpretable deviant sentence [37], repeated here:

[59] The child seems sleeping.

When we attempt to isolate the source of [59]'s deviance, we are faced with a dilemma. Is the source of [59]'s deviance (a) the occurrence of the verb *seems* in the context "The child ... sleeping," a context that can accept such verbs as *is*, *was*, and *hates*, or (b) the absence of the infinitival element *to be* before the word *sleeping*, or perhaps even (c) the occurrence of the word *sleeping* in the context "The child seems ...", a context reserved for such words as *happy* and *desperate*? Certainly the fact that we naturally interpret sentence [59] along the lines of

[60] The child seems to be sleeping

would suggest option (b), but this is not how Ziff would have us proceed. According to Ziff, we interpret an ill-formed sentence by finding the "simplest route" from the regular grammar to that sentence. Let us now consider the rules suggested by options (a), (b), and (c) above. Option (a) suggests the word-class extension rule

[61] $C_1 \rightarrow$ seems,

where $C_1$ is the class of verbs that can appear in the context "The child ... sleeping" without syntactic deviation. Option (b) suggests a deletion rule along the lines of
Option (c) suggests the word-class extension rule

\[ C_2 \rightarrow \text{sleeping}, \]

where \( C_2 \) is the class of words (e.g. adjectives) that can appear in the context "The child seems ..." without syntactic deviation. We are now bade to choose the simplest of these rules. But by what criterion of simplicity are we to do so? There is nothing about any of the rules [61], [62], or [63] that mark it as the rule relating the deviant sentence [59] to the regular grammar. Ziff's theory does not appear able to get off the ground.

Jerrold Katz's "Semi-Sentences," 1964

In this paper Katz addresses our ability to interpret the semi-sentences of a language. The class of semi-sentences includes sentences that fall short of full grammaticality as well as sentences that are "semantically ill-formed." Katz does not make clear what he means by a semantically ill-formed sentence. I will assume, for the sake of discussion, that the class of semantically ill-formed sentences includes those sentences that Chomsky labels "unsemantic," e.g. \( I \) \( knew \) \( you \) \( would \) \( come, \) \( but \) \( I \) \( was \) \( wrong \) (sentence [47]), as well as those sentences that result from violations of selectional rules, e.g. \( He \) \( had \) \( a \) \( green \) \( idea \). Nothing hinges on our defining the class of semi-sentences in this way; for our purposes, it is only important that this class include those
sentences that we previously characterized as ungrammatical.

Katz argues that it is incumbent on the linguist to provide what he calls a "theory" of semi-sentences:

A theory of semi-sentences, a theory that characterizes the set of ungrammatical strings that the speaker's knowledge of linguistic structure enables him to understand and explains why the members of this set are comprehensible, is ... to be regarded as an integral part of the description of a language, not as a bonus it is nice but not necessary to have. 42

Semi-sentences, according to Katz, differ as to the extent of their departure from full grammaticality, but nothing is a semi-sentence that "goes too far." 43 Katz therefore seeks to partition the set of ungrammatical sentences into two exclusive and jointly exhaustive subsets: the set SS of semi-sentences and the set NS of nonsense strings. 44 The set of nonsense strings is the set of those strings that depart so far from full grammaticality that they are thereby rendered incomprehensible; they are, in effect, what I have elsewhere called "word salad," a word string in which grammatical relations are obscured and the grammatical functions of the constituents are undeterminable. An example of such a string might be

[64] the the and is smokeless ashtrays.

Strings in the set NS may exhibit some structure—for example, in sentence [64] smokeless ashtrays constitutes a noun phrase—but the structure exhibited may not be enough or of the right kind to render the sentence interpretable. The sentences in the set SS, however, do
preserve enough grammatical structure to enable competent speakers of
the language to interpret them. The proposed theory of semi-sentences
must therefore account for the various ways that a speaker uses his
linguistic knowledge to interpret the sentences of SS.

Katz’s theory proceeds as follows:

A speaker knows (in the sense in which he knows the rules of
the grammar of his language) a system of rules that enables
him to associate a non-null set of grammatical sentences with
each semi-sentence. This association is performed on the
basis of the structure that the semi-sentence has and the
speaker’s understanding of the semi-sentence is nothing other
than his understanding of the sentences in the set in with
which the semi-sentence is associated ... Let us call the set
of sentences that have their meaning(s) transferred to a
semi-sentence its comprehension set and let us call the rules
which accomplish this transference transfer rules.

A system of transfer rules may be regarded as containing at
most a rule for each rule of the grammar (the same transfer
rule may relate to more than one rule of the grammar). Each
transfer rule tells how the rule(s) of the grammar to which it
corresponds can be violated without leading to derivations
whose terminal lines are strings in NS. A theory of
semi-sentences thus generates semi-sentences (without
generating any nonsense strings) when the transfer rules are
used to construct their “semi-derivations.” The comprehension
set for a semi-sentence must be constructible from its
semi-derivation.

Katz’s theory is fairly complex, and it is worth being clear about its
general structure: (1) Each element of the set SS is derivable from a
combination of transfer rules and rules of the grammar. We saw the
reason for this stipulation earlier: the derivability of the sentence will insures that hearer will be able to recover its deep structure. (2) The comprehension set for the derived semi-sentence is mechanically constructible from its semi-derivation. (3) Each element of the comprehension set is a possible interpretation of the semi-sentence in question.

I see several problems with Katz's proposed theory as outlined. Assuming that the members of \( SS \) are in some way identifiable and derivable, why should it be expected that the invoked transfer rules will bear some relation to the normal rules of a grammar_\( e \)? According to Katz, each transfer rule will correspond to one or more rules of the regular grammar_\( e \), and it will indicate how these rules of the grammar_\( e \) can be violated without generating a derivation whose terminus is a string in \( NS \). But how do we know a priori that each element of \( SS \) can be derived by violating (or relaxing) one or more of the normal rules of the grammar_\( e \)? What if relaxing the appropriate rules amounts to little more than the license to "derive" whatever string we want? Furthermore, what licenses the assumption that we will be able to construct the comprehension set for a semi-sentence from its semi-derivation? These are just some of the questions that Katz leaves unanswered.
Summary and discussion

In each of the papers discussed above, an attempt was made to associate an ungrammatical sentence with one or more grammatical, or well-formed, sentences. Chomsky spoke of our "imposing" an interpretation on the ungrammatical sentence by exploiting whatever grammatical structure the sentence preserved, and by using "whatever analogies we [could] construct with perfectly well-formed utterances"; Ziff showed how we might attempt to find the "simplest path" from the deviant sentence to one that was perfectly well-formed; and Katz sketched a theory wherein each "semi-sentence" was associated with a "comprehension set" of well-formed sentences, each sentence of this set providing a possible interpretation of the given deviant sentence.

Behind these attempts to beat a path back to the grammar lay, I believe, something like the following chain of reasoning: (1) The sentence perception device, or parser, would take a well-formed phone sequence or word string as input and deliver as output something that was much "closer" to a surface structure than a deep structure; (2) our knowledge of the grammar would then enable us to recover the deep structure, and hence the meaning, of the given sentence; (3) if, however, the input to the parser was an ungrammatical sentence, and the parser delivered a surface structure that was not well-formed, then we might not be able to use our knowledge of the grammar to recover the necessary deep structure; therefore (4) it was first necessary to "fix up" or repair the input to or the output of the parser, by somehow replacing the ill-formed sentence by
one that was fully grammatical.

Note that both Chomsky and Ziff appealed to abilities that were left further unanalyzed—namely, (for Chomsky) the ability to construct the relevant "analogies," and (for Ziff) the ability to intuit the "simplest route" back to the regular grammar. Our skeptic may well point to these special talents as evidence, for example, that our ability to recover the meanings of utterances is rather less "mechanistic" and rather more a product of "wit, luck, and wisdom" than our theorist will allow—at least in the domain of the ungrammatical.

Katz had a different approach altogether. His suggestion was simply to create a whole new theory of semi-sentences, one that would supposedly work side by side with the theory for well-formed sentences. Katz was therefore in the business of placing constraints on a theory that he was not even sure existed—a move that does not seem well-motivated.

All of these approaches assumed that there was an essential difference between the process of interpreting a deviant sentence and one that was fully grammatical. This assumption was later dropped, as we shall see in the next section.
Current theory

The complexion of grammatical theorizing is very different today from what it was 25 years ago when those early papers on ungrammaticality were published, and I will not attempt to reconstruct all the relevant shifts of focus and methodology. Higginbotham has already expressed what I believe to be the going assumption among theorists, and I will repeat his words here:

... non-sentences must have definite meanings, as full-blooded as those of ordinary sentences, if the source of their intuitive uninterpretability (or merely partial interpretability) is just the violation of a rule of formal grammar. 48

Thus, contrary to Katz's original proposal, you shouldn't need two interpretive mechanisms to accommodate both sentence and non-sentences, but one. There have been a number of significant changes in grammatical theory that have made this consolidation possible, and I will very briefly sketch some of these below.

My aim in this section is to defend the mechanistic account of interpretation against the skeptic who argues that this account cannot explain our ability to interpret deviant utterances. I will show how current grammatical theory can accommodate this ability, and how this accommodation might necessitate a shift in commonly-held notions of a grammar and a grammatical rule.

One of the most salient features of the overall organization of UG
is the addition of the LF, or "logical form," level of syntactic representation—the LF-representation of a sentence being, roughly, a representation of the contribution of the grammar to the meaning of that sentence. It is the task of the linguist to determine which aspects of meaning are represented at LF and how they are represented. In many cases, the LF-representation of a sentence will be the same as its S-structure representation. Consider, for example, the sentence

[65]  George eats poor people for breakfast,

with S-structure representation

[66]  [NP George] INFL [VP eat [NP poor people] [PP for breakfast]].

(The abstract element INFL represents the tense and agreement properties of the verb eat.) Loosely speaking, the S-structure [66] contains all the information relevant to the interpretation of sentence [65]. The noun phrase George is assigned the 0-role agent by the transitive verb eat, and the noun phrase poor people is assigned the 0-role patient; the prepositional phrase for breakfast modifies the verb phrase. The grammatical functions of all the elements in sentence [65] are indicated in the S-structure representation [66], and the mapping from S-structure to LF is correspondingly trivial: [66] is also the LF-representation of sentence [65]. Other sentences do not map so straightforwardly from S-structure to LF. The sentence

[67]  Everybody loves somebody sometime,

for example, is assigned the S-structure
[68]  \([NP \{QP \text{Everybody}\}] \text{INFL} [VP \text{love} [NP \{QP \text{somebody}\}] [\text{ADVP sometime}]]\),

where the QPs indicate that the elements everybody and somebody are quantifier phrases. Sentence [67] is ambiguous with either the universal quantifier everybody or the existential quantifier somebody enjoying wide scope. The quantifier movement (QR) rules of the LF component thus map the S-structure [68] into two corresponding LF-representations,

[69]  everybody\(x\) [S somebody\(y\) [S \(x\) loves \(y\) [ADVP sometime]]]

[70]  somebody\(y\) [S everybody\(x\) [S \(x\) loves \(y\) [ADVP sometime]]],

where, for the sake of clarity, I have omitted the abstract element INFL. [69] and [70] receive their interpretations in accordance with the usual rules of quantifier-variable binding: [69] asserts that each person will love some (unspecified) person sometime; [70] asserts that there is at least one person that everybody will love sometime.

Another salient difference between previous incarnations of generative grammars and their more modern counterparts is the elimination, or perhaps more accurately, the reinterpretation of the transformational component:

In early work in generative grammar it was assumed, as in traditional grammar, that there are rule such as "passive," "relativization," "question-formation," etc. These rules were considered to be decomposable into more fundamental elements: elementary transformations that can compound in various ways, and structural conditions ... that are themselves
formed from elementary constituents. In subsequent work, in accordance with the sound methodological principles of reducing the range and variety of possible grammars to the minimum, these possibilities of compounding were gradually reduced, approaching the rule Move-α as a limit. 49

Chomsky is here suggesting that the rules of the transformational component might essentially reduce to one: Move-α, meaning "move any category anywhere," with some parametric variation on the choice of and possible landing sites for the moved category. Chomsky calls this a "reasonable hypothesis," and further claims that "many particular cases appear well-substantiated." 50 An example of the application of the rule Move-α may be seen in the derivation of the S-structure [72] from the D-structure [71]:

[71] $np$ is believed [George to be an evil capitalist]

[72] $George_i$ is believed [$t_i$ to be an evil capitalist],

where $np$ is a position reserved for a noun phrase, and $t_i$ is the coindexed trace left by the movement of the noun phrase $George$ from the embedded clause to the position occupied by $np$.

One dominant perspective on the organization of UG construes the grammar as a complex system of interacting principles. These principles divide naturally into a number of subsystems:

[73] [a] X-bar theory
    [b] θ-theory
I will assume that X-bar theory is familiar to the reader in one or another variant. Loosely speaking, X-bar theory, together with the $\theta$-marking properties of the lexical entries, determines the categorial constituents of representations at D-structure and their ordering.

$\theta$-theory is concerned with the rules governing the proper assignment of thematic roles.

Case theory is concerned with those conditions that govern the assignment of abstract Case to those elements that appear in Case-marking positions. (When used in this technical sense, the word *case* is capitalized.) Consider, for example, an application of J.-R. Vergnaud's *Case Filter*. This principle states that every noun phrase with phonological content must be assigned Case. The Case Filter explains why the noun phrase *George* in D-structure [71] must be moved to the subject position of the matrix clause: infinitivals are not Case assigners, and the noun phrase must move to a Case-marked position in order to receive Case.

The binding theory is concerned with the relation of anaphors and pronominals to their antecedents.
Bounding theory places constraints on possible applications of the rule Move-α. In particular, there appear to be constraints on "how far" certain elements can be moved in a given structure, and what positions these elements can be moved to.

The theory of control determines possible antecedents for the abstract element PRO. I will have little to say about this theory.

The concept of government plays a central role in most of the subsystems discussed above. Without getting bogged down in the technical details, the theory of government explores the relation between the head of a construction and the elements or categories governed by it. A verb, for example, governs its complements, and in the prepositional phrase [pp P NP], P governs NP.

We can best appreciate the functioning of this elaborate machine by considering an example. Suppose that our speaker were to utter the ill-formed:

[74] *It is believed George to be an evil capitalist.

We interpret the speaker to mean that George is believed to be an evil capitalist. How do we do this? Looking at our Sentence Interpretation Device (figure A), we can begin by making some very minimal assumptions about the structure delivered by the parser. Let us suppose that the parser simply scans [74] from left to right and assigns to it a constituent structure along the lines of:
It is believed [S: [S George to be an evil capitalist]],

where I have suppressed labeling those constituents not strictly relevant to our discussion. By simply "accessing" the lexicon, the parser would know that the verb believe takes a clausal complement, and little else apart from some basic X-bar theory would be needed by the parser to yield a structure like [75]. If we were now to assume that [75] mapped isomorphically to the LF level of syntactic representation, so that [75] was also the LF-representation of [74], we could account directly for the recovered meaning. As an LF-structure, [75] shows that George is assigned the θ-role agent or subject by the embedded verb phrase. The pleonastic interpretation of it would be suggested by the fact that it was not coindexed with any trace in the structure [75]. And this, it appears, would be enough to account for the recovered meaning. Notice that the structure [75] is not licensed by the Case Filter discussed earlier: George is not in a Case-marked position and must therefore move from the subject position of the embedded clause to receive Case. Nevertheless, this violation of the Case Filter does not seem to affect the interpretation of [74]. This would be possible, for example, if there existed general semantic principles that applied to linguistic structures independently of their well-formedness. We can put this observation in the form of a putative operative principle, one which, on the basis of empirical evidence, we might later find cause to modify or drop altogether:

Unlicensed LF-structures can, under certain conditions, still be processed by the semantic component (i.e. interpreted).
I should mention that from the semanticist's point of view, it is simply sound methodology to assume an operative principle of the form [76]. If one is going to specify the semantics for a language in the most general way possible, then one does not a priori want to exclude the semantic processing of syntactically ill-formed structures.

Alternatively, we could have assumed that the pleonastic *it* in structure [75] first loses its phonetic content, yielding a structure like [71], repeated here:

[77]  *np* is believed [S [S George to be an evil capitalist]].

"Normal" grammatical processes (e.g. an application of the rule Move-α) would then convert [77] to the well-formed LF-structure:

[78]  George_i is believed [S [S t_i to be an evil capitalist]].

The choice between these two alternatives, it must be stressed, is not a matter of stipulation, but rather one of empirical discovery.

It is certainly true that our lack of knowledge about the parser and its relation to the grammar makes our account of sentence interpretation somewhat provisional. I think it is fair to say that this relationship is little understood at present. One of the central questions in this domain is the following: To what extent does the parser employ the kind of information that is represented by grammatical rules? A recent proposal by Clifton and Frazier54 has it that identifying the the
grammatical structure of a sentence "is accomplished by identifying and categorizing the lexical items of a sentence as they are read or heard and then analyzing them in terms of the possible phrase structure configurations of the language." It should be noted that a consideration of potential phrase structure configurations is not enough to rule out a structure like [75] as the possible "output" of this parser, unless, of course, we stipulate that constraints on possible structures (like the Case Filter) also be included among the parser's "considerations."

But why mention the parser at all? Chomsky and other linguists have already made it clear that the grammar is not to be taken as a model of sentence production or interpretation. If, as theorists have suggested, the interpretation of ungrammatical sentences is to be treated no differently than the interpretation of fully grammatical ones, then why worry over the details of the parser in the former case any more than in the latter?

Part of the special worry over the parsing or processing of interpretable ungrammatical sentences concerns the fact that they are not "constrained" in the same way that grammatical sentences are. An example will help make my meaning clear. I have already shown how a sentence like [74] might very straightforwardly be assigned an LF-structure that is fully "interpretable." But now consider the ill-formed sentence:

[79] *Love are in we.
When I first encountered this sentence, I interpreted the author to mean "We are in love." Sentence [79] is a very odd kind of poetic inversion, this inversion being more commonly expressed by the sentence, "In love are we." Given our arsenal of grammatical devices we might say that the meaning we recover from sentence [79] is derived from an LF-structure like:

\[ [\text{NP}_{\text{Love}} t_i \ [\text{VP are [in [NP_{\text{we}} t_j]]}]], \]

where the traces mark the positions from which the noun phrases have been moved. Structure [80] is certainly not licensed by the grammar—it is, in fact, monstrous—but we can appeal to our operative principle [76] and say that this does not affect the ability of the semantic component to "interpret" the structure. The parser, for all we know, may simply deliver the well-formed structure:

\[ [\text{NP We} \ [\text{VP are [\alpha in love]]}]]. \]

The issue that I am pursuing here is not the ability of the semantic component to process a structure like [80], but rather the question of how we might come to assign an LF-structure like [80] or [81] to sentence [79] in the first place. We have assumed that the LF-structure of a sentence is what determines its interpretation, and that the parser, in one way or another, accesses the information coded in the grammar to assign to sentence [79] a structure (not necessarily an LF-structure). It appears that on the basis of some very simple knowledge about the lexicon, and a little X-bar theory, our parser might have assigned the structure
to sentence [79]; whereupon we might have recovered the interpretation "Love is in us." We note, for example, that a lack of agreement between subject and verb does not affect the interpretation of such paradigms as

[83] Love are a many splendored thing;

and a pronoun with incorrect case-marking does not affect the interpretation of

[84] He gave we the flowers.

Why, then, do we come to interpret [79] the way that we do?

We might suppose that there is something about sentence [79] that prompts the parser (or the "processor" of the SID) to modify the structure that it would, under normal circumstances, assign to this sentence. The lack of agreement in number between the subject and the verb; the wrong case-marking on the pronoun; and the fact that both these shortcomings can be put aright by exchanging the positions of the noun phrases—all of these factors can co-conspire to determine an LF-structure substantially different from the one that we might expect (viz. [82]) if the parser were simply to consider the structure-forming rules of the grammar. Here we move away from the view that there exists a kind of direct mapping from the processes of the grammar to the rules and heuristics of the parser. It appears that the perceptual model must be "fleshed out" with certain processing mechanisms that have no clear
counterpart in the grammar. Stemberger writes that "Many researchers ... assume a one-to-one relationship between elements of a syntactic theory and the production of utterances," and a review of the literature will show that a similar assumption is often made in the case of sentence interpretation. This kind of perceptual model may work well for a sentence like [74], where simply ignoring a licensing principle can yield the appropriate LF-structure. But it does not appear to work for the more complex case of sentence [79], where other processing rules may apply.

Many questions still remain unanswered in this domain. Why, for example, can we recover a meaning for sentence [79], but cannot do the same for

[85] John Bill loves?

I am assuming, of course, that [85] is not ambiguous as between the two readings "John loves Bill" and "Bill loves John"; I am simply assuming that [85] is quite literally meaningless. I now ask why our sentence perception device cannot assign, for example, a structure like

[86] [NP John] [NP Bill] [VP loves tj],

to sentence [85]. Why not simply construe the noun phrase Bill (or John) as a moved element? There is, after all, a gap after the verb loves to license this assumption.

These considerations, far from posing any kind of threat to the mechanist program, actually help to define it. If, as in the generation of a
structure like [75], the parser or sentence processor must act as if one or more rules of the grammar has been "relaxed," then the theorist must ask what limits there are on the operation of "relaxing a rule of the grammar." It then becomes part of his research program to inquire which rules of the grammar can be relaxed, when they can be relaxed, and how.

If, in order to recover the LF-structure of an ungrammatical sentence, the parser or processor must employ rules or heuristics that have no clear counterpart in the grammar, then the theorist might still profitably ask the extent to which the parser or processor accesses the kind of information that the grammar provides. In the case of sentence [79], for example, the grammar might be able to "inform" the sentence processor that the mismatch of features (number and case) can be put aright by exchanging the positions of the two noun phrases.

Our concern over the parser is, of course, a concern over what has been generally called the "psychological reality" of our linguistic theory. Joan Bresnan (Bresnan, 1983) has suggested that a "realistic grammar" should be "realizable" in the following sense:

...we should be able to define for it explicit realization mappings to psychological models of language use. These realizations should map distinct grammatical rules and units into distinct processing operations and informational units in such a way that different rule types of the grammar are associated with different processing functions.58

If these "realization mappings" were not possible, then, Bresnan argues, "the grammar could not be said to represent the knowledge of the language
user in any psychologically interesting sense. This would, of course, severely compromise the explanatory value of any such grammar vis-à-vis the problem of sentence production and interpretation, and it would be unclear what role, if any, such a grammar would have in our linguistic theory. We might, on this basis, assume that the intuitions of the early grammarians were not far off: if we wish to preserve the value of the grammar in explaining our ability to interpret sentences—ungrammatical or otherwise—then we might suppose that ungrammatical sentences are interpreted in part by first establishing how they have departed from full grammaticality. We assume, of course, that detecting these departures from full grammaticality is an automatic (unconscious) process—a process that it will take no little amount of empirical research to uncover.

We can now focus somewhat our inquiry into our ability to interpret ungrammatical sentences: Supposing that we can recover a (literal) meaning \( M \) from an ungrammatical sentence \( U \), will the SID (cf. Part III, §2) be able to deliver an LF-structure \( S \) for \( U \) that yields recovered meaning \( M \)? We of course want the SID to deliver these LF-structures in some some purely mechanical fashion. We would, in other words, like to avoid a homoncular view of the SID in which, for example, the interpretation of an ungrammatical sentence \( U \) depends on our being able to "intuit" a close grammatical neighbor of \( U \) (leaving further unanalyzed the nature of this "intuition"). This is not to deny that it may be possible to find a mechanism that yields the close grammatical neighbor, but the theorist's account is made all the more plausible if he has no need of invoking these adjunct mechanisms.
It is true, in a sense, that grammar-based theories of our linguistic competence cannot (at present) account for our ability to recover the meanings of ungrammatical utterances. But in the same sense it is also true that these theories cannot account for our ability to recover the meanings of sentences that are fully grammatical. The grammar is merely a repository of our knowledge of the language—a knowledge of the ways it can go right and the ways it can go wrong—and too little is known about the actual mechanisms that access this knowledge and produce interpretations.

As to our skeptic, we must distinguish between two claims that he might be making: (1) Current theory is unable to account for the LF-structures that we apparently recover from ungrammatical sentences; and (2) no future natural extension of this theory will be able to account for the recovery of these meanings. The first claim is certainly true—but then again, no linguist ever denied that it was. The second claim, it appears, flies in the face of the considerable progress that has already been made in this domain.
Part V
Summary
In this paper I set out to explore one of a fairly broad class of skeptical arguments that seek to undermine rule-governed accounts of our ability to produce and understand language. This argument, proposed by Donald Davidson in his paper "A Nice Derangement of Epitaphs," attempts to show that our interpretive abilities reach far beyond what standard accounts of our linguistic competence can explain. Davidson attempts to motivate his skeptical argument by pointing to such phenomena as our ability to recover the intended meaning of a malaprop. This ability on the part of the hearer, he argues, depends not on any internalized "interpreting machine" that the hearer shares with the speaker, but rather on the hearer's possession of a good measure of "wit, luck, and wisdom," which allow him to construct correct "passing theories" of what the speaker's words mean. The hearer's ability to construct proper passing theories is therefore not any different in kind from his ability to find his way around in the world generally.

In Part I of the paper I presented Davidson's argument in abstraction from those views that we might have been tempted to attribute to him on the basis of his other works on language. My aim in that chapter was not to play the part of the philosophical historian, but rather to isolate the kind of skeptical argument that would become the focus of this paper. Davidson's claim was that phenomena like our ability to interpret malaprops threatened what he called standard descriptions of linguistic competence, including, by his own admission, descriptions for which he was himself responsible. (I take this to be an allusion to his championing the use of Tarski's Convention T in descriptions of our
semantic competence.) In this section, I sketched the assumed model of our linguistic competence—what I called the Three Plausible Principles model—and I presented Davidson's objections to it.

In Part II I assessed Davidson's objections to the standard account of our linguistic competence, and I argued that his skepticism was not well-motivated. I attacked Davidson's argument on several fronts: (1) I defended the mechanist framework of linguistic inquiry championed by Chomsky and other theorists of the generative grammar school, arguing that this framework could accommodate the interpretive phenomena upon which Davidson based his skeptical challenge; (2) I suggested that mechanistic accounts of our ability to use language need not be held accountable for every aspect of ability to recover utterance meanings, and that there could in fact be phenomena (malaprop interpretation included) that lay beyond the grasp of this theory; (2) I argued, contra Davidson, that there are no good grounds for extending a skepticism about rule-governed accounts of interpretation from malaprops to utterances in general.

In Part III, I explored some of the details of the mechanist response to Davidson's argument, as well as some of the assumptions underlying this response. Davidson sought to argue that rule-governed accounts of interpretation could not account for our ability to recover certain meanings from speaker utterances. In §1 I explored a little more carefully the notion of "recovering a meaning" from an utterance; and in §2 I "fleshed out" the interpretation machine which I had briefly sketched in
Part II, and which, I suggested, constituted the proper foil to Davidson’s skeptical claims.

In Part IV, §1, I took a closer look at Davidson’s list of "Class M" phenomena—phenomena which he claimed threatened the standard model. In §2 I examined our ability to recover the meaning of a grammatically garbled sentence, one of the phenomena which Davidson claimed threatened standard descriptions of our linguistic competence. After defining the notion of ungrammaticality, I gave a "historical snapshot" of theorists’ attempts to accommodate the phenomenon. I showed how these attempts appealed to abilities that were left further unanalyzed by these theorists, and which therefore provided no real explanation of the phenomenon. These shortcomings, I argued, disappeared with the subsequent development of grammatical theory, in which the interpretation of ungrammatical sentences was put on par with the interpretation of fully grammatical sentences. I showed how several features of the new theory made this consolidation possible. I concluded that our ability to interpret ungrammatical utterances, which, at the end of §1, I had argued constituted the great white hope of the skeptic, fell squarely within the purview of current linguistic methodology, and as such was no real threat to the mechanist program.

I conclude, therefore, that there is very little to recommend the kind of skeptical argument that Davidson has outlined in his Epitaphs paper. If the skeptic were to try arguing along these lines he would face a number of formidable challenges.
It is difficult to see, for example, how the skeptic might successfully argue that language-users can recover utterer meanings that the mechanistic picture cannot account for. It is not clear, at present, that the mechanist can account for any of the utterance meanings that we in fact appear to recover: too little is known about the relevant mechanisms operating at the pragmatic end of the UID (figure A, p. 103). Furthermore, the mechanist's enterprise would appear well-motivated even if all it could do was account for our intuitions about literal meanings.

At the SID-end of our interpretation machine, the skeptic's task looks equally thankless. The entire enterprise of the linguist is geared to providing an explanation of our ability to recover the literal meanings that we do in fact recover. That an ungrammatical string can mean such-and-such is not something that the theorist vaguely hopes his theory of our linguistic competence will illuminate. It is for him a primary datum that will shape and constrain his grammar-based theory of linguistic processes.

I take it to be an interesting feature of grammar-based theories of our linguistic competence that they can so naturally accommodate our ability to interpret the ungrammatical—a feature which, I believe, provides good evidence that our theory is on the right track.

Finally, the existence of a scientifically adequate mechanistic
model of language use is consistent with the existence of interpretive phenomena which fall outside the range of this theory. It is unreasonable to expect that any model will do full justice to a real-world phenomenon as complex and richly variegated as language.
Notes

Introduction

2. Ibid., p. 444.
3. Ibid.
4. Ibid.
6. Ibid., p. 31, note.
7. Ibid.
10. Ibid., p. 437.
11. Ibid.

Part I: Presentation of Davidson's Skeptical Argument

2. Ibid., p. 446.
3. Mrs. Malaprop is a character in R. B. Sheridan's play The Rivals.
5. Ibid.
6. Ibid., p. 434.
7. Ibid.
8. Ibid.
10. Ibid., p. 435.
11. Ibid., p. 436.
12. Ibid., p. 437.
18. *Ibid*.
25. *Ibid*.
27. *Ibid*.
29. The double bracket convention will be used as follows: \[ [s] \] refers to the literal meaning of sentence *s*. Thus when I say that Mrs. Malaprop's interpreter takes her to mean \[ [\text{What a nice arrangement of epitaphs}] \], I mean that her interpreter takes her to mean whatever she would mean literally by uttering the sentence "What a nice arrangement of epithets."
34. *Ibid*.
35. *Ibid*.
37. *Ibid*.
38. *Ibid*.
39. Dummett has a stronger reading of the PT model in mind: "The apparatus of prior and passing theories ... is required to explain the phenomena, of malapropisms and of deviant and unfamiliar uses, that interest him. It is also sufficient to explain linguistic communication in general." Dummett (1986), p. 466, my italics. Given any reasonable
interpretation of the notion of explanation, this is much too strong a reading of Davidson's claims for the PT model.

50. George (forthcoming).

52. Dummett has a much stronger reading of Davidson's argument: "Davidson would like to believe that our whole understanding of another's speech is effected without our having to know anything: 'there is no such thing' as a language, he says, 'to be learned or mastered', and the implication is that there is nothing to be learned or mastered." Dummett (1986), p. 474. Davidson's argument does not, I believe, support this "implication."

**Part II: Assessing Davidson's argument**

11. Ibid., p. 279, italics mine.
12. Some of these are described in Davidson (1984b).
14. At the very least, the speaker does not intend that the hearer should take him to be speaking non-literally.
15. Many thanks to Jim Higginbotham for suggesting this example.
16. Alternatively, Revere might assume that it is most likely that the British troops will come by land, there being only a small chance that they will come by sea.
18. Ibid.
20. How would our semantic theory for a language-user be different if, for example, he belonged to a speech community in which utterances were true only 25% of the time?
21. I am not suggesting that these are the actual features in question.
23. Ibid., p. 438.
25. Ibid., p. 453.
28. Ibid., p. 125, italics mine.
29. Ibid., p. 127.
32. This is somewhat of a deviation from standard usage.
34. Ibid.
35. Irene Heim has pointed out that if the speaker utters [13] and means it literally (that is, if he really does mean to say "Familiarity breeds attempt," for comic effect, say), we nevertheless get an echo of the expression in [14]. It appears that we cannot help making certain phonological associations.
38. Ibid., pp. 231-232.
39. Ibid., p. 233.
40. Ibid., pp. 218 ff.
41. This is much too strong a claim and must be softened somewhat by the addition of ceteris paribus clauses.
43. Ibid., p. 442.
44. Ibid.
45. Ibid., p. 443.
46. Ibid.
47. Ibid., p. 446.
52. Chomsky (1975), p. 43.
55. Ibid., p. 443.
56. Ibid., p. 446.
57. Ibid.
58. Ibid., p. 445.
59. Ibid., p. 444.
60. Ibid.
61. Ibid.
62. Ibid.
63. Ibid., p. 443.

Part III: Some details of the mechanist response to Davidson

2. Ibid., p. 445.
5. Ibid., p. 68.
7. A key question here is whether or not we can motivate the analytic/synthetic distinction at some other level of representation.
11. Ibid.
12. Ibid.
14. Ibid.

Part IV: Class M phenomena

2. Ibid.
12. Ibid., pp. 75-76.
13. Ibid., p. 77.
16. Ibid., p. 124.
18. Ibid., p. 20.
19. Ibid.
21. Ibid., p. 548.
22. Ibid., p. 550.
26. Ibid., p. 17.
27. Ibid.
28. Ibid.
29. Ibid., p. 16.
30. Ibid., p. 138.
31. Ibid., pp. 138-139.
32. Ibid., p. 140.
34. Ibid., p. 385.
35. Sometimes referred to as "type crossings" or "category mistakes" in the literature.
37. Ibid.
38. Ibid., p. 395.
39. Ibid., p. 398. Ziff explains the word-class contraction rule as follows: "... the expression to the right of the vertical line is invoked on the supposition expressed by the expression to the left of the vertical line."
Ziff's examples are none too clear.
40. Ibid., p. 396.
42. Ibid.
43. Ibid., p. 402.
44. Ibid.
45. Ibid., p. 411.
46. Ibid., p. 412.
51. I follow the order of exposition in Chomsky (1982).
53. Assuming that 0-roles can be assigned at this level of representation.
55. Ibid., p. 273.
56. Cummings, op. cit., p. 115.
59. Ibid.
Bibliography


