



National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file - Votre référence

Our file - Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Donald Davidson's Theory of Linguistic Competence

Beata Gallay

A Thesis

in

The Department

of

Philosophy

Presented in Partial Fulfilment of the Requirements

for the Degree of Master of Arts at

Concordia University

Montreal, Quebec, Canada

June 1992

Copyright: Beata Gallay, 1992



National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file - Votre référence

Our file - Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

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

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-80962-0

Canada

ABSTRACT

Donald Davidson's Theory of Linguistic Competence

Beata Gallay

In his most recent paper on language, Donald Davidson made the controversial claim that "there is no such thing as a language." (Davidson 1986) This claim is seen by many philosophers as evidence for Davidson's theoretical conversion vis-à-vis his early views. Davidson spent more than two decades explaining our ability to interpret speakers' utterances. He maintained that a Tarski-type truth theory would provide the truth-conditions of sentences, the knowledge of which would enable us to interpret any sentence a speaker may utter. The knowledge of truth-conditions requires the employment of the Principle of Charity, Davidson's central theoretical tool for interpretation.

In this thesis the view is defended that, contrary to the general opinion, Davidson's theory of linguistic competence did not undergo any radical change; rather, Davidson's recent claim with respect to language is a necessary outcome of his views, consistently held throughout his work. Given his fundamental notions on the interdependency of beliefs and meaning, Davidson was forced to reject the significance customarily attributed to the role conventional and its linguistic manifestation, language plays in interpretation.

*As philosophers, we are peculiarly tolerant of
systematic malapropism,
and practised at interpreting the result*

Donald Davidson

TABLE OF CONTENTS

Introduction	1
 PART ONE THE THEORY	
1 The Tarskian Concept of Truth in Formalized Languages	10
1.1 The Concept of A True Sentence in Formalized Languages	11
1.2 The Concept of A True Sentence in Natural Languages	18
 2 Davidson's Theory of Meaning	 25
2.1 Rejecting Building Block Theories	27
2.2 Extensionality Versus Intensionality	31
2.3 Reference - A Useless Concept	35
2.4 Why Choose A Truth Theory	39
2.5 The Theory of Meaning Relativized To A Specific Language	42
2.6 The Problem of Demonstratives	43
2.7 Recursive Definitions	44
2.8 Metalanguage Versus Object Language	47
 3 The Theory At Work	 50
3.1 Radical Translation	51
3.2 Radical Interpretation	56
3.3 The Principle of Charity	62

PART TWO THE IMPLICATIONS OF THE THEORY

4 The Possibility of Completing The Theory	70
4.1 Responses to Tarski's Objections	72
4.2 Commitment to Extensionality	75
4.3 The Problem of Indirect Discourse	82
5 Rejecting The Concept 'Language': A Necessary Outcome of The Theory	90
5.1 The Empirical Testing of The Theory	92
5.2 Developments in Davidson's Employment of The Principle of Charity	99
5.3 The Rejection of Linguistic Conventions	104
5.4 The Rejection of 'Language'	108
5.5 The Interpretability of The Theory	113
Conclusion	116
Bibliography	118

INTRODUCTION

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.¹

This is probably one of the most frequently quoted passages in current papers concerning the philosophy of language. It contains a "downright astonishing conclusion," an "iconoclasm," to use Ian Hacking's somewhat melodramatic words. The conclusion is astonishing on its own, because it goes contrary to our common sense notions of language. To Ian Hacking (and to many others), it is astonishing also because its author is Donald Davidson, one of the greatest philosophers of language within the twentieth century Anglo-American tradition. Among his peers, Donald Davidson is well known for his efforts to construct a theory of meaning that would depict the kind of knowledge that would enable one to interpret any sentence a speaker may utter. The seemingly controversial conclusion is the result of Davidson's efforts extended over a period of more than twenty-five years to explain our success in communication. Hacking, among many others, claims that although "Davidson is not intending philosophical suicide, ... there is an element of retraction in his paper."²

¹ Donald Davidson (1986) "A Nice Derangement of Epitaphs", in E. LePore, ed. Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson, pp. 433-446. Oxford: Basil Blackwell, 1986. p.446.

² Ian Hacking (1986) "The Parody of Conversation", in E. LePore, ed. Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson, pp. 447-458. Oxford: Basil Blackwell, 1986. p.448.

My goal in this thesis is to evaluate this element of retraction in Davidson's recent conclusion concerning language. Through a very detailed reading of several of Davidson's relevant papers, I intend to show that, although Davidson's philosophy of language did change over a period of more than two decades, the change that had occurred was certainly not into the direction of defeat, or of despair. One of the possible reasons why Davidson may be seen by some as giving up his earlier held views could be if these critics, just like Hacking, would selectively attend to the above quoted part of Davidson's recent conclusion, while ignoring the rest of the relevant paragraph. The reader will forgive me for quoting the entire paragraph as it appears in Davidson's paper, since it is the most authentic way to summarize what, I believe, Davidson's *actual* conclusion amounts to. Here it goes:

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 ther fore 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.³

If Davidson's theoretical goal is defined as explaining how communication succeeds, his conclusion means merely that, according to him, our communicative success is not attributable "in any important sense" to conventions. I must stress here that I do not wish to evaluate whether or not this conclusion is plausible or even correct. What I would like to show in my thesis is, that this conclusion *follows from* Davidson's consistently held

³ Donald Davidson (1986) op. cit. p.446.

views on beliefs and meaning, formulated from the very beginning of his work on language. Davidson retracted his own earlier appeal to conventions precisely because that appeal was inconsistent with his fundamental views on meaning. His recent claims on language are not a sign of defeat or despair, but rather, a necessary outcome of his consistently held theoretical views. In this sense, my thesis is an attempt to defend Davidson's theory against one single *type* of criticism, the one that claims that Davidson's above conclusion is an "iconoclasm" in light of his early versus recent views concerning linguistic competence.

My thesis is divided into two main parts. **Part One: The Theory**, contains an exposition of Davidson's theory of meaning. Davidson proposed a new way of constructing such a theory, mainly because he had found the already existing alternatives inadequate. He identified four conditions leading both to the possibility and the necessity of formulating an adequate theory of meaning for natural languages (NL):

- a) The meaning of any sentence in NL is a function of a finite stock of features contained in the sentence.
- b) NL is learnable by virtue of its compositional nature.
- c) Mature speakers are competent in understanding infinitely many sentences of NL based on their knowledge of a finite vocabulary and a finite set of syntactical rules.
- d) Even though no speaker of NL has ever been exposed to all possible sentences of NL at any given time, all competent speakers of NL are capable of understanding arbitrarily many novel sentences.

Davidson claimed that a Tarskian theory of truth, modified to apply to a given NL, can be used to describe in a finite form the kind of knowledge that would enable a speaker to interpret any sentence of that NL, and as such it would be adequate as a theory of

meaning. The first chapter of my thesis describes Alfred Tarski's theory of truth, in so far as it has been relevant to Davidson's theory of meaning. Tarski's aim has been to construct a theory of truth exclusively aimed at formalized languages, since Tarski's outlook on the possibility of constructing such a theory for natural, everyday languages has been extremely negative. Despite Tarski's pessimism with respect to the outcome of such an endeavour, Davidson had attempted to use the theory of truth as a theory of meaning for NL. The motivating forces behind Davidson's choice of theory are explained in chapter two, together with a detailed description of Davidson's ways to adapt Tarski's theory to suit his purposes. Chapter three describes how the Davidsonian theory of meaning is expected to work. It turns out that, rather than defining the meaning of sentences of a language in advance, the theory is expected to define the relevant truth-conditions that prevail on actual occasions of utterances. The theory of meaning amounts to a theory of interpretative action, dynamically geared to the occasion. The Principle of Charity is invoked as the central posit in Davidson's theory, without which interpretation is unthinkable to him. This normative principle prescribes that the interpreter assume that the speaker holds true the sentences uttered.

Part Two: The Implications of The Theory, contains a critical analysis of the theory. It borrows the outline suggested by Davidson himself in the form of three questions:

1. Is it reasonable to think that a theory of truth of the sort described can be given for a natural language?
2. Would it be possible to tell that such a theory was correct on the basis of evidence plausibly available to an interpreter with no prior knowledge of the language to be interpreted?
3. If the theory were known to be true, would it be possible to interpret utterances of speakers of the language?⁴

Davidson himself had answered these three questions that essentially deal with the *possibility of completing* the theory, with its *testability* against evidence, and with its *interpretability* as a theory of meaning, respectively. Even though my aim is not to evaluate Davidson's answers to these questions in detail, or possible objections to Davidson's answers, I find it convenient to follow the sequence of issues raised by them. In chapter four, I critically analyse a selection of Davidson's answers to the first question, with the particular purpose in mind, namely, to demonstrate, that Davidson, far from being pessimistic about the possible completion of the theory, made ingenious attempts to solve the problems inherent in his choice of an extensional truth theory. In chapter five I finally come to the actual defence of my thesis, namely, that Davidson's rejection of the concept 'language' was a necessary outcome of his consistent adherence to his fundamental views on meaning; since it is extremely important for me to show how the argument works, here is my somewhat lengthy summary of Davidson's moves that I believe support my thesis:

⁴ *Inquiries* 131.

Some homophonous sentences do not have the same meaning, thus, the mere uttering of homophonous sentences does not prove that the speakers share a language. If the mere fact that both speakers utter homophonous sentences is not sufficient to prove that they speak the same language, the assumption of their sharing a common language cannot be the relevant principle allowing the interpretation of one another's utterances. Thus, it is not the case that two speakers interpret the meaning of one another's utterances by virtue of sharing a common language. Davidson maintains that if it is supposed to play a decisive role in interpretation, the assumption of a shared language had to be proven, and that the proof of that assumption would be extremely difficult to obtain. There are two implications of the above conclusion: 1) Interpretation of one another's utterances can in principle succeed *regardless* of whether or not the speakers' sharing the same language can clearly be established. That is, if interpretation is successful, it is not because of the assumption of a shared language between speaker and interpreter. This claim weakens customarily held notions with respect to the usefulness of language as a theoretical tool aiding interpretation, and at the same time, it invokes the Davidsonian notion that interpretation is always radical; 2) Another principle allowing the possibility of interpretation must be sought after.

The relevant principle, both allowing the possibility of interpretation and providing evidence for the correctness of the theory is the Principle of Charity. The Principle of Charity is a normative principle prescribing that in interpreting a speaker's uttered sentence, the interpreter assume at first that a speaker's sentence is true when the speaker

holds it true, and moreover, that the speaker uses methods of reasoning, similar to those of the interpreter's in deciding under what conditions to hold a sentence true. However, the Principle of Charity as stated above could be misinterpreted as not allowing the attribution of false beliefs and their consequences, mistake or error in speakers' utterances as well as in listeners' interpretations. There has been a phase in Davidson's theory when, in explaining why the Principle of Charity would not exclude error, he appealed to linguistic convention, defined as a set of publicly standardized beliefs concerning the use of sentences within a speech-community. It is, he claimed, relative to the publicly standardized beliefs that mistaken meanings can be made intelligible.

More recently, however, Davidson rejected the significance of the role convention plays in interpretation, and together with convention, he also rejected the concept 'language', customarily taken as the linguistic manifestation of convention. I shall argue that his earlier appeal to convention *had to be* rejected by him because it was inconsistent with his fundamental views on the interdependency of meaning and beliefs. Davidson's recent rejection of the role of convention and of its linguistic manifestation, language, is not a defeatist move, but, on the contrary, it is strong evidence for Davidson's commitment to his consistently held views on beliefs and meaning. The relevant underlying principle, providing both the evidence for the correctness of the theory, as well as allowing the possibility of linguistic interpretation can be found in Davidson's more recent refinements of the Principle of Charity. If the interpreter assumes that the sentence is a sincere expression of what the speaker believes, this leads her to attribute beliefs to

the speaker that agree with her own, where the agreement is reached through their *shared rationality*. It is the speakers' capacity to reason in similar ways that enables them to communicate successfully. The assumption of widespread agreement in the interlocutors' rationality explains how interpretation succeeds despite idiosyncracies in persons' linguistic behavior. It is the assumed common background of rationality, prescribed by the Principle of Charity, against which both agreement concerning beliefs and meaning as well as disputes and mistakes can be interpreted.

I hope to convince the reader that Davidson's theory of interpretation, *taken in its entirety* can be read as a consistent, plausible explanation of our communicative success.

PART ONE

THE THEORY

CHAPTER ONE

THE TARSKIAN CONCEPT OF TRUTH IN FORMALIZED LANGUAGES

Alfred Tarski's theory of truth⁵ has provided the basis on which Davidson's theory of meaning was built. It is therefore important to give a short outline of what Tarski's theory is about. Tarski's main goal is to provide a *materially adequate* and *formally correct* definition of the term 'true sentence' (what he precisely means by 'materially adequate' and 'formally correct' will be explained within the discussion of his theory). The concept of truth is understood by Tarski as correspondence with reality in the classical, Aristotelian sense⁶; further, this concept is to be interpreted relative to the language in question. Two points are of particular interest here: first, the actual details of Tarski's theory in so far as they have been relevant for Davidson's purposes; second, Tarski's insistence that a correct definition of the concept of truth can only be construed for formalized languages and not for natural, everyday languages⁷. Tarski emphatically restricts his analysis to formalized languages (FL) with a limited stock of grammatical forms, since, according to him, it would be impossible to construct a correct definition

⁵ I am using two of Alfred Tarski's papers that deal with this subject: the earlier, quite technical one, entitled "The Concept of Truth in Formalized Languages" (CTFL for short), first published in the Polish original in 1933, in English in 1956, and the more recent "The Semantic Conception of Truth and the Foundations of Semantics" (SCTFS for short), which is aimed mainly at philosophers, first published in 1944. For complete references see in the Bibliography, Tarski (1933) for CTFL and Tarski (1944) for SCTFS. Page numbers in the notes refer to pages of the volumes that contain these articles as cited in the Bibliography.

⁶ Tarski himself refers to Aristotle's Metaphysica in SCTFS 49.

⁷ Tarski's arguments to that effect have been partially ignored by Davidson, or rather, Davidson did not seem to share Tarski's worries concerning some seemingly insurmountable difficulties with regard to everyday language.

of the concept of truth for either formalized languages with a richer stock of grammatical forms, or for "colloquial" or natural languages (NL). "In that language [NL] it seems to be impossible to define the notion of truth or even to use this notion in a consistent manner and in agreement with the laws of logic."⁸ I shall first outline Tarski's steps to provide a definition of the concept of true sentence in formalized languages, and after that I shall look at Tarski's arguments against the possibility of constructing a truth theory for NL.

1.1 The Concept of A True Sentence in Formalized Languages

It is necessary to start the discussion of the concept of true sentence in FL with some definitions. A language is said to be *formalized* if in specifying its structure one refers exclusively to the form of the expressions involved. Three components specify the structure of the Tarskian formalized language (a and b are formal components, while c is a semantic definition):

- a) *Primitive symbols*
the class of all words or expressions considered meaningful in FL (must be unambiguously defined)
and
the so-called "primitive terms" of the language that are assumed without being defined.
- b) *Rules of formation*
must be given for defining complex expressions. Clear criteria are needed to distinguish the class of expressions to be considered as sentences.
- c) *Definition*
of the conditions under which a sentence of the language obtains its truth-value.

⁸ CTFL 153.

Theorems are provable sentences of a language. In an FL theorems are the only sentences that can be asserted. Tarski stresses that "the problem of the definition of truth obtains a precise meaning and can be solved in a rigorous way only for those languages whose structure has been exactly specified."⁹ In the case of NL, says Tarski, the meaning of the problem is so vague that its solution can only be approximate.

In the introduction to this chapter it was said that Tarski's aim was to provide a *materially adequate* and *formally correct* definition of the concept 'true sentence' in FL. The schematic form of his intuitive definition of a true sentence, which Tarski calls the form (T), is as follows:

(T) x is true iff p

where p is a place holder for the sentence in the extension of the truth-predicate 'is true', and x stands for the name of the sentence. The logical relation between the two sentences ' x is true' and p is that of *equivalence*. The definition of the term 'true' is considered *materially adequate* by Tarski if this term is used in such a way that all equivalences of the form (T) can be asserted (i.e., all such equivalences are implied by the definition). The substitution of any particular sentence of FL for p and of the name of that sentence for x will yield a partial definition of truth that explains the truth-conditions for the given sentence only. The general definition of the concept 'true sentence' in FL will result from the logical conjunction of all particular instances of the form (T). In the following I shall describe the criteria of a *formally correct* definition.

⁹ SCTFS 52.

The semantic notion of truth as in 'true sentence' expresses a property of certain expressions of a language, namely, of sentences. Consequently, a semantic truth-theory will be *about* sentences of a language. The language, about which the theory will make various assertions is called the *object language*. Theory construction in general being a linguistic endeavor that requires the use of a language, a semantic truth-theory, too, will *use* a language in which to make theoretical assertions about the object language. The language used to "talk about" the object language is called the *metalanguage*.

When the aim is to provide the correct definition for the concept of true sentence in FL, a metalanguage, separate from the FL in question has to be employed. The previously defined criterion of material adequacy explains in what sense this metalanguage has to be separate from the object language FL. It has been stated that the correct definition must imply all the equivalences of the form (T) in FL. The definition itself and all the equivalences implied by it are part of the theory, so that these have to be stated in the metalanguage. However, as the symbol p in the form (T) stands for any arbitrary sentence of FL, any of these sentences of FL must be translatable into the metalanguage of the theory, or, in other words, the metalanguage must contain the object language as a part. Moreover, the metalanguage of the theory has to be rich enough to provide a name for every sentence of the object language (x in form (T) stands for the name of the sentence). Finally, since the form (T) contains the general logical term 'if, and only if', the metalanguage must obviously contain terms of logic. All terms used in the metalanguage must be clearly and unequivocally defined.

Tarski points out that the metalanguage in its logical part must be "essentially richer" than the object language. This means that the metalanguage "must contain variables of a higher logical type than those of the object language."¹⁰ Were this condition of "essential richness" not satisfied, the object language could be used to interpret the metalanguage, which would result in contradictions.¹¹ However, according to Tarski, just in case the condition of an "essentially rich" metalanguage holds, this will be not only *necessary*, but at the same time *sufficient* for the construction of a correct definition of 'true sentence' in FL.

It has been stated that the general definition of the concept 'true sentence' in FL will result from the logical conjunction of all particular instances of the form (T). It follows that whenever a language contains infinitely many sentences, the general definition constructed according to form (T) would have to have infinitely many conjuncts. In FL there is a solution for this apparently serious complication, as in FL the sense (or meaning) of every expression is uniquely determined by its form (which is a concatenation of a finite number of primitive terms). Tarski suggests the use of a recursive method to define ways in which all sentences of FL could be constructed (the recursive method gives precise definitions of the atomic parts of FL, and of the rules connecting the atomic parts into sentences; this method provides a finite set of rules, with the help of which an arbitrary number of sentences of FL can be generated). Since FL

¹⁰ SCTFS 55.

¹¹ This is particularly problematic in NL. See the detailed discussion of *the antinomy of the liar* on p.22 of this chapter.

has the characteristic feature of *compositionality*, i.e., simple expressions are combined into composite ones where the truth or falsity of composite sentences depends exclusively on the truth or falsity of the simple ones contained in them, after having divided the simple sentences of FL into true and false ones through partial definitions, the operations necessary for combining the elementary sentences into composite ones can be described using the recursive method. The notion of sentential function in FL can be described through the recursive method, by first defining sentential functions with the simplest structure (containing no logical connectives) and then indicating the operations by means of which compound sentential functions can be constructed from simpler ones (e.g., using 'and' to construct a conjunction of simpler sentential functions). However, as Tarski points out, in general, composite sentences are in no way compounds of simple *sentences*. Sentential functions do in fact arise in this way from elementary functions, i.e., from inclusions; sentences on the contrary, are certain special cases of sentential functions. For example

(1) x sees y

is a sentential function of the *open sentence* kind, thus, it is neither true, nor false. On the other hand,

(1*) John sees Mary

is a sentential function of the *closed sentence* kind, and it is either true or false.

Expressions such as these:

(2) x is white

(3) x is greater than y

(4) x is between y and z

contain *free variables* x , y , and z . Sentences are a special class of sentential functions that contain *zero free variables*. Truth does not enter the structural-descriptive theory until the free variables are replaced by constants. In view of this fact, "no method can be given which would enable us to define the required concept *directly* by recursive means."¹² The introduction of a new concept is required in order to get from open sentential functions to closed ones, i.e., to sentences, that fulfils the following conditions: it must be applicable to any sentential function, be recursively definable, and, when applied to sentences, lead directly to the concept of truth. These requirements are met by the concept of *the satisfaction of a given sentential function by given objects*. An informal definition of the concept of *satisfaction* is provided by Tarski as follows: given objects satisfy a given function if the latter becomes a true sentence when we replace in it free variables by names of given objects. The free variables in the previous examples can be replaced by the following names of objects:

- | | | |
|------|--|---|
| (2*) | snow is white | snow/ x |
| (3*) | a mountain is greater than a hill | a mountain/ x ; a hill/ y |
| (4*) | the number 3 is between
the number 4 and the number 5 | the number 3/ x ;
the number 4/ y ;
the number 5/ z |

The formal definition of the concept of satisfaction is given through the recursive method, starting with the statement of conditions under which certain objects satisfy the simplest sentential functions, and then defining the conditions under which compound

¹² CTFL 189.

sentential functions are satisfied. A concrete example will help to clarify the above statement. The sentential function Fx is satisfied by an object iff that object is F . So, if F is interpreted as 'is red,' then Fx will be satisfied by any object iff that object is red. If G is the two place predicate 'loves,' then Gxy will be satisfied by any ordered pair of things such that the first loves the second. So it will be satisfied by $\langle \text{Romeo, Juliet} \rangle$, but not by $\langle \text{Iago, Othello} \rangle$. The interpretations for predicates specified which objects satisfy these simple sentential functions. The axioms for satisfaction of complex sentential functions can now be given. For example, something satisfies $\neg Fx$ iff it does not satisfy Fx . Grass satisfies $\neg Fx$, since it does not satisfy Fx , where F is interpreted as 'is red.' The general formal definition of the concept of satisfaction will automatically apply to those specific cases of sentential functions that do not contain any free variables, namely, to sentences. For example, the sentence $(\exists x)Fx$ says, given the previous interpretation of F , that there is something red. This is a true sentence, as long as there is something red, that is, as long as there is something to satisfy the corresponding sentential function Fx . As long as something satisfies Fx , everything will satisfy the sentence $(\exists x)Fx$, and if nothing satisfies Fx , then nothing will satisfy $(\exists x)Fx$. Since x is bound by the existential quantifier, it does not matter what value is assigned to it, as long as there is something that satisfies Fx . At this point the definition of 'true sentence' in FL can be given as follows: "a sentence is true if it is satisfied by all objects, and false otherwise" (since a closed sentence has zero free variables).¹³

¹³ SCTFS 56.

1.2 The Concept of A True Sentence in Natural Languages

Tarski discusses the plausibility of formulating the correct definition for the concept of a true sentence in NL: he contemplates *semantic* and *structural* definitions and in both cases his final conclusion is negative. In search of a semantic definition of the concept of true sentence in NL, Tarski goes about it in the following way: intuitively,

- (5) a true sentence is one which says that the state of affairs is so and so, and the state of affairs indeed is so and so

or in a more general, schematic form (the form (T))

- (6) x is a true sentence iff p

where the symbol x stands for any individual name of the sentence, and the symbol p for the sentence itself.

Tarski mentions two ways in which the individual name of any particular sentence (substituting for the symbol x) can be depicted. For example, as an individual name for a particular sentence, a so-called *structural-descriptive name* could be used, where the words constituting the expression denoted by the name are broken down into the signs of which each single word is composed and the order in which these signs and words follow one another is given. For example, scheme (6) applied to the expression 'It is snowing', employing the structural descriptive name of the expression would look like this:

an expression consisting of three words, of which the first is composed of the two letters I and Te (in that order), the second of the two letters I and Es (in that order), and the third of the seven letters Es, En, O, Double-U, I, En, and Ge (in that order), is a true sentence if and only if it is snowing.¹⁴

Or, one may use the so-called *quotation-mark names* to name individual sentences, consisting of quotation marks, left-, and right-hand, and of the expression between the marks, depicting the object (in this case the object being the sentence) denoted by the name in question. Example:

'It is snowing' is a true sentence if and only if it is snowing.

Tarski points out various difficulties encountered by attempts to correctly indicate names for individual sentences for the purpose of formulating a general semantic definition of true sentence in NL.¹⁵ Problems arise from the use of quotation-mark names, as every quotation-mark name, when taken as a syntactically *simple* expression, becomes a constant individual name of a definite expression (enclosed by the marks) and functions *quasi* as the proper name of that expression. According to this interpretation, no generalizable semantic definition can be constructed using scheme (6) by replacing the symbol *x* with any quotation-mark name, as no substitution for constants is permitted. This way, the complete definition of 'true sentence' for NL had to be given by the conjunction of all possible sentences, and considering the possibility of generating infinitely many sentences in NL, this project could never be completed. In order to avoid

¹⁴ CTFL 157.

¹⁵ I admit that the following paragraph is an extremely dense summary of Tarski's objections. For details, see CTFL 159-164.

this problem, Tarski suggests an alternative solution, namely, to interpret quotation-mark names as syntactically *composite* expressions. A full quotation-mark name such as '*p*' would be regarded then as a function, where *p* is a sentential variable that can be substituted by names of individual sentences.¹⁶ However, Tarski calls our attention to the fact that no precise and clear meaning can be given as to the sense of such quotation-functions. Moreover, certain quotation-expressions may prove ambiguous (i.e., one could not tell whether the quotation-name '*p*' is to be interpreted as a function with *p* as a variable argument or as a constant name denoting a particular letter of the alphabet). In light of the encountered difficulties in depicting names for sentences, Tarski holds the possibility of constructing a correct semantic definition for true sentences in NL "extremely doubtful."

The alternative suggested by Tarski is to define the concept of true sentence in NL in *structural* terms. Such an approach would mean that the truth or falsehood of sentences would be defined by considering their relevant structural properties, such as for example the logical connectives contained in the sentences. With the help of sentential calculus "sufficiently numerous, powerful and general laws" would be drawn for sentences in NL, based on which a general structural definition could be reached. However, given the vagueness inherent in NL, says Tarski, NL, in its colloquial form, is not amenable to formalization because of the difficulties in defining the correct logical form of its sentences. The seriousness of this problem is best demonstrated by quoting Tarski's own

¹⁶ CTFL 161.

passage at length, in which he suggests reforms NL had to undergo before formal semantical methods could be applied to it:

...the concept of truth (as well as other semantical concepts) when applied to colloquial language in conjunction with the normal laws of logic leads inevitably to confusions and contradictions. Whoever wishes, in spite of all difficulties, to pursue the semantics of colloquial language with the help of exact methods will be driven first to undertake the thankless task of a reform of this language. He will find it necessary to define its structure, to overcome the ambiguity of the terms which occur in it, and finally to split the language into a series of languages of greater and greater extent, each of which stands in the same relation to the next in which a formalized language stands to its metalanguage. It may, however be doubted whether the language of everyday life, after being 'rationalized' this way, would still preserve its naturalness and whether it would not rather take on the characteristic features of formalized languages.¹⁷

I have left the most important difficulty discussed by Tarski until last, since this one provides the decisive proof for Tarski in support of his claim, namely, that "the very possibility of a consistent use of the expression 'true sentence' which is in harmony with the laws of logic and the spirit of everyday language seems to be very questionable, and consequently the same doubt attaches to the possibility of constructing a correct definition of this expression."¹⁸

¹⁷ CTFL 267.

¹⁸ CTFL 165.

The problem, known as *the antinomy of the liar* arises, when (using a quotation-mark name) the quotation-expression itself contains the term 'true sentence', which is also a part of the definition.¹⁹ If the sentence in question happened to be:

(7) (7) is not a true sentence

then the T-sentence would look like this:

(T) '(7) is not a true sentence' is a true sentence iff (7) is not a true sentence

where '(7) is not a true sentence' is the name of sentence (7). It is empirically evident that symbol (7) denotes the sentence that we named '(7) is not a true sentence'. That is, the sentence denoted by symbol (7), and named '(7) is not a true sentence', *says about itself* that it is not a true sentence. A contradiction such as the following can now be demonstrated:

(8) (7) is a true sentence iff (7) is not a true sentence

Tarski explains the source of this and other antinomies in NL by pointing out NL's characteristic feature, namely, the feature of *universality*. The universality of NL is understood as the possibility of using NL to say things about sentences of the same NL, i.e., the possibility of *using* NL to describe the same NL sentence *mentioned* in the course of the description. In the above example, English has been both *used* as the language in which the example was described, as well as it has been *mentioned*, in the sense that the sentence (7) *itself, about which* explanatory statements were made was an English sentence.²⁰ Part of the explanation was to generate a name for (7) and that was done

¹⁹ Tarski discusses the case of *the antinomy of the liar* in both papers; see CTFL 158, and SCTFS 52.

²⁰ This distinction between the various purposes served by NL will prove crucial in Davidson's attempts to construct a truth theory. See pages 47; 70; 93 of this thesis.

also by using English. Further, semantic expressions such as the term 'true sentence' were used in English. Tarski calls a language with such properties "semantically closed" (the properties being that the language, in addition to its expressions, also contains the names of these expressions, as well as semantic terms such as 'true,' referring to sentences of that language, - sentences, that may contain the same expression 'true' as their component part). In order to show how the problem of antinomy arises, ordinary laws of logic have been employed. English has been *used* to state the contradiction, and *mentioned* in terms of the English sample sentence (7), about which the contradiction was established. The universal character of NL has been shown to produce a contradiction in terms of the laws of logic, and so, NL was found *inconsistent*.

NL's two important characteristics, namely, its *generativity* (the possibility to generate arbitrarily many new sentences), and its *universality* (the possibility to generate statements about its own sentences) are emphasized by Tarski as the greatest obstacles that prevent a successful theory formation as to the correct definition of the concept 'true sentence' in NL. We shall see next, in the discussion of Davidson's theory, how the generativity of NL became Davidson's chief motivation for seeking the possibility of constructing a formal theory of truth for NL. The problem of universality in NL, while acknowledged by Davidson, was not given the same serious consideration as by Tarski. Davidson was confident that Tarski's truth theory for formalized languages (with some necessary modifications) could be used for the theory of meaning for NL. Tarski's concept of truth in formalized languages has found new application in Donald Davidson's

theory of meaning for natural languages. The following chapter will introduce Davidson's reasons for considering Tarski's theory as a viable option for the fulfilment of his plans.

CHAPTER TWO

DAVIDSON'S THEORY OF MEANING

Davidson's main concern in his philosophy of language²¹ has been to explain what it is for a theory to give a satisfactory account of linguistic competence, i.e., the ability to interpret arbitrarily many novel sentences in NL. A NL may contain an infinite number of sentences. Given that, at any point, the number of sentences one has *actually* used or understood is *finite*, there will always be an *infinite* number of potential sentences which would be *new* to the speaker of the language, but which, when actually encountered, she is capable of understanding and using immediately. One could not do so, Davidson argues, unless the new sentences were composed in familiar ways from familiar words. Therefore, the meaning of any sentence must be determined by the meanings of the words that make up the form of the sentence.

The fact that a competent speaker of NL, after having acquired at most a finite vocabulary and a finite set of rules of NL, is capable of understanding and using arbitrarily many novel sentences points towards the need (and at the same time the possibility) for a formal theory of meaning for NL. A further argument for the need of constructing a formal theory for NL stems from the fact that NL is learnable. Only if a finite statement of how the meaning of an arbitrary sentence depends on its elements drawn from a finite stock is given, allowing the possibility of a constructive account for

²¹ Donald Davidson, Inquiries Into Truth and Interpretation, (1986) Since my main source of reference in the discussion of Davidson's theory of meaning will be this volume, containing several of Davidson's relevant papers, in the following chapters, notes referring to this volume will be indicated by the short form '*Inquiries*,' followed by appropriate page numbers.

the meaning of all sentences of NL, can the language described by the theory be considered *learnable*.²²

Davidson identifies two basic requirements a satisfactory theory of meaning has to fulfil: First, it must be powerful enough to provide an interpretation for any sentence in NL, i.e., a successful theory of meaning would have to tell us, what it is we need to know that would enable us to successfully interpret any novel utterance a speaker of NL might make.²³ Second, the theory must be empirically testable against evidence, without detailed knowledge of the speaker's beliefs, and independently of the linguistic concepts of the language.²⁴ Even though he refers to the theory to be construed as a theory of meaning, Davidson eventually discards the classical notion of meaning as reference, as not having any "demonstratable" explanatory value for his theory.²⁵ Instead, Davidson claims that a Tarski type truth theory will function as an adequate explanation of our

²² The condition of learnability is directly related to the compositional character of NL. Davidson sums this up in the following way:

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

²³ Davidson stresses that his aim is not to describe what *actually* goes on in a person's mind when interpreting spoken language, 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." *Inquiries* 125.

²⁴ *Inquiries* XIII; see also *Inquiries* 215.

²⁵ *Inquiries* 20-21. The assumption of a strong connection between meaning and reference is fundamental to classical theories of meaning.

linguistic competence.

This chapter is divided into three large blocks: the first one deals with Davidson's explanations as to why classical theories of meaning are inadequate, the second one introduces Davidson's reasons for favoring a Tarski type truth theory as a theory of meaning in NL, and finally, the third block will describe Davidson's ways to adapt Tarski's theory to suit his own purposes. The global outline of the first block, concerning Davidson's rejection of the classical theories of meaning and leading to the adoption of a Tarskian truth theory is as follows.

a) Rejecting Building Block Theories

While in agreement with the notion of *compositionality*, Davidson rejects the so-called *building block theories* based on the notion of *meaning as reference*, as these type of theories are inadequate in explaining how the meaning of a sentence is a function of the meaning of its parts.

b) Extensionality Versus Intensionality

Still considering meaning as reference, problems with *intensionality* arise that can be avoided only by adopting a purely *extensional* approach to meaning.

c) Reference - A Useless Concept

The concept of reference cannot fulfil the two basic requirements for a satisfactory theory of meaning.

d) Why Choose a Truth Theory

Meaning as reference, not having any demonstratable use, is discarded. Enter the compositional, axiomatizable, extensional, empirically verifiable truth theory.

2.1 Rejecting Building Block Theories

Davidson agrees with the widely held assumption that the fact that, having mastered a *finite vocabulary* and a *finite set of rules* enables us to produce and understand

arbitrarily many sentences of a natural language, must be explained by reference to the *compositional* character of NL: in that sense, the meaning of a sentence in NL depends upon the meaning of expressions contained in it.²⁶ However, Davidson disagrees with the classical theories of meaning about how the compositional character of the language plays a part in conveying meaning in NL. In the so-called *building block theories*, to use the Fregean metaphor, the explanation of a complex linguistic phenome P begins with an account for the simple linguistic phenomena Q contained in P.

The smallest building blocks (or atomic parts) of sentences in NL are for example names, predicates, connectives, quantifiers etc., called 'words' in general. The building block theory makes use of the notion of meaning as reference, that is, the meaning of every word is to be defined by pointing out the object to which the word refers; in general, the schematic form of such a definition looks like this:

(1) *t* refers to *x*

where *t* is replaced by the structural description of the word to be defined, and *x* is replaced by that word itself pointing out the referent object. For example: the word composed of the letters t-h-e-o-r-y refers to theory.

Davidson argues against the purported usefulness of the building block interpretation of meaning as reference, and his specific target in "Truth and Meaning"²⁷

²⁶ *Inquiries* 8.

²⁷ *Inquiries* 17-36.

is Frege's account of complex singular terms. Frege assigns a reference to each significant atomic part of any sentence in NL in such a way that the reference of any complex expression is determined as the value of a function for some given argument or arguments. Davidson's first objection is that the reference of a complex term can equally be explained without assigning meaning to the constituent atomic parts of the term. For example, the expression 'the father of Annette' refers to x iff x is the father of Annette. Nothing is added to this definition, according to Davidson, by an assignment of the Fregean concept of referent to 'the father of', i.e., the assignment of referents has no explanatory value in the case of incomplete ("unsaturated") expressions.²⁸

Davidson's second objection concerns Frege's treatment of sentences as complex singular terms. For the purpose of an adequate theory of meaning, Davidson holds that it is useless to assign meaning to all atomic parts of a sentence in isolation. For Davidson, the meaning of sentences arises from a mutual semantic affect between the sentences and their constituents. The atomic parts acquire meaning only by virtue of their being embedded in sentences and in return, the meaning of sentences results from the systematic contribution of meaning by the atomic parts occurring in them. Thus, mere postulating meanings of atomic parts in isolation would not explain how the meanings of sentences are generated from them. The trouble is that the machinery of function and argument which is supposed to explain the construction of sentences cannot do so.

²⁸ 'unsaturated expressions' in Frege's terminology are equivalent to Davidson's open sentences or to Tarski's sentential functions containing free variables to be replaced with constants; e.g., 'x loves y'.

Davidson illustrates the difficulty by considering the sentence 'Theaetetus flies': "Given the meaning of 'Theaetetus' as argument, the meaning of 'flies' yields the meaning of 'Theaetetus flies' as value."²⁹ This so-called analysis remains completely vacuous, Davidson claims, because we cannot know what the sentence parts yield, unless we already know what they mean. But if we ask what the argument and what the function should be (given in the subject and the predicate term, respectively), we are again referred to the complete sentence. The explanation is blatantly circular, in the sense that fixing the reference of a given expression presupposes an account of its meaning that is prior to the way in which the expression is actually used. In other words, the referent that is supposed to be the meaning of the term, cannot be picked out without already knowing what the term means.

While the notion of compositionality will prove indispensable in Davidson's theory of meaning, he rejects the atomic view of building block theories (where the focus is on the component parts of the composite sentences, i.e., on the *what*) and adopts instead a *holistic* view (where the focus is on the structural relationship of the component parts, i.e., on the *how*) within the domain of the semantics of NL. The building block approach does not assign any special significance to language as a whole; on this account, language is an aggregate: once one had the reference worked out, atomically, so to speak, one could go on to put together sentences, the truth-conditions of which would be worked out on the basis of the references of the parts. In a holistic theory of meaning, language as

²⁹ *Inquiries* 20.

a totality is crucial, for meaning is seen as defined by its structure.

If sentences depend for their meaning on their structure, and we understand the meaning of each item in the structure only as an abstraction from the totality of sentences in which it features, then we can give the meaning of any sentence (or word) only by giving the meaning of every sentence (and word) in the language. Frege said that only in the context of a sentence does a word have meaning; in the same vein he might have added that only in the context of the language does a sentence (and therefore a word) have meaning.^{30,31}

2.2 Extensionality Versus Intensionality

Another difficulty with the notion of meaning as reference is that pointing out the referent of an expression is not always a sensitive enough procedure to express the precise meaning of that expression. It could happen, and quite often it does happen in NL that two or more expressions share a common referent, and so their meaning, if derived exclusively from their referent would be the same. However, common sense observation tells us that sometimes such expressions may not mean the same thing. The famous

³⁰ *Inquiries 22*. Davidson has been directly influenced by his mentor, Quine's views, expressed among others in Quine's paper, "The Two Dogmas of Empiricism", see Quine, (1951). The Fregean dictum mentioned by Davidson originates from Frege's The Foundations of Arithmetic. Many philosophers after Frege and under Frege's influence have taken up Frege's principle. John Wallace (1977) lists among others Wittgenstein's Tractatus and Philosophical Investigations, Russell's theory of definite descriptions, and Ramsey's paper "Universals", as some of the major works influenced by Frege's principle: "only in the context of a sentence do words have any meaning."

³¹ It is interesting to observe that dictionaries such as the Oxford Dictionary, in their attempt to interpret the meaning(s) of words or of other significant syntactical features in NL, are making use of the holistic approach, insofar as they define the meaning of a word by providing sample sentences of NL, in which the word in question is used meaningfully. The underlying assumption is that the reader of the dictionary has already mastered the NL in question, to the extent where she is capable of understanding the overall meaning of these sample sentences, and thus, by virtue of her understanding the sentence as a whole, she will be able to figure out the meaning of the word that is new to her. In a sense, the reader of the dictionary demonstrates a reverse order of linguistic understanding based on compositionality, as opposed to the atomic view of building block theories: She is capable of understanding novel parts of sentences, based on her understanding of familiar, previously encountered sentences. This method obviously will have no applicability in case the person consulting the dictionary has no mastery of the NL in which the dictionary has been written.

example of 'Morning Star' versus 'Evening Star' illustrates this point.

There was a scientifically "benighted" era when people were not aware of the simple fact that the star sighted in the morning and named Morning Star was the same star seen in the evening and named Evening Star. The star thus observed is in fact the planet Venus, and so, both expressions refer to Venus, i.e., share their referent. If meaning is taken to be reference, the following argument could be made:

- (2) 'Morning Star' refers to Venus (in the morning)
- (3) 'Evening Star' refers to Venus (in the evening)
- (4) Venus in the morning is the same planet as Venus in the evening

Therefore, the meaning of both 'Morning Star' and 'Evening Star' is the same.

Common sense tells us that, while observers of the "benighted" era would have held the above conclusion blatantly false, (because of their lack of knowledge as to the premises of the argument), even we, in the possession of such knowledge feel that something is strange about the conclusion. The explanation to that intuition lies in the theoretical distinction between *extensional* versus *intensional* meaning.³²

The *extension* of an expression is its referent. The extension of both expressions 'Morning Star' and 'Evening Star' is the planet Venus. In other words, the two expressions are *co-extensive terms*. In adherence to the notion of compositionality, the extension of a complex expression can be given by determining the extensions of its

³² This distinction is made (among others) by Gottlob Frege, in "Sense and Nominatum", see Frege, (1892).

component parts. That is, given the extensions of the parts of a complex expression, one can uniquely determine the extension of the complex expression itself. If the meaning of a sentence is considered in purely extensional terms, then by definition, the extension of a sentence remains unchanged if we substitute for any of its constituents co-extensive terms, and the truth-value of the sentence remains constant.

Accordingly, if the sentence

(5) The Morning Star is Venus

is true, then so is

(6) The Evening Star is Venus

The *intension* of an expression could be conceived of as the thought, or sense it expresses. Accordingly, the expressions 'Morning Star' and 'Evening Star' express different senses, namely, that of Venus observed in the morning versus Venus observed in the evening, respectively. Since the people of the "benighted" culture did not know that both expressions had the same referent, namely, Venus, they failed to interpret these two terms as co-extensive. Instead, they have taken the two different expressions to have separate meanings according to their respective senses, and that is the reason why they would have rejected the previous conclusion. In the following the particular relevance of the intensional-extensional distinction to the Davidsonian theory of meaning will be discussed.

In general, a plausible proposal concerning a theory of meaning for NL could be that it issue in true theorems of the form:

(7) s means that p

where s is replaced by a structural description of a sentence in NL and p is replaced by the entity that gives the meaning of that sentence. In other words, if we suppose that meanings are the referents of expressions of the form 'that p ', then a theory which entailed all and only true instances of (7) for sentences of NL, would yield the extension of 'means'. There is, however, a problem with this way of formulating a theory of meaning. If one attempts to extend the notion of meaning as reference to sentences, a serious difficulty arises, due to the distinction between the intensional versus extensional meaning. In the purported definition of meaning

(7) s means that p

the sentential operator ' s means that...' does not seem to allow the substitution of co-extensive terms *salva veritate*, i.e., without affecting truth value.

For example, from

(8) 'six is three times two' means that six is three times two

and

(9) 'six' and 'the number of continents' have the same extension

it does not follow that

(10) 'six is three times two' means that the number of continents is three times two

since 'six' and 'the number of continents' do not express the same concept, and therefore

(10) is clearly false.³³

Even though the evidence seems to point towards the plausibility of the notion of meaning as intensions, Davidson rejects it, and instead, he commits himself firmly to construct a theory of meaning with the exclusive reliance on extensions.³⁴ In fact, as it will be shown below, the whole notion of meaning in the classical sense "falls out" of use in Davidson's theory as deemed entirely useless.

2.3 Reference - A Useless Concept

Let us reiterate the two requirements Davidson wishes to impose on an adequate theory of meaning in his own words:

I propose to call a theory a theory of meaning for a natural language *L* if it is such that (a) knowledge of the theory suffices for understanding the utterances of speakers of *L* and (b) the theory can be given empirical application by appeal to evidence described without using linguistic concepts, or at least without using linguistic concepts specific to the sentences and words of *L*.³⁵

³³ The solution offered by Frege (1892) is that in such intensional contexts, a term takes as its referent what is normally its *sense* (its intension). Intensional meanings are objectified, that is, these meanings become somewhat obscure objects, to which words and sentences can refer. This solution allows Frege to continue to hold that co-extensive expressions are everywhere substitutable *salva veritate*. Davidson, sympathetic to Quine's views on nominalism, rejects the positing of ill-defined things, such as meanings. Quine's major objection (1951) against the entities called meanings is that it is notoriously difficult to individuate them, i.e., to tell, when they are different from, or the same as one another. As it turns out, apart from the obscurity surrounding meanings as entities, Davidson is confident in claiming that his theory of meaning can get by entirely without them; see *Inquiries* 22-23.

³⁴ Davidson seems to believe that the relevant truth-conditions of sentences will sufficiently indicate the different intensional senses in which sentences are being used. See chapter four of this thesis for further details.

³⁵ *Inquiries* 215.

Summarizing Davidson's arguments against the notion of meaning as reference so far discussed yields the following list.

- the meaning of complex terms containing predicates can be defined by determining which objects *satisfy* the given predicate (i.e., without having to define the referent of the predicate).
- assigning meaning to the atomic constituents of a sentence in NL (by pointing out their referent) will not explain how the meaning of the sentence will depend on its structure and constituents.
- substitution of co-extensive terms affects truth-value, so that the notion of reference cannot fully account for the meaning of sentences.

In light of these arguments the notion of meaning as reference does not fulfil the first of Davidson's requirements for the theory of meaning, as it appears either *useless* or *insufficient*.

Elsewhere, Davidson adds further arguments as to why he is "reluctant to live with" the notion of meaning as reference.³⁶ Apart from considering the hopelessness of either providing a non-linguistic characterization of the concept 'reference', or reducing it to another semantic concept, Davidson seriously doubts that the concept of reference is "*the*, or at least one, place where there is direct contact between linguistic theory and events, actions, or objects described in non-linguistic terms."³⁷ The above claim refers to the second requirement vis-à-vis the theory of meaning, namely, to the verifiability condition. Davidson, following in his mentor Quine's footsteps, considers the notion of

³⁶ See especially his papers entitled "Reality Without Reference" and "The Inscrutability of Reference" in *Inquiries* 215-241.

³⁷ *Inquiries* 219. The italics in the quote are Davidson's.

reference *inscrutable* and as such, not amenable to empirical verification. This means the following.

An adequate theory of meaning for NL must include an account of truth, - i.e., a statement of the conditions under which an arbitrary sentence of NL is true. Evidence for or against a theory of truth comes in the form of facts about events or situations in the world. Events or situations in the world are what they are, that is, their ontology is assumed to be fixed.³⁸ Facts about the world cause the speakers of NL to assent to or dissent from each sentence uttered by them. However, it is possible for speakers of NL to apply alternative reference schemes to the same facts.³⁹

In Davidson's example, in (an arbitrarily chosen) reference scheme (a) the name 'Wilt' refers to Wilt and the predicate 'is tall' refers to tall things; in the alternative reference scheme (b) 'Wilt' refers to the shadow of Wilt (assuming of course, that everyone has one and only one shadow) and 'is tall' refers to the shadow of tall things.

³⁸ While Davidson agrees with Quine with respect to the inscrutability of reference, they disagree about the fixedness of the ontology of the objects in the world. This disagreement is discussed in detail in "The Inscrutability of Reference"; (see *Inquiries* 227-242), however, for the purposes of my thesis it is not necessary to elaborate on that issue.

³⁹ In Quine's words: "... two men could be just alike in all their dispositions to verbal behavior under all possible sensory stimulations, and yet the meanings or ideas expressed in their identically triggered and identically sounded utterances could diverge radically, for the two men, in a wide range of cases." See Quine, (1960) *Word and Object*, p.26. The Quinean connections of Davidson's theory are especially relevant in terms of its verifiability, and will be discussed in greater detail in chapter three. Quine's views on the indeterminacy of reference are echoed in Wittgenstein's remarks about the ambiguity of pointing in his *Investigations*. Are we to interpret the outstretched arm and extended finger as pointing in the direction from shoulder to fingertip or in the reverse direction, or along some zigzag path? Neither the devices of ostension nor the devices of description are uniquely attached to objects.

According to scheme (a), the sentence 'Wilt is tall' is true iff Wilt is tall; according to scheme (b), 'Wilt is tall' is true iff the shadow of Wilt is the shadow of a tall thing. We recognize that the truth-conditions defined by the empirical evidence are equivalent for the alternative schemes (a) and (b). It cannot be decided on purely empirical grounds, which reference scheme is being employed by the speaker who utters the sentence 'Wilt is tall.' The interpreter is only able to tell that the speaker's schemes are different from one another (and probably different from the interpreter's own schemes), but she will never be able to pick the uniquely correct scheme matching the speaker's words to objects. Reference is not an empirical feature of the utterance. Davidson gives here an early hint of his underlying assumption (that later became one of his arguments for rejecting language as a useful theoretical concept) that an interpreter noticing that the speaker uses a different reference scheme must assume that the speaker is speaking a language different from that of the interpreter's, rather than attributing irrationality to the speaker, in the sense that she does not know what she is talking about. "...some languages are identical in that their speakers' dispositions to utter sentences under specified conditions are identical. There is no way to tell which of these languages a person is speaking."⁴⁰ If such is the case, the theory of meaning as reference cannot be given empirical application.⁴¹

⁴⁰ *Inquiries* 239.

⁴¹ *Inquiries* 235.

2.4 Why Choose a Truth Theory

Here we arrive at the second large issue of this chapter: Davidson's reasons for favoring a Tarski type truth theory as the theory of meaning for NL. Given the constraints a satisfactory theory must have according to Davidson, the notion of meaning as reference has been found insufficient, not testable by purely empirical means and in general useless for the purpose of explaining linguistic competence. Therefore, the notion of meaning as reference ('*s* means that *p*') is discarded as a possible candidate for his theory. Davidson "sweeps away" the obscure 'means that', by offering a "simple, and radical" solution: his suggestion is to treat the

position occupied by '*p*' extensionally: to implement this ..., provide the sentence that replaces '*p*' with a proper sentential connective, and supply the description that replaces '*s*' with its own predicate.⁴²

The result is

(T) *s* is T if and only if *p*.

Here is how it follows: For the "proper sentential connective", Davidson chooses the material biconditional 'if, and only if', which, when connecting two sentences *P* and *Q* will yield a true compound sentence if *P* and *Q* are either both true, or both false, and a false compound sentence otherwise. Since in (T), *s* stands for the *name* of a sentence (of the object language) only, and not for a sentence itself, the name *s*, in order to become a sentence, will have to have a predicate attached to it. This is accomplished by Davidson through adding the predicate 'is T' to *s*. The result, '*s* is T' becomes now the left-hand side or sentence *P* of the material biconditional. On the right-hand side of the

⁴² *Inquiries* 23.

material biconditional, p stands for Q , the sentence of the object language itself, or for a translation of it, as in Tarski's form (T), or, as Tarski himself called it, *Convention T*. Using Tarski's truth-predicate 'T' is largely a matter of convenience. Though, according to Davidson, the choice has certainly something to do with the recognition that all competent speakers of a NL have an intuitive notion of what it means for a sentence to be true. This commonsensical familiarity with the concept of truth proves quite practical when the empirical testability of such a theory becomes an issue.⁴³

What Davidson expects from his theory is that it should specify information such that if someone had this information she would be in the position to understand NL, and in order to understand NL it suffices to know what it is for a sentence in NL to be true.⁴⁴ Before discussing the specific details of how to adapt the Tarskian truth theory for NL, I shall summarize Davidson's reasons for choosing a truth theory for his theory of meaning.

a) The concept of truth can be used to explain what our sentences mean.

Reference cannot be the source of the explanatory import of a theory of truth no matter how one would attempt to construe it, because the only way to test any such construal were to trace its effects on the truth value of sentences. If meaning lies in the difference something's being true or false makes, then it is in the definition of truth conditions that the connection between speakers and the world must be made.

⁴³ *Inquiries* 23-24.

⁴⁴ *Inquiries* 24.

- b) **The truth theory is axiomatizable and preserves the notion of compositionality.**

"From a finite set of axioms it is possible to prove, for each sentence of the language to be interpreted, a theorem that states truth conditions for that sentence. Further, the proof of such a theorem amounts to an analysis of how the truth or falsity of the sentence depends on how it is composed from elements drawn from the basic vocabulary."⁴⁵

- c) **The truth theory is extensionally correct.**

Since the totality of T sentences uniquely determines the extension of a truth predicate for a given NL, a theory that entails all such sentences must be extensionally correct.⁴⁶

- d) **The truth theory is able to provide lawlike generalizations.**

The generalizations of any successful theory must be lawlike. It will not do simply to provide a list of causal chains between the *actual* phenomena to be explained by the theory and its theoretical representation; the success of the theory lies in its ability to support counter-factuals. The theory must be able to tell us how the meaning of a sentence would be defined by its truth-conditions, were certain causal relations between speakers and the world to hold. For a theory in which the generalizations are not lawlike, has no predictive power, and without predictive power, we have no explanation of anything.⁴⁷

- e) **The truth theory accounts for learnability.**

Our ability to use and understand an infinite number of sentences on the basis of a finite vocabulary and limited semantic experience suggests a form of theory that has a finite number of axioms which give the semantic properties of the words of NL and the semantic effects of their permissible modes of combination, which together entail an infinite number of further theorems which give the semantic properties of all the sentences of NL.⁴⁸

⁴⁵ *Inquiries XIV.*

⁴⁶ *Inquiries XV.*

⁴⁷ *Inquiries 171-179.*

⁴⁸ See for example *Inquiries 1-15*. Davidson argues that it is a necessary feature of a learnable language that "it must be possible to give a constructive account of the meaning of the sentences in the language"; and that "a learnable language has a finite number of semantical primitives". As meaning taken

f) The truth theory is empirically verifiable.

Being able to interpret a sentence is knowing that that sentence is true iff some condition obtains and that that sentence is composed out of parts which feature in other sentences which are true iff other specified conditions obtain.

Davidson, despite Tarski's pessimistic outlook as to the possible application of the truth theory for natural languages, has chosen the Tarski style truth theory to provide the basis on which his theory of meaning can be built. While aware of the problems that had to be resolved if the theory were to be completed, Davidson seemed very optimistic with respect to the success of his endeavor. In the remainder of this chapter, I shall describe the ways in which Davidson has planned to adapt the Tarskian truth theory for the purpose of providing a theory of meaning for natural languages.

2.5 The Theory of Meaning Relativized To a Specific Language

The theory of meaning for Davidson, just like the theory of truth for Tarski, has to be relativized to a specific language, since it is obvious that there can be *homophonic* sentences in two different natural languages, the meaning, i.e., their respective truth-conditions of which are different. Davidson uses the example of the sentence in German 'Empedokles liebt' (translated as: Empedocles is in love) versus the English sentence

as truth value is something that is available to (shared) public scrutiny, a truth-functional semantics seems a plausible, even though not exclusive candidate for a theory of meaning.

'Empedocles leaped' to demonstrate the case of such homophony.⁴⁹ Consequently, when constructing T sentences for, say, English, these will have to be understood as stating

(11) s is true in English, iff p

2.6 The Problem of Demonstratives

One of the necessary modifications to be implemented in the Davidsonian theory of meaning deals with an integral feature of NL, namely, its use of demonstratives. These include words such as 'I', 'here' and 'now'. Since demonstratives cannot be eliminated from NL "without loss or radical change", the theory of meaning for NL must accommodate them. The problem with demonstratives is that they cause the same sentence to turn out true at one time/place or spoken by one person, and false when spoken at another time/place and/or by another person. The suggestion to treat demonstratives as constants is rejected by Davidson on the grounds that assimilating demonstratives to constant terms obliterates their unique feature, namely, that part of understanding demonstratives is knowing the rules by which they adjust their reference to circumstances. Instead, Davidson's proposal is "to view truth as a relation between a sentence, a person, and a time".⁵⁰ Thus, new phrases must be added systematically to the theory of meaning, corresponding to each expression with a demonstrative element, in order to relate the truth-conditions of sentences in which the expression occurs to

⁴⁹ *Inquiries* 98. We shall see later the central importance of Davidson's observation, namely, that the mere utterance of homophonous sentences by two speakers is not enough evidence for assuming that they speak the same language. See chapter five of this thesis.

⁵⁰ *Inquiries* 34.

changing times and speakers. For example

- (12) 'I am 26 years old' is true as (potentially) spoken by p at t iff p is 26 years old at t .

It should be noted that in the case of sentences containing demonstratives, the right-hand side of the biconditional will no longer merely depict the sentence named on the left or a translation of it. Nevertheless, Davidson is confident that the resulting "relativized" T sentence will allow the used sentence on the right to give the meaning of the named sentence on the left. The use of demonstratives shifts Davidson's attention from sentences (semantic components of language) to utterances (behaviour), which can be interpreted as an early hint with regards to the weakening of the concept of 'language' as a useful theoretical tool in explaining communicative competence (see also chapter five of this thesis for further details).

2.7 Recursive Definitions

In order to "get the theory working", Davidson requires of a theory of meaning

that without appeal to any (further) semantical notions it place enough restrictions on the predicate 'is T' to entail all sentences got from schema T when 's' is replaced by a structural description of a sentence of L and 'p' by that sentence. Any two predicates satisfying this condition have the same extension, so if the metalanguage is rich enough, nothing stands in the way of putting what I am calling a theory of meaning into the form of an explicit definition of a predicate 'is T'.⁵¹

The theory of meaning (in the form of an *explicit* definition of the predicate 'is T') for a specific NL, for example for English, would simply have to list the truth-

⁵¹ *Inquiries* 23.

conditions of all sentences in English like this:

- (13) 'snow is white' is true in English if, and only if snow is white.
- (14) 'grass is green' is true in English if, and only if grass is green.
- (15) 'university life is sometimes strenuous' is true in English if, and only if university life is sometimes strenuous.

and so on, *ad infinitum*, until all (true) sentences of English were included.⁵² The totality of such sentences would then uniquely determine the extension of the truth predicate. However, to list all (true) sentences of English would yield no insight into the structure of the language and would thus provide no demonstration of how the meaning of a sentence depends on the systematic contribution of its parts. In order to exploit the compositional character of English, the theory of meaning for English would have to be characterised *recursively*. The recursive method works like this: first, the axioms depicting the semantic properties of the atomic components of the sentences must be formulated in extensional language. For example

- (16) 'theory' denotes theory
- (17) 'Julliet' denotes Julliet

The truth of a sentence will depend on the extensions of the component expressions, but since none of the parts of a complex sentence in NL needs to be a sentence itself, the truth of a complex sentence could not simply be accounted for in terms of the truth of its parts. Tarski has solved this problem (that is equally present in the case of FL), by defining truth via the relation of satisfaction.⁵³ For Davidson, too,

⁵² Naturally, the left side of the T-sentences substituting for *s*, would not look that simple; instead, it would contain the "canonical notation" of the sentence in question, using first-order quantificational logic. Canonical notation for certain types of sentences in English has proved notoriously difficult. See chapter four dealing with specific problems of the theory.

⁵³ see p.16 of this thesis.

the recursive characterization of satisfaction must run through every primitive predicate.⁵⁴ Davidson calls a Tarskian sentential function such as ' x loves y ' an "open sentence" (it contains the free variables x and y), that will be satisfied by a function f just in case the entity that f assigns to x loves the entity that f assigns to y (e.g., Romeo loves Juliet). Accordingly, the base clause of the recursion will have to specify what things satisfy atomic open sentences.

(18) 'is boring' applies to all and only things that are boring

(19) 'loves' applies to all and only ordered pairs of which the first member loves the second etc.

The recursive portion of the theory will specify for each kind of complex sentence how the things which satisfy a sentence of the kind are determined by the things which satisfy its immediate components. If the sentence is "closed", i.e., if it does not contain free variables, then it must be satisfied by every function just in case it is true, and by no function, if it is false. The semantic theory of truth tells us, how closed sentences are constructed from open ones by the assignment of entities to free variables. The recursive account of satisfaction appropriate to each particular sentence constitutes a proof of a theorem for that sentence in the form of a T-sentence. The theory would give a T sentence for each elementary sentence, and a rule corresponding to each sentential connective.

T-sentences provide the definition of the conditions under which sentences of English are asserted (held true). In the extension of the predicate 'is T', happen to be all

⁵⁴ *Inquiries* 47.

true (assertible) sentences of English. "It was in the nature of a discovery that such a predicate would apply exactly to the true sentences."⁵⁵ Davidson claims that the connection between the Tarskian definition of truth in the form of Convention T and the concept of meaning is obvious: to be a competent speaker of, say, English, (i.e., to be able to understand arbitrarily many sentences in English) amounts to knowing the necessary and sufficient conditions under which sentences of English are (truthfully) asserted.⁵⁶

2.8 Metalanguage Versus Object Language

Davidson discusses two cases of constructing a theory of meaning for a given NL. The NL considered as the object language (for which truth is being characterized) could be the language used and understood by the person constructing the theory. In case of an English speaking "theorizer" characterising the truth-conditions for sentences in English, this would mean that English would be used both as the metalanguage of the theory as well as the object language about which theoretical claims are being made. Davidson remarks of the convenience of having a metalanguage with a sentence guaranteed equivalent to each sentence in the object language. Apart from the obvious convenience of being able to make assertions about a NL by using the same NL (an empirical test for such an application of the theory would be a trivial matter), there are some problems with this feature that seem to threaten the completion of the theory of

⁵⁵ *Inquiries* 24-25.

⁵⁶ *Inquiries* 24.

meaning for that NL. I shall return to these and other problems later in this thesis.

The other case mentioned by Davidson⁵⁷ is when a speaker of one NL constructs a theory of meaning for the speaker of another NL. In this case, there can be no guarantee for complete equivalences between sentences of the meta-, and of the object language, so that the empirical test for the theory would no longer be trivial. The way to go about this "imperfect fit" between the two languages would be for the theorizer to find out what sentences the speaker of the object language holds true, and then to attempt a characterisation for the extension of the truth predicate in that language, and finally, to map the sentences held true in the object language to sentences held true in the theorizer's language, i.e., in the metalanguage. As Davidson notes, "supposing no perfect fit is found, the residue of sentences held true translated by sentences held false (and vice versa) is the margin for error (foreign and domestic)."⁵⁸ Davidson's holding the problem of interpretation for domestic -- as well as foreign --, is a very important characteristic of his views on meaning, since it (again) forecasts Davidson's doubts with respect to the plausibility of a uniquely definable concept of language. Due to its crucial relevance, this issue will be discussed in more detail in the following chapters.

⁵⁷ *Inquiries* 27.

⁵⁸ *Inquiries* 27.

My introduction of Davidson's optimistic and programmatic view on the possibilities for a formal characterization of a truth predicate for a natural language has dealt with the various issues involving the fulfilment of his first requirement vis-à-vis an adequate theory of meaning, namely to provide an interpretation of any sentence a speaker of NL may utter. Naturally, there are numerous problems with some of the more "recalcitrant" features of sentences in NL, not yet touched upon. Since these problems were seen by many as standing in the way of completing the theory of meaning, they deserve a more detailed, separate discussion, which will be provided in chapter four of this thesis. But first, in the following chapter I shall turn to the discussion concerning the second requirement, namely, the problem of how to tell whether the theory works, i.e., the empirical application of the theory of meaning.

CHAPTER THREE

THE THEORY AT WORK

If one were to apply the theory of meaning to a particular NL, one would have to find a way to tell whether or not the theory really applies to that language, whether it is capable of providing the meanings of arbitrarily many sentences of that NL, in other words, whether or not the theory works. What would count as evidence for the correctness of the theory of meaning? In this chapter the problem of empirical testability for Davidson's theory of meaning will be discussed. As I previously mentioned, there are strong connections between Davidson's and Quine's views with respect to meaning and understanding. The Quinean concept of *the inscrutability of reference* bore relevance to Davidson's rejection of the classical notion of meaning, and to his adoption of a truth-theoretical semantics for NL. Quine's theory of *radical translation*, too, has its affects on Davidson's theory, in particular on his method of verification. It seems therefore necessary to begin the discussion with Quine's theory of radical translation, as it appears in the famous Chapter Two of Word and Object.⁵⁹ The second segment of my discussion compares radical translation with Davidson's *radical interpretation*, and the third one introduces the Principle of Charity, the central notion within Davidson's philosophy of language.

⁵⁹ W. V. O. Quine (1960) Word and Object, Cambridge, Mass.: M.I.T. pp.26-79.

3.1 Radical Translation

Quine's program is to find out how much of NL can be made sense of in terms of its stimulus conditions. These stimulus conditions are conditions of the world that prompt us to respond verbally. Objects and events of the world "impact at our nerve endings", thus, sense data is produced. Some stimuli, e.g., physical objects, are socially, *intersubjectively* observable. Some other stimuli, such as a person's feeling pain, are internal to that person, i.e., *intrasubjective*, not directly observable for others, although even such a stimulus can be accompanied with observable behavioral cues, like wincing. Language is acquired socially, that is, we learn from intersubjectively available cues, when and how to use it correctly.

In so far as linguistic meaning is concerned, the *only* justification for it is found in the dispositions of speakers of a linguistic community to respond overtly to socially observable stimulations. All one can observe is, what is being uttered by speakers of a given NL, and under what observable conditions. Our observations concerning the conditions under which utterances are made are idiosyncratic, dependant on our personal history of situations, in which our knowledge of that language has been acquired, and on our particular vantage point (both literally and figuratively speaking), from which we presently observe the prevailing stimulus conditions.⁶⁰ Quine's thought-experiment is meant to illuminate the fact that linguistic meaning is underdetermined by available evidence, consequently, "manuals for translating one language into another can be set up

⁶⁰ Ibid., 8-9.

in divergent ways, all compatible with the totality of speech dispositions, yet incompatible with one another."⁶¹

A case of radical translation arises according to Quine, when for example an English speaking observer (the field linguist) wishes to create a *translation manual* for some totally unknown language. The manual is expected to contain for each native sentence its English equivalent. In order to solicit evidence for her translation manual, all the field linguist can do is to test the native speakers for assent to and dissent from sentences in response to various stimulation conditions.⁶²

The utterances most likely to be translated first, are the ones referring to objects or events, conspicuous to both the linguist and the native speaker. To use Quine's famous example, the field linguist may repeatedly observe the native speaker uttering the one-word sentence 'Gavagai' whenever a rabbit runs by.⁶³ The field linguist would hypothesize that 'Gavagai' may be translated as 'There's a rabbit' and set up some test situations to verify her hypothesis. Some of these situations would include a rabbit running by, while the control situations would exclude such event. If the field linguist asked the native speaker in each of these situations 'Gavagai?', responses assenting to the sentence when a rabbit is present and dissenting from it in the absence of the rabbit

⁶¹ Ibid., 27.

⁶² For the behavioristically inclined Quine, the above description exhausts the totality of linguistic behavior.

⁶³ Ibid., 29.

would help to deem the translation tentatively as acceptable.⁶⁴

Quine defines the *affirmative stimulus meaning* of a sentence for a given speaker S, as "the class of all the stimulations ... that would prompt his assent."⁶⁵ According to this definition, 'There's a rabbit' in English and 'Gavagai' in the native language can be equated insofar as they both have the same stimulus meaning (relative to time *t*, the time of the observations, and relative to speaker S). If there were sentences to which speakers *always* gave assent or dissent on the basis of their sensory stimulations, then stimulus meaning could serve as *the* meaning of these sentences. However, for Quine, no sentence is connected merely to sense experience, but rather to a host of other sentences as well.

Quine, supporting *semantic holism*, conceives of language as a web, the nodes of which represent sentences.⁶⁶ The identity of each node depends on its relations with both other nodes and with experience. The nodes at the outer edge of the web (Quine's *observation sentences*) are directly connected to sense experience, while sentences further to the middle of the web (e.g., 'my father's sister is my aunt'), having less direct contact with experience, obtain their identity through their connections with other sentences. Even though stimulus meaning pulls considerable weight in the case of observation

⁶⁴ It is important to note, that it is not the rabbit itself, but the stimulation ("the pattern of chromatic irradiation of the eye") that prompts the native speaker to be "affirmatively disposed" towards 'Gavagai'. So, even at the appearance of a cleverly disguised mechanical toy that looks like a rabbit, the native would be expected to assent to 'Gavagai'.

⁶⁵ Ibid., 32.

⁶⁶ Ibid., 42.

sentences, it clearly cannot serve as the meaning of all sentences.

In order to solve this problem, Quine proposes that the field linguist engaged in radical translation adopt "analytical hypotheses."⁶⁷ The field linguist observing that certain segments of observation sentences, for which she has already found a tentative translation, constitute parts of other sentences not yet translated, can form hypotheses about how the already translated segments may contribute to the meaning of the sentences in which they occur. These analytical hypotheses however, are just unverifiable hypotheses; that is, they go beyond all behavioral evidence there is available for translation.

Quine holds that translation and meaning exceed all possible evidence (i.e., are indeterminate) and therefore no fact of the matter concerning the uniquely correct way to translate (or to interpret) some segment of an NL can be found in the totality of evidence.⁶⁸ This means that, precisely because the analytical hypotheses of the field linguist are underdetermined by evidence, it would be possible for her to form several, non-equivalent sets of analytical hypotheses for the native language under scrutiny. In Quine's own words: "... rival systems of analytical hypotheses can conform to all speech

⁶⁷ Ibid., 68.

⁶⁸ Even though Quine's thought-experiment focusses on *translation*, it is in fact meant to sharpen our understanding on how *meaning* is interpreted. Since for Quine, behavioral dispositions constitute the meanings of the sentences of a NL, not only translation, but meaning itself is indeterminate. Quine rejects the "uncritical, mentalistic theory of ideas," according to which "each sentence and its admissible translations express an identical idea." See Ibid., 74.

dispositions within each of the languages concerned and yet dictate, in countless cases, utterly disparate translations;...⁶⁹ This would lead to the result that for some native sentence *n*, one set of analytical hypotheses could give as the translation for *n* the English sentence *e*, while another set could yield *e**, where *e* and *e** are different, non-synonymous sentences.⁷⁰ In the case of the native sentence 'Gavagai' there would be no fact of the matter about whether the English translation 'There's a rabbit', or its rival translations, such as for example 'There are some undetached rabbit parts,' or 'A rabbit phase just passed by' would count as the uniquely correct translation.⁷¹ Linguistic meaning is characterised by a systematic indeterminacy: "Beneath the uniformity that unites us in communication there is a chaotic personal diversity of connections, and, for each of us, the connections continue to evolve. No two of us learn our language alike, nor, in a sense, does any finish learning it while he lives."⁷² To summarize Quine's argument:

1. all genuine semantic facts are constituted by facts about behaviour.
 2. facts about behaviour do not uniquely determine facts about meaning or translation.
- therefore, meaning and translation are indeterminate.

Quine's holistic view on language, together with his notion of indeterminacy has been adopted with some modifications by Davidson in his theory of meaning. In the following

⁶⁹ Ibid., 73.

⁷⁰ Ibid., 51. Quine makes the distinction between *stimulus-synonymy* of expressions, meaning that the situations prompting the utterance of expressions are the same, and *intrasubjective stimulus-synonymy*, by which he refers to speakers' idiosyncratic (and inscrutable) beliefs concerning the situation prompting an utterance. The latter may not coincide with the former, which constitutes the main source of indeterminacy.

⁷¹ Ibid., 52.

⁷² Ibid., 13.

section I shall elaborate on Davidson's ways to make use of the concept of radical translation.

3.2 Radical Interpretation

Quine's question was: to what extent can a theory that translates all the sentences of a native NL rely on behavioral evidence; his radical translation theory answered this question by testing speakers' assent to and dissent from sentences under certain stimulus conditions. Davidson's question is: what is it a speaker of NL might know in order to successfully interpret any of the infinity of sentences in NL, or, in other words, how can a theory that is "specifically semantical in nature" and based on evidence "described in non-semantical terms", be construed.⁷³

In chapter two of this thesis I have given a detailed account concerning Davidson's disenchantment with the idea that corresponding to each meaningful expression there is an entity: the meaning of that expression, so that all the interpreter needs to know is this entity in order to interpret the expressions. Whether or not his truth theory can indeed give the meaning of all sentences of NL is a matter of a separate evaluation. Meanwhile, I am concerned with what knowledge Davidson considers as a viable basis of interpretation.

⁷³ *Inquiries* 142. A specifically semantical theory would be capable of telling what all sentences of an NL mean, and the requirement of evidence described in non-semantical terms would prevent circularity, for this way the evidence were available independently of the theory to be confirmed by it. In a way, Quine and Davidson differ in their conceptions of what the evidence for their theories would look like. Quine insists on exclusively behavioural evidence, while for Davidson it is important that it be in non-semantical terms.

Intention based semantics (as for example the Gricean project, the great rival theory to formal semantics) considers the meaning of sentences as dependant on the *intentions* with which they are used.⁷⁴ Davidson rejects this approach for two reasons: intention based semantics cannot deal with the structural, recursive feature of NL, considered by Davidson essential to explaining how new sentences are understood. What is even more important for Davidson is that he holds the speaker's intentions, beliefs and the meaning of the speaker's sentences to be interdependent parts of a single project, neither of which, taken in isolation, can serve as the evidential base of interpretation.⁷⁵ Attributions of attitudes and beliefs will essentially rest on the same evidence as interpretation:

beliefs and meanings conspire to account for utterances. A speaker who holds a sentence to be true on an occasion does so in part because of what he means, or would mean, by an utterance of that sentence, and in part because of what he believes. If all we have to go on is the fact of honest utterance, we cannot infer the belief without knowing the meaning, and have no chance of inferring the meaning without the belief.⁷⁶

The first and most obvious suggestion answering the question posed by Davidson would be that speaking *the same* language would enable speakers to interpret one another, since the same utterances would be interpreted in the same way. Davidson rejects this

⁷⁴ See for example, Grice, H.P. (1957) "Meaning", reprinted in A.P. Martinich, ed. The Philosophy of Language, second edition, pp.72-78.

⁷⁵ *Inquiries* 127.

⁷⁶ *Inquiries* 142. The interdependency of meaning and beliefs has been stressed by Davidson consistently, throughout his work concerning interpretation. I shall show in chapter five of this thesis, how this view had to lead to Davidson's rejection of linguistic conventions and later of 'language' itself.

suggestion claiming that since the same utterances can mean radically different things in different languages, it is a *mere assumption that both speakers are speaking the same language* when encountering the same expressions.⁷⁷ For Davidson, *all* understanding of the speech of one another involves *radical interpretation*, since the assumption that two speakers of a given NL speak the same language is just an unproven assumption. Accordingly, "the evidence must be of a sort that would be available to someone who does not already know how to interpret utterances the theory is designed to cover: it must be evidence that can be stated without essential use of such linguistic concepts as meaning, interpretation, synonymy, and the like."⁷⁸ This evidence for Davidson is just the attitude of speakers of a given NL *to hold or to accept various sentences of NL to be true* at certain times/places, uttered by certain speakers under specified circumstances.⁷⁹

This approach solves the interdependence of belief and meaning, insofar as it holds the particular belief 'holding true' *constant*. The interpreter assigns truth conditions to utterances to be interpreted according to what she holds to be true. Radical interpretation can only proceed based on the assumptions of the interpreter that

- a) the speaker is saying what she holds true
- b) speaker and interpreter are in agreement about sentences they hold true

⁷⁷ *Inquiries* 125.

⁷⁸ *Inquiries* 128.

⁷⁹ *Inquiries* 135.

Since meaning is a function of specified conditions determined by the objective features of the world, meaning will *vary* according to the changes in these features. Moreover, individual speakers of a given NL may at times be mistaken about what they hold true about the features of the world, although massive error is implausible. The crucial assumption for interpretation is that of agreement between speaker and interpreter about sentences held true by them.⁸⁰ Davidson's own concrete example clarifies this point.⁸¹ The interpreter in the example is English speaking, the native speaker named Kurt belongs to the German speech community; speaker and interpreter are sharing their immediate, observable environment, the time is Saturday at noon, and it is raining. Kurt utters the following sentence:

(1) Es regnet.

The interpreter, *assuming* that Kurt uttered a sentence he holds true, and *knowing the conditions under which sentence (1) has been uttered*, construes the following T-sentence:

(T) 'Es regnet' is true-in-German when spoken by speaker x at time t iff it is raining near x at t .

The evidence for the T-sentence above, available to the interpreter is

(E) Kurt belongs to the German speech community and Kurt holds true 'Es regnet' on Saturday at noon and it is raining near Kurt on Saturday at noon.

⁸⁰ This view concerning agreement between speakers about sentences held true, underwent various modifications in Davidson's theory. In more recent papers he emphasized that the agreement has to be the "right sort" of agreement, guided by the shared ways of speakers' rational reasoning. For more details on this central issue of Davidson's theory, see chapter five of this thesis.

⁸¹ *Inquiries* 135.

The resulting partial theory based on empirical evidence in the form of truth-conditions given by the T-sentence would be, according to Davidson

(GE) $(x)(t)$ (if x belongs to the German speech community then x holds true 'Es regnet' at t iff it is raining near x at t)).

The totality of sentences like (GE) defining the truth-conditions for all sentences in German would constitute a theory of meaning for German.⁸²

The Quinean notion of indeterminacy of meaning is adopted by Davidson insofar as he suggests that agreement of beliefs concerning sentences held true must be assumed but cannot uniquely be reconstructed from speech behavior: consequently, there may be different ways of assigning truth-conditions to sentences, that is, there may be different, non-equivalent sentences in the metalanguage of the theory, which are true iff various sentences of the object language are true, such that the different truth theories they imply will equally well conform to all the formal and empirical requirements of the theory. The aim of the interpretation is to get the best fit, "as far as we can tell, as often as possible."⁸³

In fact, however, there are at least two ways in which Davidson's form of indeterminacy may be less disconcerting than that of Quine's: while Quine chose to

⁸² Here, Davidson clearly appeals to the linguistic convention of a particular speech-community for deciding whether the theory concerning the utterance of 'Es regnet' is correct. Later, Davidson realized that this appeal to conventions could not be the relevant principle of how interpretation succeeds, and he rejected the attribution of importance to the role of convention in communication. For further details, see chapter five of this thesis.

⁸³ *Inquiries* 136.

construe stimulus conditions in terms of sensory states, Davidson rejects Quine's notion of stimulus meaning. Instead, Davidson views things and events in the world at large as being the relevant conditions under which sentences are asserted. If the conditions under which sentences are held true are construed in terms of the external, *intersubjectively* observable world, rather than in terms of *intrasubjective* sensory states, the possibility of indeterminacy becomes less alarming. At the same time, Davidson's theory requires (by virtue of being based on the Tarskian truth theory) that first-order quantificational logic be applicable to the object language, while Quine's theory did not impose any theoretical constraint on how the sentences are to be analysed. Davidson's restriction would probably diminish the problem of indeterminacy as well.

Quine's holistic view considered the meaning of sentences in terms of their being situated in a web of other sentences constituting the whole of language. In his early paper, "Truth and Meaning", written in 1967, Davidson committed himself to holism only in the Quinean sense.⁸⁴ In his more recent paper from 1973, "Radical Interpretation", (on which my present discussion is based) the meaning of speakers' utterances is interdependent with their beliefs concerning the meaning of sentences uttered. Quine's doctrine of holism has been expanded here by Davidson in terms of connecting the so-called literal meaning of sentences with speakers' beliefs, i.e., what the speakers mean by the sentences on the particular occasion of the utterance. Clearly, the Quinean, purely behavioristic evidence will not do for Davidson; for Davidson, the crucial evidence will

⁸⁴ See his statement to that effect quoted verbatim on page 31 of this thesis.

be essentially non-behavioral, encompassing the mental, the domain of beliefs. Davidson differs from Quine insofar as he, but not Quine allows his theory to rely on a merely theoretical posit, namely, that of the particular belief of 'holding true', without which Davidson's theory does not seem able to do its task.

I would like to point out one further difference between the Quinean and Davidsonian projects: the method of radical translation suggested by Quine would be inadequate as a solution to Davidson's problem, since the former is not intended to *give the meaning* of sentences of the native NL, rather, it is merely intended to *map sentences* of the native NL (the object language) to sentences of an NL already mastered by the field linguist (the subject language), while describing the procedure in the metalanguage of the theory (which may or may not coincide with the subject language used by the field linguist). Quine's translation manual could show, using English as the metalanguage, that the French sentence 'La neige est blanche' translates the German sentence 'Schnee ist weiss' without thereby telling what either sentence means. In Davidson's case, if one understands the metalanguage of the theory, one understands the object language for which it is a theory.⁸⁵

3.3 The Principle of Charity

This principle has already been alluded to in the previous segment of this chapter, when the radical interpreter's assumptions concerning the speaker's beliefs vis-à-vis

⁸⁵ *Inquiries* 129.

sentences held true were discussed. The Principle of Charity is one of the central notions of Davidson's theory of meaning, both in its early as well as in its recent formulations. Having been stated somewhat vaguely in the early theory, it underwent several modifications and refinements provided by Davidson himself, and played a crucially important role in Davidson's recent, controversial views on language. Given its considerable significance, I shall discuss the Principle of Charity here only as it first appeared in Davidson's early theory, while its developments will be elaborated on in chapter five of this thesis. As often in Davidson's theory, in this case too, his mentor Quine's footsteps are clearly retraceable.

Quine's use of the Principle of Charity is limited. He provides a good example of this principle when discussing the radical translation of truth-functional logical constants.⁸⁶ He claims that behavioural evidence (in the form of the native speaker's assent or dissent) would suffice for us to identify and translate the truth-functional operators of a previously unknown native NL. For example, in the case of *negation*, a short sentence originally assented to will turn out to be dissented from when negated. Another operator, *conjunction* produces conjuncts to which the speaker would assent always and only if she were to assent to each component.

Should the speaker deviate from the expected response, says Quine, it must be attributed to her temporary confusion, rather than to extreme differences of her native

⁸⁶ Quine, (1960) Word and Object, p.57.

logic compared to our own. If a native speaker were observed to assent regularly to sentences translatable as ' p and *not* p ', instead of assuming that she is flouting the law of non - contradiction, we had to conclude that she is using her language differently, that is, our translation of it was mistaken. Quine extends this view to the domestic case as well, where, if a speaker espouses a logic whose laws are ostensibly contrary to our own, we will conclude that the speaker is using familiar words by giving them new meaning, rather than assuming that the speaker does not share our laws of logic. "The common sense behind the maxim is that one's interlocutor's silliness, beyond a certain point, is less likely than bad translation - or, in the domestic case, linguistic divergence."⁸⁷

The maxim referred to by Quine was named by Neil Wilson the Principle of Charity.⁸⁸ Davidson applies the principle, not only to assert that one's interlocutors are not likely to believe in logical contradictions, but by claiming that interpretation is possible only if the interpreter takes other people to believe what she thinks it is right to believe. Only in that case can one use information, concerning the conditions under which sentences are held true, as evidence for the theory of meaning. In other words, in order to use facts about conditions under which speakers hold sentences true as evidence for a truth theory, one must assume that they believe what the interpreter thinks they should believe. In the early application of the Principle of Charity, Davidson emphasizes agreement by the interpreter's standards. In a simplistic way, it can be taken as a general

⁸⁷ Ibid., 59.

⁸⁸ Wilson, N.L. (1959) "Substance Without Substrata", *Review of Metaphysics* 12 pp. 521-539.

guideline to rational behaviour: believe only things you take to be true, act according to your best judgement on the basis of all relevant, available evidence.

With this segment Part One of this thesis has been completed. Davidson's early theory of meaning has been explicated, its origins and motivating forces discussed, its workings analysed. So far, no attempt on my part has been made to provide any critique to the theory's more controversial elements. My aim was merely to introduce the reader to my understanding of Davidson's theory, since any critique on my part added henceforth will be based on my particular interpretation. I believe that the reader will be in a better position to judge whether or not my comments to Davidson's theory are justified, after having been familiarized with my reading of it.

Davidson proposed a new way of constructing such a theory, since he had found the already existing alternatives inadequate. Four conditions have been identified as leading to both the possibility and the necessity of formulating an adequate theory of meaning:

- a) The meaning of any sentence in NL is a function of a finite stock of features contained in the sentence.
- b) NL is learnable by virtue of its compositional nature.
- c) Mature speakers are competent in understanding infinitely many sentences of NL based on their knowledge of a finite vocabulary and a finite set of syntactical rules.
- d) Even though no speaker of NL has ever been exposed to all possible sentences of NL at any given time, all competent speakers of NL are capable of understanding arbitrarily many novel sentences.

Davidson claims that a theory of truth, modified to apply to a given NL, can be used to describe in a finite form the kind of knowledge that would enable a speaker to interpret any sentence of that NL, and as such it would be adequate as a theory of meaning. This proposal raises three questions (suggested by Davidson himself) that must be answered when defending the choice of such a theory:

1. Is it reasonable to think that a theory of truth of the sort described can be given for a natural language?
2. Would it be possible to tell that such a theory was correct on the basis of evidence plausibly available to an interpreter with no prior knowledge of the language to be interpreted?
3. If the theory were known to be true, would it be possible to interpret utterances of speakers of the language?⁸⁹

Davidson himself had answered these three questions, that essentially deal with the *possibility of completing* the theory, with its *testability* against evidence, and with its *interpretability* as a theory of meaning, respectively. I too, shall follow the above outline in my discussion concerning the implications of the theory in Part Two of this thesis. However, since my aim is not to provide a defence of Davidson's theory against all possible objections, I shall not provide an exhaustive discussion of the pro's and con's of these three issues.

My aim in this thesis concerns one single objection against which I wish to defend Davidson's theory and that objection can be described as follows. Precisely while discussing the pro's and con's of the above three issues, some philosophers misinterpreted Davidson's proposal of a theory of truth to be employed for a theory of meaning, and

⁸⁹ *Inquiries* 131.

concluded falsely that Davidson, seeing the impossibility of his theoretical endeavour, had finally abandoned his early views, declaring defeat by rejecting the very intelligibility of the theoretical concept 'language'.

My main focus in this thesis is to defend the point of view, that, contrary to some philosophers' opinions (of which Hacking's critique mentioned in the introduction of my thesis is a representative example), Davidson did not undergo any radical conversion concerning the notion of language as a theoretically useful concept. In other words, I attempt to show that there is a consistency in his early and recent views concerning this issue and all that has been changed is his way of spelling out more explicitly his consistently held view, present in his earliest papers. Chapter four is intended by me as a mere demonstration of Davidson's optimism concerning the possible completion of the theory. He has spent a significant period of time tirelessly working on attempts to formalize NL. Even though he clearly saw the difficulties such attempts would mean, he remained confident that the significant progress witnessed in the study of formal logic would eventually resolve these problems.

The most important development in Davidson's theory of meaning occurred as he made his views concerning the empirical constraints of his theory more explicit. These are his most relevant considerations that eventually lead to his recent, controversial conclusions concerning language. In a sense, he can be said to have shifted his focus from the logical form of NL to the act of interpretation, but that shift is not to be

interpreted as an abandonment of his early views. As we saw in the present chapter, from very early on, Davidson consistently held certain views about how interpretation of utterances is possible. To show that he remained consistent with regard to these views, is the main task of my thesis, to be discussed in chapter five. That discussion will hopefully shed light on the issue too, whether Davidson's theory is interpretative. With regard to this issue, I believe that it is important to evaluate Davidson's theory in its entirety, including not only the role of T-sentences, but also the role of the Principle of Charity. Most of those, who object to the claim that the truth theory possesses an interpretative force, tend to focus mistakenly only on the formal constraints of the theory, while conveniently neglecting its equally, (if not more) relevant empirical claims.

PART TWO

THE IMPLICATIONS OF THE THEORY

CHAPTER FOUR

THE POSSIBILITY OF COMPLETING THE THEORY

My discussion, following Davidson's own outline for a defence of his theory, begins with addressing problems concerning the successful completion of the theory. These problems have been taken as contributing heavily to the "inevitable abandonment" of the theory' resulting from Davidson's allegedly growing pessimism regarding the possible completion of his early theory. Having found the traditional theories of meaning inadequate, Davidson opted for a Tarskian truth theory that, with some necessary modifications, was to be employed in constructing his theory of meaning for NL. Tarski held the prospects for a formal semantical theory of NL very poor. He pointed out three major problems that would render the completion of a truth theory for NL impossible: first, said Tarski, it is essential for the metalanguage of the theory to have richer ontological resources than the object language about which the theory is construed. Consequently, it would be impossible to use NL as the metalanguage of the theory when the same NL is also the object language. Tarski's other two objections to a formal theory of NL were that the universal character of NL leads to contradictions, and that the ambiguities inherent in NL do not permit the direct application of formal methods. Obviously, since Davidson has chosen Tarski's theory as the basis of his theory of meaning, he must have had some suggestions with regard to overcoming the difficulties surrounding such a project. In the first segment of this chapter I shall summarize Davidson's responses to the problems mentioned by Tarski.

The remainder of this chapter deals with the problems implied by Davidson's commitment to a purely extensional semantics (acknowledged by Davidson himself). Two types of problems arise because of Davidson's commitment to a purely extensional semantics: on one hand, Davidson has to show that his commitment to extensionality will not entail some untoward consequences rendering the theory useless, on the other hand, he has to prove that an extensional truth theory can successfully account for the interpretability of *any* type of sentence a speaker of NL may utter, if the theory is to be completed.

In particular, I shall discuss an often cited objection against Davidson's extensional truth theory, namely, that the truth-value of T-sentences would be unaffected by pairing any two sentences as long as these are both true, so that, instead of having interpretative force, T-sentences may result in grotesque pairs such as for example, '*snow is white*' is true iff *grass is green*. I shall argue that this objection is not a genuine objection against his theory. In the last segment of the chapter, a paradigm case of the other type of problem is introduced (the particular problem of *indirect discourse*, or *oratio obliqua*), as it appears in the paper "On Saying That"⁹⁰. Davidson offers a solution to the problem of indirect discourse that is to be the key to a correct analysis of psychological sentences in general. He explains how the characteristically intensional idioms of NL can be handled by an extensional theory of truth. As I have mentioned in the conclusion of the previous chapter, my aim is not to include all the problems a completed theory would

⁹⁰ *Inquiries* 93-108.

have to resolve, since that would go beyond the scope of this thesis. My task here is merely to demonstrate that, far from being pessimistic about the outcome of his proposal, Davidson has been most ingenious in offering some creative solutions to the problems encountered.

4.1 Responses to Tarski's Objections

The seeming problem with constructing a truth theory for an NL (say English) as the object language, when the same NL is to be used as the metalanguage of the theory, "where a significant fragment of a language (plus one or two semantical predicates) is used to state its own theory of truth",⁹¹ is the problem of infinite regress. Any attempt to construct theorems *in* English in which to state the theory of truth *for* English will expand the list of sentences of English to be interpreted. For example, the T-sentence stated in English

- (1) 'snow is white' is true iff snow is white

would also have to be interpreted in terms of giving *its* truth-conditions:

- (2) The sentence "'snow is white' is true iff snow is white' is true in English iff 'snow is white' is true iff snow is white

and so on, *ad infinitum*, so that the truth-theory for NL would have to remain forever unfinished. Davidson does not seem to worry about the regress problem. According to him, the suspected incompleteness of the theory does not in itself render the theory useless as a theory of interpretation. Davidson's suggestion is to simply require that the

⁹¹ *Inquiries* 132.

T-sentences of the theory *be true*, without any further analysis into what makes them true.⁹²

The explanation for this stipulative move on Davidson's part is given by him as follows. Tarski's aim was to provide a full characterisation of the truth predicate in terms of the axioms and of the rules of inference of the theory (for FL, this aim could be achieved through reliance on a separate metalanguage that has richer resources than the FL about which the theory is to be construed). Davidson's response to the pessimistic claim that no explicit definition of truth for NL can be given using the Tarskian method is, that, while this claim is true, it does not endanger his project, since his and Tarski's aims are not the same:

While Tarski intended to analyse the concept of truth by appealing (in Convention T) to the concept of meaning (in the guise of sameness of meaning, or translation), I have the reverse in mind. I considered truth to be the central primitive concept, and hoped, by detailing truth's structure, to get at meaning.⁹³

In other words, Davidson's theory of meaning *uses* truth as an unanalysed primitive term, rather than aiming at an explicit definition of the concept of truth itself. The necessary "increase in ideology" for NL used as the metalanguage of the theory vis-à-vis the same NL as the object language is restricted to the use of the truth-predicate (and of the satisfaction-predicate), otherwise the ontology of NL as a metalanguage can be the same

⁹² *Inquiries* 73.

⁹³ *Inquiries* XIV.

as that of the same NL as object language.⁹⁴ Again, we must remember that the theory of meaning for NL is to provide a description of the knowledge that would enable one to interpret sentences of NL. According to Davidson, the competent speaker's ability to grasp the concept of truth in NL intuitively, even without being able to define it explicitly, would enable her to successfully utilize the T-sentences for interpretation.⁹⁵

In his early paper, "Truth and Meaning",⁹⁶ Davidson comments on the two other objections raised by Tarski concerning the insurmountable difficulties of constructing a formal semantical theory of NL, namely, that the universal character of NL leads to contradictions (to antinomies), and that the ambiguities inherent in NL do not permit the direct application of formal methods. The gist of Davidson's answer to the problem that universality leads to paradox is that, even though he cannot seem to come up with a solution to the problem, none the less, this problem refers only to a small fragment of the relevant NL, and so, he will not consider it a serious obstacle to his theory.⁹⁷ In actual practice, semanticists only work on a single fragment of NL at a time; so it can be assumed that this kind of piece meal work would eventually be completed.

⁹⁴ *Inquiries* 72. See also Quine's comment concerning this issue: "... the demand for a stronger metalanguage arises, in general, only when we undertake to transform the recursive definition of truth into a direct definition. This we need not insist on doing." W.V.O. Quine (1980) "Reply to Davidson", in D. Davidson and J. Hintikka, eds. Words and Objections: Essays on the Work of W.V. Quine, pp. 333-335. Dordrecht: D. Reidel Publishing Company. p.334.

⁹⁵ *Inquiries* 29.

⁹⁶ *Inquiries* 28-29.

⁹⁷ *Inquiries* 28.

Tarski's second objection, namely, that unless NL undergoes a normalization in which all its ambiguities are eliminated, no formal semantics can be constructed for it, is given here a somewhat more serious reply by Davidson: some parts of NL, for example of English, are amenable to being formalized, i.e., to be given a canonical notation, and for these parts his theory will do the work.⁹⁸ As far as the rest of English is concerned, Davidson has repeatedly attempted the "semantic taming" of some of the most problematic idioms of English.⁹⁹ A paradigm case of such attempts will be discussed in the last segment of this chapter. Before that, however, I shall be concerned with a more general implication of Davidson's commitment to an extensional semantics for NL.

4.2 Commitment to Extensionality

As I have attempted to show in Chapter Two, Davidson had rejected the Fregean intensionalist concept of meaning in his theory, for a variety of reasons. Apart from deeming the use of the concept 'meaning' superfluous, two further reasons had been identified, explaining Davidson's preference for an extensional, truth-theoretical semantics instead of an intensional, Fregean one. Following Quine's nominalism, Davidson rejected the positing of meanings as these, being obscure entities, their proper identification could not be done on purely empirical grounds (see p.34). Davidson also adopted Quine's theses on indeterminacy and on semantic holism, which played a significant role in his

⁹⁸ *Inquiries* 29. Here and elsewhere Davidson appeals to Chomsky's notion of deep structure; NL's surface structure does not always reveal the possibilities of formalization for NL; these however become evident once one considers the underlying deep structure of NL.

⁹⁹ Even though Davidson emphasized that his aim was not to reform NL, but to describe and explain its workings. See, for example *Inquiries* 29.

rejection of meanings:

Indeterminacy shows ... that if there is one way of getting it right there are other ways that differ substantially in that non-synonymous sentences are used ... And this is enough to justify our feeling that there is something bogus about the sharpness questions of meaning must in principle have if meanings are entities¹⁰⁰

With regard to Davidson's commitment to extensionality, the question has often been raised, whether mere knowledge of truth-conditions can indeed elucidate the meaning of a sentence, that is, whether Davidson is justified in replacing '*s* means that *p*' with '*s* is true iff *p*'. According to Tarski's theory, in order to apply Convention T to elucidate the concept of truth, one must know what *s* means, since only if one already has an interpretation for *s*, can one tell whether or not *p* means the same as *s* or a translation of *s*. Davidson's aim is just the opposite: he presumes an understanding of the concept of truth that will lead to the ability to interpret sentences of NL. The exchange of the intensionality of meaning for the extensionality of truth will have the seemingly serious consequence that two sentences could be true under the same conditions and yet none the less not be synonymous. For example while the sentence

(1) 'Snow is white' is true iff grass is green

is true,

the sentence

(2) 'Snow is white' means that grass is green

is obviously false, so that the mere truth of theorem (1) used to give the meaning of 'snow is white' cannot do a satisfactory service for a theory of meaning. Here the truth-

¹⁰⁰ *Inquiries* 101.

value of theorem (1) is taken as depending solely on the extension (truth-value) of its component sentences, and, according to standard logic, a material biconditional is true as long as both of its component sentences are true (or both are false, for that matter).

However, a special feature of the truth-theory is that the truth-conditions of the sentences are to be accounted for in terms of compositionality, that is, on the basis of the component expressions and connectives occurring in the sentence. Each T-sentence of English is related to the totality of possible T-sentences of English, in so far as they all share various component parts. Davidson considers it a mistake

to think that all we can learn from a theory of truth about the meaning of a particular sentence is contained in the biconditional demanded by Convention T. What we can learn is brought out rather in the *proof* of such a biconditional, for the proof must demonstrate, step by step, how the truth value of the sentence depends upon a recursively given structure.¹⁰¹

Thus, the theory of meaning should not be interpreted as mere invariable pairing of truths with truths and falsehoods with falsehoods: "we can have confidence in a characterization of the truth predicate only if it pairs that sentence [the sentence to be interpreted] with one we have good reason to believe equivalent."¹⁰² Just what "having good reason to believe" would mean becomes clearer if one considers that for Davidson, interpretation is always radical and thus, it necessitates the employment of the Principle of Charity. To cite an example of radical interpretation, if one were to interpret a novel sentence uttered by an alien speaker, it would not be enough to have knowledge of a theorem in the form

¹⁰¹ *Inquiries* 61.

¹⁰² *Inquiries* 26.

of s is true iff p . Since the object language is alien, the interpreter would have no way to tell, whether the pairing of sentences on the right-, and on the left-hand side of the biconditional is not resulting in the kind of grotesque theorem in (1). It is therefore crucial for the interpreter to assume that the alien speaker is to be interpreted *on the same rational basis* as the one possessed by the interpreter herself, under the assumption that the sentence s uttered was *held true* by the speaker. This constraint on the theory concerning Charity applies both to foreign and domestic radical interpretation. In other words, the interpreter (domestic or alien) must know that the T-sentences are part of a theory that conforms to the Principle of Charity. To summarize the argument so far analysed, interpretation must be provided by T-sentences taken as parts of a whole theory that conforms to the Principle of Charity and the requisite formal constraints.

The final argument supporting the view that insofar as the above objection against extensionality is concerned, Davidson's design is not in ruins, is that Davidson has subscribed to Quine's indeterminacy thesis from early on. Consequently, he allows that the above constraints cannot determine a uniquely correct theory.¹⁰³ In a more recent essay "The Material Mind" he echoes his early view: "In the case of meaning, the constraints cannot uniquely fix the theory of interpretation. The reason, as Quine has convincingly argued, is that the sentences a speaker holds to be true are determined, in ways we can only partly disentangle, by what the speaker means by his words and what he believes about the world ... The remaining indeterminacy should not be judged as a

¹⁰³ *Inquiries* 27.

failure of interpretation, but rather as a logical consequence of the nature of theories of meaning."¹⁰⁴

In my opinion, the above objection concerning Davidson's extensional theory *does not* render the theory useless for the purpose of interpretation. If one (like Davidson) subscribes to the indeterminacy thesis, the generation of extensionally equivalent, rival interpretations must be tolerated to a certain extent. The preference for accepting one interpretation over its extensionally equivalent counterpart is made easier by the formal and empirical constraints of the theory that would limit the acceptance of grotesque theorems like (1) as interpretatively useful.¹⁰⁵ The important notion emerges here that if we give up the unjustified expectation that Davidson's aim is to define the uniquely correct meaning for each sentence of NL a speaker may utter, the theory in *this sense* need never be actually completed. Since the number of potential sentences in NL is infinite, the pairing of sentences of NL with sentences stating their truth-conditions can be expanded indefinitely, based on relevant empirical evidence. Conflicting theorems can always be discriminated through a revision using new evidence. Only if "all evidence is in", can one plausibly speak of completing the theory of meaning. This however, will never be the case, due to the generative (expansive) character of NL.

¹⁰⁴ Donald Davidson (1973) "The Material Mind", reprinted in D. Davidson, (1980) Essays on Actions and Events, pp. 245-259. Oxford: Clarendon Press. pp.256-7.

¹⁰⁵ *Inquiries* 224-25. Davidson takes indeterminacy to be analogous to reflecting different ways of "measuring" the same thing, like Fahrenheit and Centigrade are both used to measure the same temperature in different ways.

The expectation, that (within Davidson's theoretical framework), for a speaker, in order to be able to interpret sentences of NL, it is necessary to identify some unique meaning belonging to a particular sentence of NL, is unjustified for two main reasons: firstly, it misidentifies the aims of the theory as giving the meanings of sentences in NL.¹⁰⁶ Instead, Davidson's theory has been clearly directed towards describing the knowledge needed to interpret sentences of NL. From the very beginning, the focus has been on defining the general underlying ability of interpreters of NL, and not on *actually* giving the meaning of sentences in NL. The other reason for which the above expectation vis-à-vis Davidson's theory can be seen as mistaken is that (as shown above), this expectation ignores Davidson's consistently held views about the indeterminacy of meaning on one hand, and his holism extended to language as a whole to be interpreted on the other hand. The holistic view of language does not mean that there are clearly defined boundaries restricting what belongs to the domain of a certain language, or even to language as a theoretical concept. NL ought not be seen as a mere sum of its component elements, where the work of the theorizer would be finished, once all these elements are explicitly defined. Rather, the work of the theorizer can be considered completed, once the underlying principle accounting for the possibility of generating infinitely many sentences of NL could be correctly defined. So far, together with Davidson, we can be optimistic about having found the underlying principle of linguistic interpretation. Knowing the truth-conditions of a sentence, defined via recursion, based on the compositional character of NL, thus involving NL as a whole, does seem to depict

¹⁰⁶ The proper task of any particular theorem is, as Davidson puts it, to assign "the sentence a semantic location in the pattern of sentences that comprise the language". *Inquiries* 225.

the underlying principle of how interpretation of sentences in NL is possible.

However, the definition of truth-conditions for sentences in NL via recursion is possible only if the correct logical form or, in Davidson's words, the "canonical notation" of these sentences can be given, to replace *s* in the T sentences of the theory. Tarski shows that the truth of sentences in FL could be defined on the basis of the relation between expressions and objects of the world which can satisfy the expressions of FL. The definition of truth for Tarski is extensional and its logic only requires first-order predicate calculus. NL however contains many linguistic features that go beyond the resources of first-order logic. If Davidson's theory of meaning shall make use of a Tarskian truth theory, Davidson has to show that the extensional resources of that theory would suffice to depict the knowledge needed to interpret *any* sentence a speaker of NL may utter.

Among the most problematic features of NL are, as admitted by Davidson in "Truth and Meaning", counterfactual or subjunctive sentences; sentences about probabilities and about causal relations; adverbial modification, attributive adjectives and mass terms; sentences expressing propositional attitudes, and finally sentences constituting various speech-acts, such as for example imperatives, or interrogatives and so on.¹⁰⁷ My aim here is not to recite the various solutions offered by Davidson to these obvious

¹⁰⁷ To quote the footnote by Davidson, "For attempted solutions to some of these problems see *Essays 6-10 of Essays on Actions and Events*, and *Essays 6-8 of this book [Inquiries]*. There is a further discussion in *Essays 3,4,9, and 10*, and reference to some progress in section 1 of *Essay 9. Inquiries 34-35.*

problems, but rather to show his conviction that these problems could eventually be resolved, so that they would not obstruct the recursion of a Tarskian truth theory and thus, there is no reason to regard them with utter pessimism. Davidson envisioned two major phases an actual theory construction would take: first, only the segment of NL that can be put into the logical form or canonical notation, using first-order predicate calculus would be given truth-theoretical treatment. After that, each of the remaining types of sentences in NL would be matched with sentences, already put into canonical notation, in order to fall within the scope of the truth theory.¹⁰⁸

4.3 The Problem of Indirect Discourse

In the introduction of this chapter, I have referred to the specific problem of indirect discourse, to be discussed as part of a general problem involving Davidson's commitment to extensionality. NL is generally seen by many philosophers (among others by C. I. Lewis, Carnap, Russell, and of course by Frege), as being characteristically intensional in nature, that is, there may be sentences of NL, in which substitution of coextensional expressions do not preserve truth-value. This problem is especially evident in the specific context of indirect discourse, and this context is often taken as the paradigm example for the intensional nature of NL, flouting the principle of substitutivity with coextensionals *salva veritate*. Frege's solution to this problem was to consider the relevant expressions as having different semantical values in different contexts, that is, the meaning of an expression being its sense, for Frege, these expressions differed in

¹⁰⁸ *Inquiries* 29.

meaning. Davidson however, does not wish to involve meanings, or senses in his approach at all, for reasons already discussed. Consequently, he will have to devise a strategy, different from that of Frege, for maintaining the substitutivity of extensionally equivalent expressions. In the following I shall briefly discuss Davidson's solution concerning the problem of indirect discourse.

Davidson proposes to give a new account of the logical form and formal semantics of sentences of NL like

(3) Galileo said that the earth moves

According to the surface grammar of the sentence, (3) contains the sentence

(3.1) the earth moves

which in turn is composed of the singular term 'the earth', and of the predicate 'moves'. The difficulty with (3) is the same difficulty we saw with regard to the intensional 'means that' (see p.34 of this thesis). If the principle of substitutivity with coextensionals *salva veritate* holds, the singular term 'the earth', ought to be replaceable with any other singular term as long as that term is coextensional with 'the earth'. Such replacement ought not result in a change of the truth-value of the sentence containing the relevant singular term. We can do such a replacement in the sentence (3.1) 'the earth moves' (e.g., 'the planet inhabited by humans moves'), but we cannot do the replacement in (3) without affecting change in its truth-value. The sentence

(3*) Galileo said that the planet inhabited by humans moves

is probably false, since Galileo may have never made a statement to that effect.

Frege would suggest that 'the earth' has different referents in the contexts of (3) and (3.1). In the former context, 'the earth' refers to the expression used by Galileo, while in the latter context 'the earth' refers to our planet in the customary sense. Frege's solution, namely, to distinguish the different referents (senses) of the expression 'the earth' in the context of (3.1), as opposed to the context of (3), will have to be rejected by Davidson by the necessity of explaining the fact that NL is learnable. NL could not be learned if its expressions had an unrestricted number of referents, since one could never be sure which referent has been referred to by the speaker, given the underdetermination of referents by evidence. Davidson's strategy is to offer an alternative grammatic division of (3), radically different from the customary analysis of surface grammar, as stated above. According to Davidson, in (3), there are two *separate* sentences involved:

(3.1) the earth moves
which contains the expression that seems to resist substitutivity, and

(3.2) Galileo said that
which attributes the case of indirect discourse.

These two sentences, when taken one at the time, escape the problem of substitutivity with coextensive expressions. Substitution with coextensive terms can be afforded without problem in (3.1), and no substitution at all is involved in (3.2). "... from a semantic point of view the content-sentence in indirect discourse is not contained in the sentence whose truth counts, i.e. the sentence that ends with 'that'."¹⁰⁹

¹⁰⁹ *Inquiries* 106.

the sentence (3.1) from the report of samesaying, we can explain why substitution with coextensional terms for 'the earth' would change the truth-value of (3), without having to rely on intensions.

An objection could immediately be raised concerning the definition of the relation 'samesaying', since it contains the expression 'mean'. Are not meanings the very things Davidson wanted to avoid at all costs? He has met this type of criticism, already in "Truth and Meaning", by distinguishing between definitions related to particular idioms in NL, and using intensions in the theory of meaning as such. The former refers to individual concepts in NL, the latter to the logical form of NL, regardless of which particular concepts are contained in NL.¹¹²

The above, crucial distinction gained new emphasis when Davidson replied to John Foster's charges against his extensional theory of meaning. In "Meaning and Truth Theory"¹¹³, Foster takes Davidson's thesis to be "that a theory of truth meeting certain formal and empirical constraints serves as a theory of meaning for one who knows that it meets these constraints" (i.e., we need to know both the facts the theory states about NL, and that, in knowing these facts, we are T-theorizing).¹¹⁴ Foster's charge is that the facts stated by the theory do not reveal that these facts are T-theoretical: if the T-

¹¹² *Inquiries* 31.

¹¹³ John A. Foster (1976) "Meaning and Truth Theory", in G. Evans and J. McDowell, eds. Truth and Meaning: Essays in Semantics, pp. 1-32. Oxford: Clarendon Press, 1976.

¹¹⁴ *Ibid.*, 18.

theoricity of Davidson's theory is merely entailed by the theory, that entailment is an intensional feature of the theory, and, since Foster mistakenly identifies Davidson's goal as being the construction of a theory of meaning that does not use intensional concepts, Davidson's "grand design is in ruins", according to Foster.

Davidson, in "Reply to Foster",¹¹⁵ proposes to use his theory of indirect discourse to meet Foster's seemingly devastating criticism. Given the following sentence commenting on the truth theory

(4) the theory of truth entails that 'snow is white' is true iff snow is white

the semantic structure of (4) will be given as a relation between an utterance of

(4.1) 'snow is white' is true iff snow is white

and the sentence

(4.2) the theory of truth entails that

where 'that' is taken to be a demonstrative. Thus, coextensive sentences can be substituted for (4.1). Davidson suggests that "If a theory T entails that 'Snow is white' is true in English if and only if snow is white, then T has as logical consequence a sentence synonymous with my utterance of "'Snow is white' is true in English if and only if snow is white."'¹¹⁶ This quote is a statement by Davidson about his theory, using the word 'synonymous'. The fact stated in my previous sentence must not be confused with the falsity, that Davidson's theory itself makes use of the intensional concepts of

¹¹⁵ *Inquiries* 171-179.

¹¹⁶ *Inquiries* 178.

entailment, or of synonymy. Statements *about* the theory are not statements *of* the theory. Since Foster's main criticism is based on this confusion, it does not refute Davidson's approach.

The project of the possible completion of the theory can be interpreted in various ways: in a sense, the theory need not ever be finished, as there can be infinitely many sentences of NL paired with their relevant truth-conditions, in order to yield infinitely many T-sentences. Convention-T of the Tarskian, extensional truth theory could in principle be applied to any sentence of NL, as long as an adequate logical form could be found for that sentence. On the other hand, the fact that NL is learnable calls for a finite statement of the underlying principle that allows infinite creativity in NL. The statement of the truth-conditions for sentences in NL, meeting certain formal and empirical constraints could in principle serve as referring to such an underlying principle. In this sense, it is possible for the truth theory for NL to be completed. Davidson has offered a programmatic view of how the details of that theory would have to be worked out, and in many cases, he himself suggested the solution for some of the details. In this chapter I have merely tried to indicate that Davidson had an optimistic outlook with regard to the possible completion of his theory. He thought from early on that the relevant principle allowing the possibility of interpreting speakers' utterances was found in the knowledge of truth-conditions, and much of his effort went into showing how this relevant knowledge could be depicted in the form of Convention T. However, from the very beginning, he clearly stated that for the interpreter in order to know the relevant truth-

conditions it was necessary to consider what conditions the speaker held relevant. The Principle of Charity has always been an indispensable element in Davidson's theory of meaning. He did shift his attention from working out the details of canonical notations to explicating the crucial relevance of the Principle of Charity (in light of his views on the interdependence of beliefs and meaning). This shift however, ought not to be interpreted as the abandonment of his early views in despair, but rather, as a change in priorities, necessitated precisely because of Davidson's theoretical commitments. The critical analysis of this change in priorities will be the topic of the next chapter. I shall follow Davidson's own outline in touching upon the issues of the empirical testability and of the interpretability of the theory. My main focus, however, shall be to show the consistency in Davidson's theoretical commitments.

CHAPTER FIVE
REJECTING THE CONCEPT 'LANGUAGE':
A NECESSARY OUTCOME OF THE THEORY

The next issue to be addressed is *the empirical testability* of Davidson's theory. In general, if a theory of meaning is meant to be a theory for a particular NL, one must have some way of deciding, whether or not for that language the theory is correct, and that is established through empirical testing. If the object language is contained in the metalanguage, say, if English is the language for which a theory of meaning is being construed, and the language of the theory happens to be English also, and moreover, the person doing the theorizing is a competent speaker of English, then, if she finds sufficiently many of the T-sentences correctly stating the truth-conditions of sentences in English, she can consider the theory of which these T-sentences are consequences as correct.

The third issue suggested by Davidson's outline is the *interpretability* of the theory, that is, whether the knowledge described by the theory is indeed the kind of knowledge that would enable a listener to interpret any sentence a speaker may utter. A specific connection between the issues of testability and of interpretability is found in Davidson's employment of the core concept of his theory, the Principle of Charity. The application of the Principle of Charity is meant to provide the evidence for the empirical correctness of his theory, and it also serves to allow the interpretativeness of the theory.

The empirical testability of the theory of meaning becomes the central issue on which the entire theory depends, since it is Davidson's search for empirical evidence that necessitates the employment of the Principle of Charity. An analysis of the search for empirical evidence also explains how Davidson has been gradually led to rejecting the concept of 'language' as a useful tool in interpretation. This involves the central claim of my own thesis, namely, that Davidson's rejection of the concept 'language' was a necessary outcome of his consistent adherence to his fundamental views on meaning. The defence of that claim is the purpose of the present chapter. In this chapter, I shall attempt to show increasingly explicit indications of Davidson's skeptical concern with respect to 'language' as a useful theoretical construct for a theory of linguistic competence, in agreement with his consistently held views with respect to beliefs and meaning.

The first segment discusses various considerations of what could provide evidence for the correctness of the theory. These considerations not only lead inevitably to the employment of Davidson's central concept of the Principle of Charity, but they also suggest that interpretation may succeed regardless of whether the theoretical concept 'language' can be given a clear definition, i.e., whether the particular language for which the theory is a theory of meaning can be clearly identified.

The Principle of Charity became the central concept of Davidson's theory of meaning, more and more explicitly referred to in Davidson's papers. Its crucially important role in the theory will be given a detailed analysis in the second segment of this

chapter. In his early work Davidson did appeal to linguistic conventions that would serve as the touchstone compared to which the correctness of the theory could be evaluated. In the third segment of this chapter I shall argue that Davidson had to reject the significance ascribed to conventions in communication, precisely because he wished to remain committed to his early views concerning meaning. Davidson consistently adhered to his fundamental views on meaning, and this commitment governed his rejection of conventions, and led ultimately to his more recent rejection of the very concept of 'language', discussed in the fourth segment. The final segment of the chapter shall touch upon the issue of the interpretability of the theory, and here again, the central importance of the Principle of Charity will be emphasized.

5.1 The Empirical Testing of The Theory

An example of the empirical testing of the theory has been given in the beginning of this chapter as follows. If both speaker and interpreter speak English, and a theory of meaning is to be given in English as the metalanguage, then, if enough T-sentences are found to be correctly depicting the truth-conditions for sentences of English as the object language, the theory can be considered correct. However, the information concerning whether or not the object language to be interpreted is contained in (or identical to) the metalanguage of the theory, *is not available* to the theorizer *in advance*, so that this information cannot be used to test the correctness of the theory, nor can it be used to interpret the speaker's utterances. If the object language *O* is foreign to the interpreter, it is not at all evident, *prior* to the theory construal, whether or not *O* is contained in the

metalanguage *M* used by the interpreter. The mere fact, that there can be sentences of *O* that have their homophone in *M*, is not sufficient for establishing whether or not *O* is contained in *M*, since, as the Empedokles-Empedocles example has shown, despite homophony, it remains a mere assumption that the languages in question are identical (see p.23 of this thesis). The only way to establish *O*'s containment in *M* would be if we knew that the two homophonous sentences of *O* and of *M* have the same meaning. But the meaning of sentences in *O* is just what we are after, so that we cannot tell in advance, without begging the question, whether a particular sentence of *O* is contained in *M*.

According to Davidson, the above reasoning applies to *domestic* interpretation as well. If the sharing of homophonous sentences spoken by the speaker and by the interpreter does not in itself prove that the language spoken by the speaker is contained in the language of the interpreter, then *there is no reason to assume that just because both speaker and interpreter "speak English", i.e., utter homophonous sentences, they are speaking the same language*. Let us recall an extremely relevant statement of Davidson, written as early as 1973:

The problem of interpretation is domestic as well as foreign: it surfaces for speakers of the same language in the form of the question, how can it be determined that the language is the same? Speakers of the same language can go on the assumption that for them the same expressions are to be interpreted in the same way, but this does not indicate what justifies the assumption. All understanding of the speech of another involves radical interpretation.¹¹⁷

¹¹⁷ *Inquiries* 125.

Davidson's reasoning with respect to domestic interpretation seems to go against our intuitions as ordinary speakers. We usually do not have doubts with regards to whether or not our interlocutor speaks the same language we speak, or intend to interpret. We simply take it for granted that our mutual interpretation of one another's utterances succeeds precisely because both of us speak, say, English. However, as Davidson's deceptively simple Empedocles example shows, our sharing of a common language is a mere assumption, which ought to be justified if it is supposed to serve as the fundamental principle of our success in interpretation. Davidson challenges us to *prove* this assumption before accepting it as a fact.

The implications of the notion of radical interpretation involving domestic verbal exchanges are twofold: on one hand, due to the lack of sufficient empirical evidence, homophonous sentences uttered by speakers of a linguistic community cannot be said to belong to the same language. Moreover, we do not even have a clear notion of what would count as a 'linguistic community'; and if we do not have a clear theoretical definition of this common sense notion, then we can hardly use it as part of the explanation. Consequently, interpretation will not be aided by the mere assumption of shared language spoken by a given linguistic community: unless this assumption is justified, interpretation must always be considered radical. On the other hand, if interpretation is always radical, this means that one must find a way to test the correctness of the theory of meaning, *regardless of* whether or not the object language is contained in the metalanguage, since this circumstance is not known to the theorizer *prior* to the

application of the theory. This second, practical consideration with respect to the empirical testability of the theory suggests that the theory may do the work it is intended to, without having to justify the assumption that the sentences to which the theory applies, belong to a particular NL. Davidson's theory of linguistic competence, if proven successful, may not have to be given up in light of the (somewhat paradoxically sounding) fact that at this point, the theoretical concept 'a particular NL' or 'language' for short, seems to have a limited use in radical interpretation.

The problem remains, what kind of evidence is needed to consider radical interpretation possible. According to Davidson,

We interpret a bit of linguistic behavior when we say what a speaker's words mean on an occasion of use. The task may be seen as one of redescription. We know that the words 'Es schneit' have been uttered on a particular occasion and we want to redescribe this uttering as an act of saying that it is snowing.¹¹⁸

Here Davidson uses a German sentence as an object of radical interpretation, but he does so just in order to "keep assumptions from going unnoticed," that is, to sharpen our understanding of what he thinks goes on in general when we interpret one another's utterances, regardless of what customarily identified language we consider.

The behaviourist Quine based his radical translation on observable bits of linguistic behaviour, namely, the assent to or dissent from sentences by speakers. Davidson's requirement with respect to evidence is that it be non-semantical. Although

¹¹⁸ *Inquiries* 141.

Davidson insists that evidence for the truth theory be non-semantic, yet, he describes the evidence as speakers' *holding sentences true* under certain conditions. 'Holding sentences true' seems to refer to a *semantic* fact concerning speakers' attitudes toward sentences they utter. Davidson's way to escape the seeming inconsistency in his requirements vis-à-vis the evidence of the theory is to take the requirement of 'holding true' as a *theoretical posit* that is needed to get his theory working. He defends his reliance on this particular belief, while rejecting the employment of any other beliefs prior to the empirical application of the theory by claiming that the interpreter can know that the speaker holds a sentence true *prior to* knowing what that sentence means, so that the semantic fact of holding the sentence true does not appear in the specific theoretical statement (i.e., the T-sentence) giving the truth-conditions for that sentence. While Davidson's exclusive preference for the belief of 'holding true' as a useful theoretical posit may be questionable, his strategy to justify its appearance among the required *non-semantic* facts of evidence is quite acceptable. Once again Davidson's strategy here is to carefully separate statements *about* the theory from statements *of* the theory. Statements concerning what would count as evidence describe how to use the "machinery" of the theory, and these statements do not appear in T-sentences depicting the truth-conditions of sentences to be interpreted.¹¹⁹

¹¹⁹ *Inquiries* 133-34.

The principle that interpretation proceed under the assumption that, given certain conditions, speakers hold sentences uttered by them true, is a normative principle, without which the theory cannot be put into action. Let us see how it works by first considering what the speaker does when uttering a sentence, and then the strategy the interpreter would have to follow in interpreting the uttered sentence. As Davidson emphasizes, a speaker holds a sentence true for two reasons: because she believes that the sentence is true, and because the sentence means what it means. If, for example a speaker S holds the sentence¹²⁰

(1) snow is white

true, it is because

(a) S believes that snow is white

and because

(b) 'snow is white' means that snow is white

So, if the speaker utters sentence (1), she does so because she believes that snow is white and because sentence (1) is the very sentence to be used to express just that belief. If we now turn to the ways the interpreter would go about interpreting the speaker's uttering sentence (1), the following strategy could be attributed to the interpreter. If we rewrite the intensionally expressed (b) in the extensional style of the theory, (b) becomes

(b*) 'snow is white' is true iff snow is white

in which case (b*) is analogous to the relevant T-sentence for the correctness of which

¹²⁰ For the sake of simplicity in this particular bit of analysis I have ignored the requirement that the truth-theory for NL must be relativized for speaker, time and place. The sample sentences are to be taken by the reader as an abbreviated version of the relativized truth-theory.

an interpreter would be seeking evidence. The correctness of (b*) cannot be justified by an appeal to the proposition expressed in (b) since this is the proposition the meaning of which the interpreter is trying to get at. The interpreter knows that snow is white (which entails that the interpreter believes that snow is white). So if the interpreter makes the assumption that the speaker holds sentence (1) true, the interpreter can consider (b*) as the relevant theorem that interprets the speaker's uttering (1). What is crucial here is that the interpreter can consider (b*) as the relevant theorem only *after* having assumed that the speaker held (1) true. Thus, since the meaning of the sentence (b) cannot be evidence for the correctness of (b*), the remaining alternative is to consider (a) as the evidence in this particular case. Accordingly, from the interpreter's point of view, says Davidson, the evidence for the following T-sentence of the theory

(T) 'snow is white' is true iff snow is white

is constituted by the interpreter's assumption concerning *the speaker S's belief that 'snow is white' is true*.¹²¹ There are at least two serious problems with this statement: first, how does the interpreter know what S's beliefs are, and second, how does the interpreter know that S's beliefs are not mistaken. Davidson's answer to these questions is to apply the Principle of Charity "across the board."¹²²

¹²¹ Davidson has stressed the interdependence of meaning and belief in several of his papers. If speakers' beliefs cannot be separated from the meanings of the sentences speakers utter, and if the meanings of the sentences cannot be clearly individuated (according to the indeterminacy thesis) then the beliefs of speakers cannot be clearly individuated either. Davidson justifies his choosing a belief as evidence for the correctness of the theory by claiming that 'holding sentences true' is one single belief applicable to all sentences, and that the speaker's relevant attitude can be identified prior to interpreting the sentence. See, for example, *Inquiries* 134-35;

¹²² *Inquiries* XVII.

5.2 Developments in Davidson's Employment of The Principle of Charity

In its first appearance, the Principle of Charity meant maximizing agreement between interlocutors in terms of what sentences they hold true. The interpreter must *assume* that her interlocutor reasons in ways significantly similar to the interpreter's own ways (e.g., not accepting logical inconsistencies), and as a result, the speaker holds true sentences in ways significantly similar to the interpreter's own ways. This means that interpretation proceeds according to the interpreter's standards, expressed by the assumption of the interpreter that the speaker believes what the interpreter thinks is true.¹²³ A decade or so later, in the introduction to the 1984 edition of the *Inquiries*, Davidson emphasizes that "the aim of interpretation is not agreement but understanding." The "right sort" of agreement serves as a means of understanding, but it is not more easily specified than "having good reasons to believe" that the speaker holds a sentence true.¹²⁴

The answer to the question, how the interpreter knows what the speaker's beliefs are, must be given in terms of indeterminacy. If meanings and beliefs are interdependent, there must be some indeterminacy about attributing beliefs to a speaker, similarly to holding the meanings of sentences indeterminate. In that sense, the interpreter cannot be said to *know* (with certainty) what the speaker believes, but merely to guess the speaker's relevant beliefs *after* having interpreted the meaning of the speaker's utterance. However,

¹²³ *Inquiries* 137.

¹²⁴ *Inquiries* XVII.

no attempt to interpret the meaning of the speaker's utterances can take place unless one *assumes* that the sentences uttered by the speaker are held true by her just in case they would be held true by the interpreter.¹²⁵ That is precisely why the employment of the Principle of Charity is indispensable for interpretation.

In the case of the so-called observation sentences the truth-conditions under which these sentences are uttered are empirically observable for both interlocutors, so that, assuming that their respective sensory apparatus functions in similar ways, agreement about what sentences they are holding true is more easily ascertained, even though the empirical evidence alone leaves open the possibility of different interpretations. The Principle of Charity functions as a constraint on possible interpretations by ruling out the alternatives that mean attributing irrationality to the speaker. It is worth quoting in full how Davidson describes the functioning of Charity:

Some disagreements are more destructive of understanding than others, and a sophisticated theory must naturally take this into account. Disagreement about theoretical matters may (in some cases) be more tolerable than disagreement about what is more evident; disagreement about how things look or appear is less tolerable than disagreement about how they are; disagreement about the truth of attributions of certain attitudes to a speaker by that same speaker may not be tolerable at all, or barely. It is impossible to simplify the considerations that are relevant, for everything we know or believe about the way evidence supports beliefs can be put to work in deciding where the theory can best allow error, and what errors are least destructive of understanding.¹²⁶

¹²⁵ *Inquiries* 137.

¹²⁶ *Inquiries* 169.

The employment of the Principle of Charity could be misinterpreted as ruling out any mistake with respect to the truth of a sentence. If one (mistakenly) thinks that the employment of the Principle of Charity means that whenever the interpreter holds a sentence true, the interpreter will assume that the speaker holds it true, too, then the question arises, how could either the speaker, or her interpreter recognize that they may be mistaken concerning the truth of that sentence? Davidson's answer, at least for some time during the seventies, has been that a theory of interpretation does not apply to isolated acts of interpretation between particular individuals, but to the general linguistic practice of entire speech-communities. This is the view suggested by his discussion of Kurt's utterance 'Es regnet', where Kurt is said to be a member of the German-speaking speech-community (see p.59). Kurt's uttering 'Es regnet' was taken to mean something to be uttered just in case the specific condition, i.e., it is raining obtains in the vicinity of the speaker, Kurt. My sentence above contains a piece of information, a competent member of the German speech-community is said to know. This bit of knowledge

- (2) Speakers of the German-speaking speech-community hold true-in-German 'Es regnet' when and only when it is raining in their vicinity.

is part of the public standard concerning the truth of sentences as held by that particular speech-community.

The publicly shared knowledge of truth-conditions serves here as the standard according to which error can be made intelligible. Individual speakers of the community can have mistaken beliefs about the truth-conditions of individual sentences *only relative to a publicly accepted standard*. They could entertain false assumptions for example, that

certain conditions (under which a sentence is held true by the public standard), do obtain, when in fact these conditions do not obtain. For example, even though the public standard of (2) may be known to the speaker, she may be mistaken in her belief that it is raining, and so, utter the above sentence without the relevant condition obtaining. This generalized form of evidence implies that the correct assignment of truth-conditions prescribed by public standards is expected to prove the speaker's utterance of the sentence mistaken (in the absence of relevant truth-conditions), and consequently her utterance *on that occasion* false. If the sentence 'Es regnet' is true if and only if it is raining, then Kurt's uttering 'Es regnet' in a sunny afternoon in Hamburg would be false. If the assumption is that in general, speakers of a speech-community are believers of truths, where would that leave the individual speaker, observed to hold sentences consistently true, where the same sentences are held false by the accepted public standards? According to Davidson, in the publicly applicable sense of the Principle of Charity, such a speaker (lest we wish to attribute to her "too much silliness"), must be assumed to be speaking a *different language*. At this earlier stage, it appears already that the Principle of Charity would counsel one to abandon the assumption of shared language rather than abandon rationality attributed to one's interlocutor. This observation, too, suggests that Davidson attributed priority to human rationality over 'language' in the theory of interpretation.

Davidson's earlier appeal to public standards, or in other words, to linguistic conventions is shared by most philosophers. By 'linguistic convention', philosophers in general mean that for successful communication to take place between interlocutors, there must be linguistic conventions governing the use of sentences to mean what they mean. Individual speakers must respect such conventions if they wish to be understood.¹²⁷ Members of a particular speech-community have acquired a set of beliefs concerning the use of sentences to communicate with one another, and this set of relevant beliefs constitutes their particular linguistic convention. According to David Lewis, convention is a regularity R in action and belief, a regularity in which more than one person must be involved. Everyone involved conforms to R and believes that others also conform. Moreover, the belief that others conform to R gives all involved a good reason to conform to R.¹²⁸ If, according to this view, the concept 'language' were taken to refer to a set of conventionally assigned meanings to sentences, communication could not succeed without language, interpreted in this sense.

Davidson, in his more recent writings clearly rejected the notion of linguistic convention as *the* relevant principle governing communication. My discussion of Davidson's theory of linguistic competence has finally reached the point where I am able to argue directly my thesis, namely, that Davidson cannot be charged with inconsistency with respect to his earlier versus his recent views on meaning. In the following I shall

¹²⁷ This is for example the view held by H.P. Grice, in Grice (1975) "Logic and Conversation."

¹²⁸ David Lewis (1975) "Languages and Language", in K. Gunderson, ed. Language, Mind and Knowledge, pp. 1-27. Minneapolis: University of Minnesota Press. pp. 5-6.

attempt to show that, in rejecting the significance of the role attributed to linguistic conventions in terms of interpreting speakers' utterances, Davidson remained consistent with his earlier theoretical commitments, and that the same rejection of conventions, *fuelled by his consistently held commitments*, had finally led to Davidson's rejection of the theoretical concept 'language' itself.

5.3 The Rejection of Linguistic Conventions

In the introduction of this chapter I have already pointed out that Davidson had to abandon his earlier appeal to linguistic convention because that appeal was inconsistent with his views concerning the interdependence of beliefs and meaning. The notion of meaning and beliefs being interdependent is of fundamental importance in Davidson's views on interpretation, consistently adhered to by him throughout his theory of meaning. Let me repeat the clearest formulation of Davidson's strong conviction concerning the interdependence of meaning and belief as found in the following quote:

beliefs and meanings conspire to account for utterances. A speaker who holds a sentence to be true on an occasion does so in part because of what he means, or would mean, by an utterance of that sentence, and in part because of what he believes. If all we have to go on is the fact of honest utterance, we cannot infer the belief without knowing the meaning, and have no chance of inferring the meaning without the belief.¹²⁹

If a linguistic convention is seen as a set of beliefs attributed to speakers of a particular speech-community (where the attribution of beliefs to speakers would take place prior to interpreting their utterances), based on which the meaning of sentences uttered by them

¹²⁹ *Inquiries* 142.

is definable, this prior attribution of beliefs to speakers is *inconsistent* with Davidson's fundamental views on the interdependency of beliefs and meaning. No particular beliefs concerning meaning can be attributed to speakers *prior to* the act of interpretation. Instead, the speaker's beliefs concerning their intended meaning, and the literal meaning of the sentence itself must be interpreted *simultaneously*. Only from having interpreted the meaning of the speaker's utterance can the listener make guesses as to what the speaker's relevant beliefs or attitudes are vis-à-vis uttering that sentence (see my previous discussion on page 97). Consequently, convention (a "prior theory"), cannot be the relevant principle that makes interpretation possible. Rather than disavowing his fundamental views concerning the employment of a truth theory as a theory of meaning, Davidson's early appeal to linguistic convention being inconsistent with the theory as a whole had to be given up. Davidson acted *precisely in the name of preserving consistency in his views*, when he, in the 1982 paper "Communication and Convention", argued that communication did not require convention.¹³⁰

According to Davidson, there is no publicly scrutinizable convention connecting the meaning of a sentence and the speaker's particular intention to utter just that sentence. For example, there is "no known, agreed upon, publicly recognizable convention for making assertions" (where making an assertion is understood as saying what one believes to be true), since all the interpreter can observe is that the speaker is representing herself as believing what she asserts, and consequently, her sincerity can only be assumed, but

¹³⁰ *Inquiries* 265-280.

not known. The actual attitude of the speaker towards the proposition represented by her act of uttering the sentence is hidden from public scrutiny, so that it cannot be operating according to conventional, public rules. In other words, the real purpose of the speaker for uttering a sentence and the meaning of that sentence cannot be linked by convention.

The widely held assumption, says Davidson, that the use of the sentence in the sense of its so-called literal meaning is conventional is mistaken, too, for similar reasons. Both the *meaning* of the sentence itself and the speaker's *intention* (i.e., *the speaker's meaning*) to utter the sentence in the sense of its literal meaning contribute to the concept of *literal* meaning. Literal meaning as such depends on the relevant intention of the speaker, which is not publicly observable (that's how liars can succeed). "The convention must pick out, in a way understood by both speaker and hearer, and in an intentionally identifiable way, those cases in which the ulterior purpose directly yields the literal meaning."¹³¹ This, however, cannot be done, since it would require that the speaker *be* sincere - for which there is neither guarantee, nor a conventionally identifiable signal. Nor is there, as often mistakenly claimed, a need for sincerity (guaranteed by convention) for the sake of language acquisition, since we are capable of acquiring the use of speech despite the fact that our parents and playmates are not always "truthful", "sober" or "genuinely assertive".

¹³¹ *Inquiries* 273.

According to Davidson, *meaning is autonomous*, in the sense that the speaker's intention and the meaning of the sentence uttered are independent, in so far as the latter cannot be derived from the former.¹³² The so-called illocutionary force of a sentence uttered can sometimes be derived from its syntax, as for example an imperative sentence "labels itself" as such, indicated by the imperative mood; but that is quite independent from the intention of the speaker, so that there is no convention connecting the speaker's intention and meaning of the sentence. No reference can be made to a so-called "standard use" of an imperative sentence; since, if convention applies at all, it applies only to the literal meaning of the sentence, which is identical in both standard and non-standard uses.

Davidson questions even the "platitude" that the literal meaning of the sentence is conventionally assigned within a speech-community. The knowledge needed for mutual understanding cannot be the sharing of common language. We have already seen the same idea suggested by the Empedocles example, and here, Davidson reiterates it by claiming that "Different speakers have ... different vocabularies, and attach somewhat different meanings to words ... as interpreters we are very good at arriving at a correct interpretation of words we have not heard before, or of words we have not heard before with meanings a speaker is giving them."¹³³ Thus, communication does not demand conformity between speakers with respect to meanings. In general, Davidson also denies that any regularity governing interpretation, necessary for communicative success could

¹³² *Inquiries* 274. See also my first mentioning of Davidson's rejection of intention-based theories of meaning, for precisely the same reasons (see p.57 of this thesis).

¹³³ *Inquiries* 277.

be found. Naturally, regularity can be found in what forms the skeleton of what one calls a language: the pattern of inference and structure created by the logical constants, and that is the part of language to which the formal description applies. But if "knowing the language" were the relevant principle allowing communication to succeed, it would mean that speakers would go about interpreting one another's utterances according to some conventional rules known *before* and *independently* of the act of interpretation. Davidson says that what is crucial to the success of communication is that speaker and interpreter *coincide* in the rules (or theories) applied *after* the utterance has been made, and those theories cannot be governed by regularity, since these theories are "made up" as the interlocutors go along interpreting one another's utterances. Davidson does not deny that convention plays a practical role in conditioning our ways of language acquisition or use; he only claims that the publicly held linguistic standards provide only the starting point to interpretation, otherwise convention does not help to explain what is basic to linguistic communication.¹³⁴

5.4 The Rejection of 'Language'

Davidson's views on the contingent role convention plays in linguistic communication have been further developed in the infamous 1986 paper, "A Nice Derangement of Epitaphs".¹³⁵ The strange title of the paper refers to a *malapropism*,

¹³⁴ *Inquiries* 280.

¹³⁵ Donald Davidson (1986) "A Nice Derangement of Epitaphs", in E. LePore, ed. Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson, pp. 433-446. Oxford: Basil Blackwell, 1986.

where the so-called literal or "first-meaning" of the expressions has been misapplied, contrary to the speaker's intentions. Davidson finds it significant that in general, the hearer of such malapropisms or of other deviant, non-standard uses of first-meanings has no difficulty understanding the speaker in the way the speaker intends. In this paper Davidson provides what he takes to be the customary definition of the concept 'language' in the following way: "...in the case of language the hearer shares a complex system or theory with the speaker, a system which makes possible the articulation of logical relations between utterances, and explains the ability to interpret novel utterances in an organized way."¹³⁶

With respect to first-, or literal meanings in a language, Davidson says that these are customarily held to be *systematic* (i.e., there must be systematic relations between the meanings of the utterances that allow interpretation on the basis of the semantic properties of the component parts; cf. Davidson's notions with respect to semantic holism and compositionality); first-meanings are held to be *shared* in the sense that *all* speakers of an NL rely on systematic methods of interpretation; and finally, first meanings are held to be *known in advance* of occasions of interpretation via linguistic conventions. While Davidson thinks that claims concerning the systematic and the shared character of first-meanings are acceptable, he has trouble combining these with the notion of conventionality, particularly when applied to the interpretation of malapropisms.

¹³⁶ Ibid., 436.

We have seen that in his earlier papers Davidson himself attempted to explain the possibility of interpreting errors of speakers by appealing to conventions. In this paper he says that conventional and prior knowledge of meanings cannot explain listeners' ability to interpret mistakes, malapropisms and other deviant linguistic behavior. Both speaker and listener may have "*prior theories*", e.g., knowing that one's interlocutor is a notorious liar, about how they would interpret a potential utterance, based on their previous experience with other speakers, but these prior theories are only *contingently* relevant to success in communication. In Davidson's words:

I have distinguished what I have been calling the *prior theory* from what I shall henceforth call the *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.¹³⁷

What is *crucial* for the success of communication, argues Davidson, is that the speaker's and listener's theories *converge* in the actual interpretative act of an actual utterance. Prior theories must be adjusted according to the circumstances prevailing during the particular communicative exchange: they turn into "*passing theories*", geared to the occasion, and these passing theories are the ones actually used to interpret the utterance, where the meaning of the uttered sentence, and the particular belief or intention of the speaker concerning the uttered sentence are interpreted together. "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

¹³⁷ Ibid., 442.

intends the interpreter to use. Only if these coincide is understanding complete."¹³⁸

By now it is obvious that neither the contingent prior-, nor the crucial passing theories can be ruled by conventional regularities. The former depend on how well one knows one's interlocutor, the latter on the particular conditions that obtain at the occasion of the utterance and its interpretation. Neither type of theory fits the customary concept of 'language' in the sense of adequately describing the language a person knows, or the competence of speakers as characteristically linguistic. The relevant ability of speakers to construct convergent passing theories for occasional speech transactions lies in "wit, luck and wisdom", in knowing "the ways people get their point across", and "rules of thumb for figuring out what deviations from the dictionary are most likely."¹³⁹ Since appeal to convention does not illuminate communicative competence, convention and its linguistic manifestation, language, must be rejected. Let us reiterate Davidson's famous concluding passage that gave rise to widespread controversy and let us see whether in light of the previous discussion that passage still seems controversial, or rather, it seems as the *natural* outcome the theory *had to* result in:

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

¹³⁸ Ibid., 442.

¹³⁹ Ibid., 446.

appeal to conventions.¹⁴⁰

As it is now evident from this quote, Davidson's rejection of 'language' applies to the customarily presumed role 'language' is supposed to play *in the specific context of interpretation*. What Davidson has concluded is that it is not by virtue of a shared, learned language that we succeed in interpreting one another's utterances.¹⁴¹ He does not wish to deny the importance of the notion of a shared language in other contexts: we indeed do master languages which we share with our interlocutors. But the skills thus acquired cannot *alone* account for our interpretive success. What is crucial for successful interpretation is not a shared language but the employment of the Principle of Charity, in the form of the so-called "passing theories" (to be understood as prior theories turned into passing theories in the actual act of interpretation). And passing theories are just *passing*: valid only for the period of time in which the interpretive act takes place, geared to the occasion. These passing theories clearly cannot be learned, mastered or born with. Speakers cannot acquire a prior, conventionally ruled, shared knowledge of passing theories. If the employment of passing theories constitutes the *decisive* principle underlying interpreters' competence, then the significance of conventions and language is clearly lessened. Convention and language, while important in other contexts, in the specific context of interpretation cannot play a significant role.

¹⁴⁰ Donald Davidson (1986) op. cit. p.446.

¹⁴¹ It is interesting to note that in fact, as early as 1975, Davidson has already expressed such views. In "Thought and Talk" he wrote: "... it is worth pointing out that the notion of a language, or of two people speaking the same language does not seem to be needed here. Two speakers could interpret each other's utterances without there being, in any ordinary sense, a common language. (I do not want to deny that in other contexts the notion of a shared language may be very important)." *Inquiries* 157.

5.5 The Interpretability of The Theory

Davidson's choice of an extensional truth-theory for a theory of (characteristically intensional) meaning has often been charged with having no interpretative force; i.e., that mere knowing of T-sentences would not suffice to interpret the meaning of sentences uttered by a speaker.¹⁴² I do not wish to go into minute details of the Dummett-Davidson debate, since, as I have emphasized before, my purpose in this thesis is not to defend Davidson's theory against specific criticisms, such as the one concerning interpretability, but rather, to defend the theory as being *consistent in its entirety*. With that aim in my mind, I would like to mention some of the reasons why a critical approach focussing on the T-sentences may be generally mistaken, or unjustified.

Whether the knowledge described by the theory is indeed the kind of knowledge that would enable a listener to interpret any sentence a speaker may utter, depends on what elements of Davidson's theory one takes to belong to the description of that knowledge. It would be unfair to arbitrarily select one particular element of the theory (such as for example the extensional, truth theoretical T-sentences) and evaluate the interpretability of the theory based on this single isolated feature, while neglecting other, crucially important features of the theory. This arbitrary practice would demonstrate a serious neglect of what Davidson emphasized throughout his relevant work, namely, that the theory ought to be evaluated in light of both the *formal*, as well as the *empirical* constraints imposed on it by Davidson himself. In the following, I shall simply list some

¹⁴² See for example M. Dummett (1976) "What is a Theory of Meaning? (II)", in G. Evans and J. McDowell eds. Truth and Meaning: Essays in Semantics, pp. 67-137. Oxford: Clarendon Press, 1976.

of the features that cannot be ignored when evaluating the interpretability of Davidson's theory.

Davidson has stressed that knowing how to interpret an utterance of the speaker is not simply knowing the appropriate T-sentence; it is knowing that some sentence is true if and only if some condition obtains and, knowing that that sentence is composed out of parts which feature in other sentences which are true if and only if other specified conditions obtain. The previous sentence appeals to Davidson's *holism* with regard to how sentences obtain their meaning by virtue of being embedded in the totality of sentences spoken by a group of speakers. Part of the relevant conditions that contribute to the sentence's meaning what it means, are the *beliefs* entertained by the speaker on occasion of uttering the sentence. Davidson's holistic view of meaning is extended to beliefs as well, since speaker's beliefs and sentence meanings are interdependent: "a belief is identified by its location in a pattern of beliefs; it is this pattern that determines the subject matter of the belief, what the belief is about."¹⁴³

T-sentences are to be taken as the consequences of *a theory that conforms to the Principle of Charity*, where said theory is a part of (but not all of) what is involved in understanding. The T-sentences are only the machinery of the theory, they cannot be expected to work by themselves. The machinery needs to be "activated" by rational agents, persons participating in the communicative act. The interpreter using the T-

¹⁴³ *Inquiries* 168.

sentence to interpret the utterance of a speaker, must make the "right sort" of assumption with respect to what conditions the speaker holds relevant for uttering the particular sentence on that occasion. In adherence to Davidson's thesis of the *indeterminacy of meaning*, there may be several right sorts of assumptions. Such assumptions, however, can be successfully discriminated by virtue of being subject to shared principles of rationality. According to Davidson, *rationality, shared by speakers must be assumed by the interpreter, because it is only relative to the assumedly shared standards of a universal human ability of reasoning, that speakers can attribute interpretations to one another's utterances*. Without considering the normative constraints subsumed by the Principle of Charity, the T-sentences are indeed useless. On the other hand, Davidson's claim that the assumption of shared rationality among all human communicators is *the* uniquely relevant principle that allows the possibility of interpretation, seems very plausible.

CONCLUSION

Davidson's recent conclusion rejecting convention and language as playing no significant role in interpreting speakers' utterances has been shown to be consistent with his earlier held view concerning the interdependency of meaning and beliefs. According to this view, no meaning can be attributed to speakers' utterances, unless the speakers' relevant beliefs are considered simultaneously with the interpretation of the sentences uttered by them. Linguistic convention understood as a set of beliefs shared by members of a linguistic community cannot play a significant role in interpretation, since the speaker's relevant beliefs are not known to the interpreter prior to having interpreted the speaker's utterance on a given occasion. Similarly, the manifestation of linguistic conventions, namely, the use of a particular language cannot be significantly contributing to interpretive success, since the interpreter's mastery of a given language alone would not explain how interpretation succeeds. Instead, what has been found crucial for interpretive competence is the employment of the Principle of Charity: the assumption on part of the interpreter that the speaker holds true the sentence uttered. The interpreter must assume the speaker's holding true the sentence, based on a more general assumption concerning their shared rationality. Only if these assumptions are made can meaning be attributed to the speaker's utterances.

The question arises just what role the compositional truth theory can be playing in a theory of interpretation, given Davidson's recent conclusions. Did Davidson reject the importance of a truth-theory as well, together with that of convention and language? The answer to that question must be negative. A compositional truth-theory was needed by Davidson in order to account for the fact that NL is learnable. Davidson never denied that there are languages that can be mastered by competent speakers, due to their compositional character. The interpreter will only be able to generate T-sentences if she knows (via recursive rules) what the component words mean and what the rules are based on which the sentences are composed. This knowledge of the interpreter is a significant part of what is involved in interpretation. This is also the kind of knowledge that can be acquired, is shared by all competent speakers and is formalizable by a finite theory. However, this knowledge alone cannot account for speakers' ability to interpret one another's utterances. In order to account for that ability, the other, crucial part of what must be involved in interpretation, namely, the Principle of Charity is invoked. The employment of the Principle of Charity is the part that cannot be acquired in a sense of speakers having knowledge of it prior to the actual acts of interpretation. The statement that all speakers must employ this principle is not analogous to saying that speakers share it, since its employment is geared to particular occasions. It is therefore unlikely that the ways speakers employ the Principle of Charity could be stated in a finite form of a theory. The decisive principle underlying interpreters' competence, namely, the assumption of shared rationality among speakers remains an assumption. It is however an extremely plausible assumption, since the negation of it is simply incomprehensible.

BIBLIOGRAPHY

- Davidson, Donald (1986) Inquiries into Truth and Understanding. Oxford: Clarendon Press.
- Davidson, Donald (1986) "A Nice Derangement of Epitaphs", in E. LePore, ed. Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson, pp. 433-446. Oxford: Basil Blackwell, 1986.
- Dummett, Michael (1976) "What is a Theory of Meaning? (II)", in G. Evans and J. McDowell eds. Truth and Meaning: Essays in Semantics, pp. 67-137. Oxford: Clarendon Press, 1976.
- Foster, John A. (1976) "Meaning and Truth Theory", in G. Evans and J. McDowell, eds. Truth and Meaning: Essays in Semantics, pp. 1-32. Oxford: Clarendon Press, 1976.
- Frege, Gottlob (1892) "On Sense and Nominatum", reprinted in A.P. Martinich, ed. The Philosophy of Language, second edition, pp. 190-202. New York, Oxford: Oxford University Press, 1990.
- Grice, H.P. (1975) "Logic and Conversation", reprinted in A.P. Martinich, ed. The Philosophy of Language, second edition, pp. 149-160. New York, Oxford: Oxford University Press, 1990.
- Grice, H.P. (1957) "Meaning", reprinted in A.P. Martinich, ed. The Philosophy of Language, second edition, pp. 72-78. New York, Oxford: Oxford University Press, 1990.
- Hacking, Ian (1986) "The Parody of Conversation", in E. LePore, ed. Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson, pp. 447-458.

Oxford: Basil Blackwell, 1986.

Lewis, David (1975) "Languages and Language", in K. Gunderson, ed. Language, Mind and Knowledge, pp. 1-27. Minneapolis: University of Minnesota Press.

Quine, Willard Van Orman (1980) "Reply to Davidson", in D. Davidson and J. Hintikka, eds. Words and Objections: Essays on the Work of W.V. Quine, pp. 333-335. Dordrecht: D. Reidel Publishing Company.

Quine, Willard Van Orman (1951) "Two Dogmas of Empiricism", reprinted in A.P. Martinich, ed. The Philosophy of Language, second edition, pp. 26-39. New York, Oxford: Oxford University Press, 1990.

Quine, Willard Van Orman (1960) Word and Object. Cambridge, Mass.: M.I.T.

Tarski, Alfred (1933) "The Concept of Truth in Formalized Languages", reprinted in A. Tarski, Logic, Semantics, Metamathematics, second edition, pp. 152-278. Indianapolis, Indiana: Hackett Publishing Company, Inc. 1983.

Tarski, Alfred (1944) "The Semantic Conception of Truth and the Foundations of Semantics", reprinted in A.P. Martinich, ed. The Philosophy of Language, second edition, pp. 48-71. New York, Oxford: Oxford University Press, 1990.

Wallace, John (1977) "Only in The Context of a Sentence Do Words Have Any Meaning", in P.A. French, T.E. Uehling Jr and H.K. Wettstein (eds) Midwest Studies in Philosophy, 2: Studies in the Philosophy of Language, pp. 144-164. Morris: University of Minnesota Press.

Wilson, N.L. (1959) "Substance Without Substrata", *Review of Metaphysics* 12 pp. 521-539.