Meaning in Donald Davidson's truth-conditional semantics

Martin Andreas Lind

Rebstein (SG) und Wien (Österreich)

Table of contents

1 Introduction

Human beings use their linguistic skills in an impressive variety of ways: they state and assert what they believe, they ask questions, and they issue commands; they communicate their wishes, hopes, and fears, and they use their languages in order to lie and to deceive. They refer linguistically to the concrete and to the tangible as well as to the abstract and to what is too small, too slow, too fast or too far away for sensory perception. Human beings use language to be funny, sarcastic, and ironic; they play with language and they use language to express themselves artistically.

This list of different kinds of language use is obviously incomplete. It suffices, however, to remind us of the fact that our linguistic abilities are exceptional, perhaps unique. This considered, it is no surprise that languages and our ways of using them have always enjoyed the philosopher's attention. The business of philosophers is, of course, the *general* and *fundamental*. Accordingly, they are not interested in specific *empirical* questions about the acquisition or the use of a particular language; their interest is rather with the question why it is possible, *in principle*, to acquire and to use a language.

Since the use of a language typically involves the production of certain sounds or written marks, there must be something, the philosophers say, that distinguishes linguistic sounds (marks) from non-linguistic ones (such as, for instance, the squeaking of a train's wheels in sharp bends, or the colourful patterns of inclusions in white blocks of marble). The relevant distinctive characteristic is, they hold, that linguistic sounds (marks) possess specific *linguistic meanings*, while other sounds (marks) do not. Accordingly, philosophers typically try to answer questions such as the following:

- What exactly is linguistic meaning?
- How could we describe a sentence's (a phrase's, a word's) particular linguistic meaning?
- How do we have to define the concept of linguistic meaning in order to describe the possibility of language-understanding and/or the possibility of language-acquisition?

A particularly interesting approach to these questions has been proposed by the American philosopher Donald Davidson (1917–2003). Davidson claims that it is possible to deliver an adequate conceptual description of the possibility of language-understanding (and, hence, of language-acquisition) if we demand nothing more of the concept of linguistic meaning than that it indicates what a speaker holds to be true by what he utters. The foundational notion of linguistic meaning (which Davidson refers to as *"literal meaning"*) has thus to be equated, he says, with the truth-conditions the speaker allocates to his language's sentences.

This proposal – it is generally referred to as the *truth-conditional approach to semantics* – is among the most influential contributions to the philosophy of language in the second half of the 20th

century. While some authors regard it as the most promising formalist approach to the present day, have others raised severe doubts about its adequacy. Their objections might be divided, roughly speaking, into two (non-exclusive) groups. There are, firstly, objections which call into question whether the concept of truth (and/or Davidson's understanding of it) provides an adequate basis for linguistic meaning *in general*; and secondly, there are objections which doubt, in one way or another, that Davidson's truth-conditional approach might be applied to *natural* languages (i.e. to the languages you and I learn, speak and understand). The former kind of objection is typically centred around the claim that it is the concept of truth which depends on the concept of linguistic meaning rather than the other way round. The objections of the latter kind, on the other hand, usually claim that it is the speaker's communicative intentions that give life to natural languages, and that these intentions therefore have to play some role for the foundational notion of linguistic meaning. The two kinds of objection are often voiced in combination, and both are generally directed against Davidson's attempt to apply his truth-conditional approach to natural languages.

It is exactly the application of the truth-conditional approach to natural languages that stands at the centre of the present study. More precisely, the study is concerned with this approach's foundational notion of linguistic meaning (i.e. the notion of *literal meaning*). Since literal meaning is, as I already mentioned, restricted to the truth-conditions of sentences, it may appear as a very uninteresting notion of meaning at first sight. This impression is, however, wrong: the notion enjoyed a noteworthy evolution during the development of Davidson's philosophy of language, acquired a number of interesting characteristics, lost certain others, and even got replaced at one point by some equally truth-conditional, but different notion of meaning (i.e. by the notion of *first meaning*[1]*)*.

However, is it really possible, as Davidson says, to arrive at an adequate conceptual description of the possibility of language-understanding if one relies on a notion of linguistic meaning that is restricted to the truth-conditions of sentences? The present study aims to clarify this question by focussing on the mentioned notions of meaning (i.e. *literal* and *first meaning*). It reconstructs their development and modifications and considers what their changing characteristics contribute to the picture in order to convince us that truth-conditional semantics might be successfully applied to our languages.

The study is divided into two parts. *Part I* (comprising the chapters 2–4) clarifies what Davidson's truth-conditional approach actually consists in, and aims to explain why it is a promising approach if applied to the semantics of natural languages. It begins in chapter 2 with a short summary of Davidson's foundational ideas. Chapter 3 subsequently defends these ideas against the communication-intentionalists' challenge. Chapter 4, finally, explains Davidson's understanding of the concept on which his approach to semantics is ultimately based: the concept of truth. *Part II* (comprising the chapters 5 and 6) explores the evolution of the notion of literal meaning (and first

---

[1] In order to avoid confusion, I ignore this replacement until chapter 6, where it is discussed in detail. Until then, I refer to Davidson's foundational notion of meaning exclusively by the term "literal meaning".

meaning). Chapter 5 contrasts the said notion with the various forms of conventional (or rule-governed) use of language, and chapter 6 does the same with regard to the various forms of non-conventional or otherwise idiosyncratic language use. Chapter 7 completes the study with a few evaluative remarks.

Before setting out, I would like to thank a number of people and institutions who helped me in one or another way in the process of writing this study:

Part I:

Defending Davidson's philosophy of language

2 Davidson's philosophy of language – a short introduction

Davidson's philosophy of language might be divided into two parts: on the one side, his proposal of a truth-conditional approach to semantics, and, on the other, the application of this approach to natural languages (i.e. the theory of Radical Interpretation). The present chapter outlines both of them, and in this order.[2]

2.1 Davidson's truth-conditional approach to semantics

Human beings are mortal beings. The amount of time available for us to acquire a language is thus necessarily *finite*. Our languages, however, typically comprise an *infinitely* large set of sentences. The logical consequence of this is that most sentences of our languages are unknown to us. There are, thus, at every moment of our life, an infinite set of sentences which we have never heard or uttered, and which accordingly we have never learned. Such sentences are easy to find. Provided that you have not read John Updike's *Toward the End of Times*, the following will be an example:

> *The sea earlier this March morning wore a look you never see in winter – a lakelike calm, a powdery blue so pale it was scarcely blue, with stripes of a darker, stirred-up color that might have marked the passage of a lobster boat an hour before* (Updike 1997, p. 89f).

The surprising fact is that you *understand* this sentence at once. But how is this possible if you have never encountered it before? Updike's sentence is no extraordinary case; it rather illustrates a ubiquitous phenomenon of language use. So how to explain your understanding?

The answer to this question cannot be found in *us*, the *language users*, for we cannot doubt or deny our own mortality: we simply *are* finite beings and can acquire finite amounts of knowledge only. The solution thus has to be searched on the part of language itself, or more precisely, in the *structures* of language. This is exactly the starting-point for Davidson's reflections on the semantics of natural languages (Davidson 1965); and that this is so is, of course, no coincidence: Davidson begins with our languages' structures rather than with the conceptual origins, or the particular nature, of the concept of linguistic meaning, because he conceives of himself as a *formal semanticist*. He thus believes that our construal of the concept of linguistic meaning has to conform to our languages' (putative) structures rather than the other way round. To proceed along these lines is important, he claims, because it is only awareness of our languages' structures that allows us to construct our

---

[2] Readers who are unfamiliar with Davidson's writings are advised to consult a more comprehensive introduction. I recommend the chapter about truth-conditional semantics in William G. Lycan's *Philosophy of Language. A contemporary introduction* (very short), Marc Joseph's *Donald Davidson* (an-easy-to-read monography), and Ernest Lepore/Kirk Ludwig's *Donald Davidson's Truth-Theoretic Semantics* or *Donald Davidson. Meaning, Truth, Language and Reality* (very detailed and quite difficult).

theories of meaning in such a way that they conform to those structures, and only if this conformity is guaranteed is it possible to claim that the said theories are really theories of meaning *for natural languages*. Observing this 'structure-first'-maxim thus prevents us from missing our target; it guarantees, according to the formal semanticist, that we arrive at where we hope to arrive. (The formal semanticists' approach draws fierce opposition from the so-called *communication-intentionalists*. The basics of their approach will be outlined in section 3.1.)

Let us return to our initial question: what makes it possible for us to understand far more sentences of our languages than those which we have actively learned? Davidson's answer is that it is due to our languages' *compositional* structure (Davidson 1965, p. 8). What he means by this is that the infinitely many sentences of our languages are *composed* of a finite number of items, and thus of a number of items that is manageable for finite beings like us.[3] The compositional structure of natural languages is supposed to look roughly as follows: among the finite number of items that constitute sentences are, on the one hand, the languages' semantically contentful items and, on the other hand, the items which explain how the semantically contentful ones are allowed to be combined with each other. The items of the latter kind (roughly) match with our languages' grammatical rules, while the semantically contentful items constitute their vocabularies. The semantically contentful items divide, furthermore, into two subcategories: those which are themselves composed from semantically contentful items, and those which are not. Among the items of the former subcategory are words such as "hairdryer", "bullet-proof" or "bullfight"; among the latter are "hair", "dryer", "bullet", "proof", "bull" and "fight". Davidson calls the latter ones the languages' *semantical primitives* (cf. Davidson 1965, p. 9). What we need to learn in order to understand a sentence – *any* sentence – of a particular language is, thus, the complete (but finite) set of the semantical primitives of that language and the complete (but finite) set of its grammatical rules. Understanding a sentence is then 'merely' a matter of *calculating* the composed meaning of the semantical primitives' respective linguistic meanings and the particular mode of their arrangement within the sentence under consideration. The compositional structure thus explains why the linguistic capacities of human beings happen to comprise so much more than what might be actively learned during our relatively short lives.

I already mentioned that it is our languages' structure – i.e. their compositional structure – upon which Davidson intends to build his approach to semantics. The problem with this idea is, however, that there exists a phenomenon that seems to foreclose such a strategy. This inconvenient phenomenon is commonly referred to as the *Principle of Contextuality*[4], which holds that a word – and thus a Davidsonian semantical primitive – lacks a clearly determined meaning when it is considered *in isolation*. The phenomenon is easily demonstrated by an everyday example. Consider, for instance, the different meanings of the word "bank" in the following sentences:

---

[3] The idea of the compositional structure of natural languages is, of course, not Davidson's own discovery, but rather one of the few consensual hypotheses among philosophers of language. The only philosopher who has ever doubted the compositional structure of natural language is, I believe, Stephen Schiffer (cf. Schiffer 1987).

[4] The Principle of Contextuality was first articulated by the German logician Gottlob Frege in the late 19th century (cf. Frege 1891 and Frege 1892).

(i)   The bank lost a lot of money in the recent economic crisis.

(ii)   The cities' harbour is at the left bank.

It is obvious that "bank" takes on completely different meanings in these sentences. The word's particular meaning is thus not determined by the word itself, but requires a sentential context in order to be so. The phenomenon referred to by the Principle of Contextuality thus seems to contravene what the compositional structure of natural languages requires: it says that a word (or a sentential primitive) possesses a determined meaning only within a sentential context, while the compositional structure of our languages require us to conceive of a sentence's meaning as the result of the determinate meanings of its constituent parts (= words, sentential primitives) and the modes of their combination. Do compositionality and contextuality thus mutually exclude each other?

Davidson's answer is that they do not: compositionality and contextuality are perfectly compatible, he says, if we take a sentence's linguistic meaning[5] to be fully determined by that sentence's particular truth-conditions. His course of reasoning goes as follows: the equation[6] of meanings with truth-conditions allows, on the one hand, to do justice to the compositional structure of our languages, for the constitution of a particular sentence's truth-conditions runs exactly parallel to the constitution of a particular sentence's meaning: a complex sentence's *meaning* is constituted by the respective meanings of its constitutive sentences plus their particular mode of combination, while the complex sentence's *truth-conditions* are constituted by the particular truth-conditions of its constitutive sentences plus their particular mode of combination. The same holds, of course, for the constitutive sentences themselves: their respective meanings – as well as their respective truth-conditions – are constituted by the particular meanings (truth-conditions) of even less complex sentences plus their respective modes of combination. The equation of meanings and truth-conditions is, on the other hand, compatible with the Principle of Contextuality, since the equation does *not* assume an isolated word (or an isolated semantical primitive) to have any determinate meaning. It does not – and cannot – since the truth-predicate is applicable to sentences only; words and semantical primitives, however, are subsentential items. It thus follows that they cannot have any truth-conditions if taken in isolation. The only statement a truth-conditional approach to linguistic meaning actually makes with respect to the subsentential items is that they have a particular linguistic meaning *if* they form a constitutive part of a sentence (Davidson 1967, p. 20ff).

---

[5] What is meant here by "linguistic meaning" is the sentence's *conceptually foundational notion of meaning*, i.e. its *literal meaning*. Davidson does not deny that a sentence may have additional (but non-foundational) other notions of meaning (say, a conventional meaning) which may go far beyond the respective sentence's truth-conditions. We shall return to this issue at length in the chapters 5 and 6.

[6] Notice that I am speaking here of the *equation* of meaning with truth-conditions, and not of the *identity* of the two. The difference is important, for the latter, but not the former, suggests a *reduction* of linguistic meaning to truth-conditions. Davidson, however, is not pursuing a reductionist project. I will come back to this issue later in the present section (cf. p. 16).

Davidson's idea to equate meanings with truth-conditions is doubtless ingenious. It provides, however, a suitable basis for a theory of meaning only if a number of tricky conditions turn out to be fulfilled. Among these conditions are (1) that Davidson is correct in claiming that linguistic meaning is completely determined by the sentences' truth-conditions, (2) that every sentence of a language possesses (or could be analyzed as possessing) a particular set of truth-conditions, and (3) that a suitable conception of truth turns out to be available.

Condition (1) obviously names the core idea of Davidson's truth-conditional approach to semantics; namely that

> *[t]o know the semantic concept of truth for a language is to know what it is for a sentence – any sentence [of a specific language] – to be true, and this amounts, in one good sense we can give to the phrase, to understanding the language* (Davidson 1967, p. 24).

I do not want to defend this assumption here, since this is an integral part of my investigations in the subsequent chapters of this study. The Davidsonian assumption has, however, *prima facie* plausibility since we can judge a sentence's truth or falsity only under the condition that we know what that sentence happens to mean. So why should it not be possible, in principle, to go the same path the other way around; i.e. to capture a sentence's linguistic meaning in terms of our knowledge of that sentence's truth-conditions?

Condition (2) is that every sentence of a language has to be such that it possesses (or might be analyzed as possessing) a particular set of truth-conditions. Its fulfilment is indispensable to Davidson's project, for it is the possession of truth-conditions that determines the limits of his truth-conditional approach. Whatever is beyond these limits is thus beyond the reach of the truth-conditional approach. The existence of sentences without truth-conditions would thus prove the inadequacy of that approach. Speculations about the existence of such sentences are not baseless: questions and commands, for instance, do seem to fall into this category. Davidson is well aware of this problem and tries to solve it by means of his *paratactic analysis* (cf. Davidson 1979). We will return to this issue in section 5.3.

Condition (3) asks for a suitable conception of truth. Such a conception is a conception that allows the construal of truth-theories which mirror our respective languages' structures. Since we cannot claim to know the structures of the natural languages in all their subtle details, it is impossible to *prove* that a particular conception of truth is the truly suitable one. We can, however, enumerate the conditions which a conception of truth has to fulfil in order to count as a *potentially* suitable one. This is the case if the conception is such that it (a) allows for the construal of individual truth-theories for each particular language (and, thus, for an individual truth-predicate for each particular language) and (b) if these truth-predicates are defined in a way compatible to the natural languages' compositional structures. The reasons for the importance of (b) should be clear by now (cf. the explanations regarding compositionality above). (a), on the other hand, is important since it ensures that the truth-

conditional approach does justice to what might be called our languages' *semantic individuality*; i.e. to the fact that one and the same pattern of sound-waves (or written marks) may have different linguistic meanings in different languages. This fact may be mirrored by a conception of truth only if it allows for the definition of an individual truth-predicate for each language individually, for if it defined only one (or a few) such predicate(s) for all languages together, it would follow that the semantically foundational notion of meaning (= literal meaning) of a particular pattern of sound-waves (or written marks) were fixed across all (or several) languages. Such a situation would be incompatible with the readily observable semantic individuality of the natural languages.

Davidson claims that a suitably modified version of Alfred Tarski's semantic conception of truth (Tarski 1944) is capable to fulfil both these conditions. It fulfils condition (a), he says, since it is a *semantic* conception of truth. Tarski thus defines the truth-predicates in terms of *linguistic meaning*. Linguistic meaning is, however, always linguistic meaning *of a particular language*. The Tarskian truth-predicates are therefore necessarily truth-predicates for *individual* languages (i.e. they take the form "true in L" where L is a specific language). Condition (b), on the other hand, is fulfilled, according to Davidson, because of the particular technique which Tarski applies in defining the languages for which he then defines his truth-predicates (more about this below).

Let me briefly outline the basics of Tarski's semantic conception of truth.[7] Tarski's aim is, as I already said, to define truth-predicates for individual languages, and to do so in terms of those languages' linguistic meanings. He believes that an adequate truth-theory for a particular language L has to be such that it fulfils what he calls the *material adequacy condition* (Tarski 1944, p. 15). This condition is fulfilled if the truth-theory for L happens to be such that it indicates, for each individual sentence of L, the conditions under which these sentences are true. This is the case, for instance, if we want to define the truth-predicate for German, and if our truth-theory (which is formulated in English) entails for each individual sentence of the infinitely many German sentences the suitable sentences of the following form:

"Schnee ist weiss" is true if and only if snow is white.
"Peter liebt Anna" is true if and only if Peter loves Anna.
"Shakespeare war ein grosser Künstler" is true if and only if Shakespeare was a great artist.

Sentences of this form are called *T-Sentences*. Their general form is called *Convention T* (Tarski 1944, p. 16) and looks as follows:

S is true if and only if p.

---

[7] The following summary is, of course, a simplified and abbreviated version. For an excellent presentation cf. Kirkham 1997, pp. 146–158.

"S" stands as the variable for any sentence of L (here: German), and "p" as variable for the sentences of the language within which the truth-predicate for L is defined (i.e. the metalanguage ML; here: English). There exist as many T-Sentences for an object language L as there are sentences in L. Tarski's material adequacy condition for a truth-theory for L is thus fulfilled if and only if that theory happens to entail *all and only those* T-Sentences which are such that they correctly indicate, for each individual sentence of L, the conditions under which these sentences are true. Each T-Sentence fulfilling this condition makes, according to Tarski's idea, its tiny contribution to the definition of L's truth-predicate: each T-Sentence is, as he puts it, a *"partial definition of truth"* for L, where the complete definition of L's truth-predicate is *"the logical conjunction of all these partial definitions"* (Tarski 1944, p. 16).

Tarski has to solve two major difficulties in order to show that it is possible to define truth-predicates along these lines. One of these difficulties concerns the languages for which he intends to define his truth-predicates, the other concerns the formulation of the truth-predicates themselves. The first difficulty arises from the fact that Tarski does not aim at defining truth-predicates for natural languages, but only for a restricted set of artificial languages. The reason behind this restriction is that natural languages contain self-referring predicates (i.e. "is true of", "refers to", etc.), and that Tarski is afraid that the paradoxes that might arise from such predicates could undermine his entire project (Tarski 1944, p. 18f). He thus starts by defining suitable languages. This, however, is a challenge of its own: a definition of a language needs to define the set of sentences which belong to it; but languages entail, as we know, an infinite set of sentences. The problem Tarski was thus confronted with is that he first had to discover a way of defining the infinitely large set of a language's sentences without mentioning each sentence individually (for otherwise, the definition would have to be infinitely long, which is impossible). Tarski's solution was to make use of a *recursive definition*-technique (Tarski 1944, p. 18f). This is a technique that allows to define infinite sets of items (for instance: sentences) by finite means. The technique achieves this by defining the finite sets of the said items' constitutive elements and the finite set of those element's admissible combinations. Tarski thus modelled his languages as having a *compositional structure*.

The second difficulty concerns, as already said, the formulation of the truth-predicates themselves. The problem that posed itself to Tarski was (roughly) that he had to find an ersatz-predicate for the "is true" that appears in his T-Sentences. His solution to this difficulty is the introduction of the *satisfaction*-predicate.[8] A sentence is true if it is *"satisfied by all objects"*, Tarski says, while it is false if it fails to be satisfied by at least one object (Tarski 1944, p. 25). What he means by "is satisfied *by all objects*" is that it is the *facts of the matter*, i.e. the entities referred to by the sentence, that determine whether that sentence happens to be satisfied or not (and whether the corresponding T-Sentence is true or not).

---

[8] The definition of the *satisfaction*-predicate is technically demanding and requires profound knowledge in predicate-logic and set theory. Its presentation would lead quite far away from the purpose of this chapter. This is why I allow myself to skip it here.

Here is an example: The open sentence "x ist weiss" is satisfied if x is replaced by a word that refers to a white object. This is the case, for instance, with the following sequences of objects (they are just a few among infinitely many other suitable sequences):

{snow, watermelon}, {milk, Shakespeare, fork}, {tooth paste, flag, milk, sundae, teddy bear}

It is not satisfied, however, by all sequences of objects that do not mention white things at their first place. For instance:

{watermelon, Shakespeare, fork, snow}, {coffee}, {coffee, milk, snow}

Accordingly, the T-Sentence:

"x ist weiss" is true if and only if ….. is white

is true if the word that replaces "…" on the right side of the biconditional (i.e. of the mentioned T-Sentence) refers to the same object as does x in "x ist weiss" on the left side of that biconditional.

The example gives only a heavily simplified and incomplete idea about the exact working of the property of satisfaction. It suffices, however, for our present needs, for it reveals the important connection between *truth and meaning* in Tarski's conception of truth: a sentence happens to be true, as we have just seen, if it is satisfied, and it is satisfied if the sequences of objects are of the right kind. The question whether a particular sequence of objects happens to be of the right kind depends, however, on the *meanings* of that sentence's predicates, for it is these predicates' meanings that determine what kind of objects are referred to by that particular sentence. It is because of this intimate connection between truth and meaning that Davidson takes the Tarskian conception of truth to be suitable for his truth-conditional approach to semantics.

A direct transfer of Tarski's work to Davidson's semantics is, however, impossible. One reason for this is that Tarski developed his conception of truth for *artificial* languages, while Davidson needs such a conception for *natural* languages. Another difficulty arises from the fact that Tarski defines truth *in terms of meaning*, while Davidson wants to describe meaning *in terms of truth*. Tarski thus starts at where Davidson wants to arrive. Davidson is well aware of both these difficulties. He is, however, very optimistic that suitable adaptations will make such a transmission possible. With respect to the problem that Tarski defines truth-predicates for artificial languages, Davidson says the following:

Let us look at the positive side. Tarski has shown the way to giving (…) [truth-] theor[ies] for interpreted formal languages (…); pick one as much like English as possible. Since this new

language has been explained in English and contains much English we not only may, but I think must, view it as part of English for those who understand it. For this fragment of English we have, ex hypothesi, *a [truth-] theory of the required sort. Not only that, but in interpreting this adjunct of English in old English [i.e. the natural language referred to as "English"] we necessarily gave hints connecting old and new. Wherever there are sentences of old English with the same truth conditions as sentences in the adjunct we may extend the theory to cover them* (Davidson 1967, p. 29).

The other difficulty – i.e. that Tarski starts at where Davidson wants to arrive – is, of course, no accidental difficulty. On the contrary: Davidson believes that the concept of truth is *fundamental* for us and thus understood by all of us. It is, he says, the understanding of the concept of truth that unites all rational beings – that makes beings *rational* beings; and it is the sharing of a particular truth-theory that unites all speakers of some particular language.[9] – It is *these* convictions that had first brought Davidson to the idea that we should try to describe a speaker's linguistic behaviour in terms of the concept of truth. It is, in other words, these convictions that stand at the beginning of his truth-conditional approach to the semantics of natural languages.

Using Tarski's work on truth as the basis for a truth-conditional theory of meaning requires us to read his work *backwards*; i.e. it requires us to read the Tarskian truth-definitions as *tools* that allow the deciphering of the meanings of sentences. Doing so gives a new function to Tarski's T-Sentences: they no longer indicate the truth-conditions of a particular sentence of the object language, but serve as a tool to determine that sentence's meaning. This is why Davidson's truth-conditional approach to semantics takes the form of a *theory of interpretation*: it is a matter of *interpretation* whether such-and-such a T-Sentence happens to express a truth for such-and-such an object language. Such an interpretation is, of course, trivial in the case of one's own language, for in this case, we already know the meaning of the object language's sentences that appear in the respective T-Sentences. The interpretation is, however, extremely demanding if we consider T-Sentences that concern languages unknown to us. Consider, for instance, a T-Sentence such as:

"Gavagai!" is true if and only if ….

What particular sentence of *our* language could make this T-Sentence true? – We come back to this question in the subsequent section. But let us focus first on Davidson's reverse reading of Tarski's truth-definitions. In order to fulfil their function, there are two additional modifications required: namely, the relativization of the T-Sentences to *speakers* and to particular *points of time*. The reason for this is that the set of true T-Sentences that 'gives' the meaning of the sentences uttered by speaker A is not necessarily the same as the set that 'gives' the meaning of the sentences uttered by speaker B.

---

[9] I will return to these claims (and Davidson's argument for them) in chapter 4.

The T-Sentences have to be relativized to points of time, on the other hand, because most sentences of a language are not always true, but only if certain conditions obtain (consider, for instance, a sentence such as "It is raining".) Davidson's T-Sentences thus take the following form (Davidson 1967, p. 34):

"Schnee ist weiss" is true as (potentially) spoken by p at t if and only if snow is white at t.

"Gras ist grün" is true as (potentially) spoken by p at t if and only if grass is green at t.

etc.

A Davidsonian interpreter is thus supposed to consider such T-Sentences and, by doing so, to discover the meanings of the sentences of the object-language that constitute the T-Sentences' left sides (i.e. the respective biconditional's antecedens). Now, can his doing so lead to a theory of meaning of the object-language? Or, to put the same question in another way: is the interpreter's search for a Tarski-style truth-theory[10] suitable to reveal the meanings of the sentences of that speaker's language? Davidson is very optimistic, and his reasons for this will be explained in the next section. Before turning to this issue, however, let me briefly mention three important characteristics of his truth-conditional approach to semantics as explicated so far.

The first such characteristic is that the approach does not make use of any *meaning entities*: They are completely replaced in Davidson's approach by the truth-conditions of sentences (Davidson 1967, p. 24). This is not to say, however, that Davidson's approach to semantics is a *reductionist* approach, for Davidson does not *reduce* meanings to truth-conditions. He only claims that a Tarski-style truth-theory for some particular language L works *as a vehicle*[11] that reveals the meanings of L's sentences. This particular characteristic of Davidson's approach is doubtless one of its most important strengths, for it is more than dubious whether it were possible to reify the said meaning entities or whether they could have any use at all in a theory of meaning.[12]

The second characteristic is that Davidsonian interpretation-theories are *externalist* theories in that they describe a language's semantics exclusively in terms of the extensional concept of truth. Whether this characteristic counts as a praiseworthy aspect of Davidson's approach or rather renders it a complete failure is, of course, a matter of one's perspective (cf. section 3.1). It clearly is a praiseworthy aspect from Davidson's own point of view, since his *holist* convictions do not allow him to build his approach on any intensional concept (like, for instance, on communicative intentions). A holist claims (roughly) that the particular contents of our propositional attitudes (beliefs, intentions, desires, etc.) are determined by the contents of the other propositional attitudes which we happen to have. According to the holist, it is thus impossible to sharply separate the content of one propositional attitude (say, a belief) from the content of another. This conviction (which is a result of Quine's

---

[10] I am speaking here about *Tarksi-style truth-theories* instead of *Tarski truth-theories* because of the above mentioned adaptations. In doing so, am I following Davidson's own usage.

[11] The wording is Lepore/Ludwig's. Cf. Lepore/Ludwig 2007, p. 27.

[12] Quine famously demonstrates the problems of reifying meanings (cf. Quine 1951 and Quine 1960, chapter 2), while Davidson argues for their complete uselessness (Davidson 1967, p. 21).

abandonment of the analytic/synthetic dichotomy (Quine 1951)) is not without consequences for the ways the understanding of a language might be explained or described: if it is impossible to isolate the contents of different beliefs (or of any other propositional attitudes), it is impossible to sharply distinguish between what a speaker, while uttering such-and-such a sentence, is believing his sentence to mean, and what he is believing to communicate with his uttering of that sentence. The speaker's communicative intentions and what he believes his uttered sentences to mean are thus, as Davidson says, *"parts of a single project, no part of which can be assumed to be complete before the rest is"* (Davidson 1973, p. 127). We cannot, therefore, first determine the speaker's communicative intention and then what he believes his uttered sentence to mean (or the other way around), but have to capture both of them in one single step. It is because of this that Davidson takes communicative intentions and beliefs about meanings to be too close to each other as to allow for an explication of one of them in terms of the other. The externalist concept of truth, on the other hand, is conceptually further away from beliefs as well as from meanings; it is, as Davidson says, *"equidistant from [both]"* (Davidson 2005b, p. 57), and thus a promising basis towards a better understanding of both of them.

The third and last characteristic I want to mention is that Davidson's approach to semantics does not allow for the determination of the meanings of *isolated* sentences. The reason for this is that T-Sentences, if considered in isolation, may look as if they belong to the required Tarski-style truth-theory for a language (say, German) although they do not belong to that theory. The T-Sentence

(i)   "Schnee ist weiss" is true as (potentially) spoken by p at t if and only if snow is white at t.

for instance, clearly belongs to that theory, while the T-Sentence

(ii)   "Schnee ist weiss" is true as (potentially) spoken by p at t if and only if grass is green at t.

obviously does not. The difference between these two T-Sentences is, however, undetectable if we regard each of them in isolation: if one of them seems to mirror (correctly or not) what is held true by a speaker, the other seems to do so too. The reason for this is that the truth-conditions of those T-Sentences remain always exactly the same. This problem might be easily removed, however, if a T-Sentence is considered in combination with many other T-Sentences. If this is ensured, it is easy to detect that

(iii)   "Gras ist grün" is true as (potentially) spoken by p at t if and only if grass is green at t.

appears as just as correct as (ii) seems to be. The obvious contradiction between (ii) and (iii) thus encourages further investigations on the interpreter's side and eventually the exclusion of the

misleading T-Sentence(s) (Davidson 1967, p. 25f). This procedure becomes even more efficient as soon as the structures of the object language's sentences happen to be known, for the knowledge of these structures allow the interpreter to recognize that those sentences' constitutive elements (words, phrases) resurface in other sentences, and that they are thus bearers of a particular meaning. It allows, in other words, to recognize that "weiss" leads to a true sentence if combined with "Schnee", but clearly not so if combined with "Gras", and that "grün" leads to a true sentence if combined with "Gras" but not if combined with "Schnee", etc.

The impossibility to determine the meanings of isolated sentences is a fortunate consequence from Davidson's point of view, for his holistic convictions are completely incompatible with the idea of isolated (sentence-) meanings. Frege recognized, Davidson comments,

> (…) that only in the context of a sentence does a word have meaning [= Principle of Contextuality, cf. p. 9f]; 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 (Davidson 1967, p. 22).

2.2 The theory of Radical Interpretation

The truth-conditional approach, as explained so far, is a purely formal model. It provides no more (but likewise no less) than a promising and cautiously outlined framework for a potentially adequate description of the semantics of natural languages. What is lacking, however, is *evidence* that this framework, if applied to natural languages, really works. This is what the *theory of Radical Interpretation* investigates. The theory of Radical Interpretation might thus be understood as a kind of *empirical test* of Davidson's model of a truth-conditional approach to semantics. It is, however, an empirical test of a quite peculiar kind, for it involves no empirical studies in the field, but rather a range of armchair considerations about the question of whether or not it is possible, *in principle*, to interpret and to understand a speaker's uttered sentences, phrases, and words, by means of searching (and finding) the particular Tarski-style truth-theory that allocates the truth-conditions of a speaker's uttered sentences in exactly the same way as the speaker is allocating them by himself.

A first obstacle to finding an answer to this question is that it asks us to imagine a situation of language interpretation that happens to be far removed from our usual daily practice. Davidson meets this difficulty by introducing a suitable thought-experiment. This is the thought-experiment of a field linguist (i.e. a *radical interpreter*) who tries to interpret the linguistic behaviour of the members of a hitherto completely unknown tribe. Lacking any specific knowledge whatsoever about the alien language, he has no choice but to rely on what Davidson's truth-conditional approach suggests. Is it possible for the field linguist to construct an adequate interpretation theory for that language?

This thought-experiment (which I henceforth refer to as *the paradigmatic situation of radical interpretation*) is not Davidson's invention, but an adaptation of a similar scenario by Willard V. O.

Quine. It differs from the Quinean version, however, in two respects: Quine imagines his field linguist to compile a translation manual of the alien language, and to do so by relying on behaviouristic data (cf. Quine 1960, chapter 2). Davidson, on the other hand, rejects Quine's behaviourism and imagines his field linguist to deliver an interpretation theory. Both differences are important: firstly, an *interpretation theory* is something completely different from a *translation manual*. A translation manual might simply translate the alien's unknown language into some equally unknown other language, while an interpretation theory is necessarily a theory that interprets the utterances of the alien speaker into sentences, phrases, and words of a known language (for otherwise, it would not count as a theory that *interprets* the alien's linguistic behaviour). A Quinean translation manual is thus completely useless in order to scrutinize whether the application of Davidson's truth-conditional approach allows to arrive at an adequate conceptual description of the possibility of language-*understanding*. The second difference between the two thought-experiments concerns Quine's behaviourism. Quine focuses on behaviouristic data because he wants to capture semantic (and other mental) concepts in purely physiological terms. Quine thus pursues a *reductionist* project. Davidson, on the other hand, explicitly rejects reductionism, as we already saw in section 2.1 (cf. p. 16).

Davidson's field linguist is supposed, as already said, to interpret the hitherto completely unknown linguistic behaviours of alien speakers along the lines suggested by Davidson's truth-conditional approach. This means that the field linguist has to start his interpretative task with nothing more than his perceptions of the particular sound-waves emitted by the speakers (i.e. their utterances), his perceptions of the multitude of readily observable entities in his and the speakers' mutual surroundings, and – ex hypothesi – the assumption that he and the speakers share the very same concept of truth, and that this allows him to apply the Davidsonian truth-conditional approach to the alien linguistic behaviours. How does the field linguist proceed? Davidson imagines him to use the tools of the truth-conditional approach in order to find out which (if any) of the alien utterances happen to refer to which (if any) of the readily observable entities in the mutual surroundings. A good opportunity to do so is, for instance, if the field linguist repeatedly observes a speaker to utter the same sounds (say, "Gavagai!") whenever there is a particularly salient entity within their mutual surroundings (say, a rabbit). Such observations allow him to formulate hypothetical T-Sentences such as the following:

"Gavagai!" is true as spoken by the alien at t if and only if there is a rabbit visible to the alien at t.

The obvious next step is to try to find out whether that T-Sentence expresses a truth about the alien's language or not. The intuitive way to do so is to wait until there is again a rabbit clearly visible in the mutual surroundings, then to utter "Gavagai!" and to observe how the alien speaker reacts to that utterance. A single test of such a 'T-Sentence-hypothesis' is, of course, not conclusive. The field

linguist may take, however, the said hypothesis – i.e. that "Gavagai!" has to be interpreted as "There is a rabbit!" – as being provisionally verified if he may repeatedly judge the alien speaker's reactions to his "Gavagai!"-utterances as affirmative. On the other hand, the hypothesis is provisionally falsified if the alien's reactions seem to be negative. Of course, to determine whether a particular reaction of the alien speaker counts as affirmative, negative, or as neither of both will present the field linguist with a huge challenge, for he does not yet know how affirmation and negation is expressed in the alien language. Davidson suggests that he may try to identify the suitable sound-patterns in the alien language from the repeatedly received reactions to his own utterances during the testing of his T-Sentence-hypotheses. Given the (not very implausible) assumption that his doing so leads to the identification of two discernible patterns of sound-waves, he may take one of them as the alien language's word for "yes", and the other one for "no".

Essentially the same strategy might be pursued by the field linguist in order to identify the semantic constituents of the alien uttered sentences. If he has evidence that, say, "Gavagai!" has to be interpreted as "There is a rabbit!" and "Tuvagai" as "There is a hyena!", he may (provisionally) conclude that "Ga" has to be interpreted as "rabbit", "Tu" as "hyena", and "…vagai" as "There is …! ". These conclusions need to be empirically tested, of course, but this might again be done by means of suitable T-Sentence-hypotheses. A further important step concerns the unveiling of the basic syntactical structures of the alien utterances. This step is supposed to go hand in hand with the field linguist's testing of the above-mentioned hypotheses concerning the ways how "Gavagai!", "Tuvagai!" and the like have to be decomposed in order to reveal their semantic constituents. The field linguist's proceeding thus leads, at a certain point of time, to his having, on the one hand, a first partial knowledge of the alien language's vocabulary (its phrases and words) and, on the other hand, a first partial knowledge of that language's syntax. This is the moment where the field linguist may start to make use of the compositional structure, which he supposes the alien language to have. He may thus start to use his current knowledge of that language in order to formulate and verify (falsify) new and more complex T-Sentence-hypotheses. Doing so will allow him to consider T-Sentence-hypotheses that go far beyond the testing of reference-relations between particular utterances and readily observable entities such as rabbits and hyenas. It will allow, on the one hand, the formulation (and verification/falsification) of T-Sentence-hypotheses about what is not readily observable, but somehow hidden, or general, or abstract and, on the other hand, about what the alien speaker did not utter, but might have uttered. It will allow, in short, for the formulation (and verification/falsification) of a far wider field of T-Sentence-hypotheses and, eventually, for the formulation of a complete interpretation theory for the alien language.

This outline of the field linguist's task illustrates how Davidson imagines the radical interpreter to proceed.[13] His success obviously depends to an important degree on his intuition and ability to find the suitable T-Sentence-hypotheses for his empirical tests. At least as important as this,

---

[13] My portrayal of the paradigmatic situation of radical interpretation is, of course, abbreviated and simplified. For more elaborated descriptions cf. Davidson 1967, Davidson 1973, Davidson 1974 and Davidson 1975.

however, is his strict adherence to the so far ignored methodological maxims of the *Principle of Coherence* and the *Principle of Correspondence* (which are often conjointly referred to as the *Principle of Charity*[14]). What do these principles demand of the field linguist? Why are they so important? And why is the field linguist asked to adhere to exactly *those* principles? I will explain and attempt to justify them in the remaining pages of this chapter.[15]

The *Principle of Coherence* organizes the field linguist's activities by asking him to construct a *coherent* interpretation theory about the alien linguistic behaviour; i.e. it asks the field linguist to apply his own *logical standards* to the alien speaker's linguistic behaviour (Davidson 1973, p. 136f). This is doubtless a surprising methodological maxim. That the radical interpreter should comply with it is, however, backed by at least two convincing reasons. One of these reasons arises from the specific characteristics of Davidson's theory of Radical Interpretation, while the other is far more general and independent from Davidson's particular approach to semantics. (It is, of course, the latter reason that serves as the principle's justification.)

Let me begin with the theory-immanent reason: the radical interpreter's primary intention is, as explained above, to find the particular Tarski-style truth-theory that happens to allocate the truth-conditions to the alien utterances in exactly the way the alien speaker is himself allocating them to his utterances (i.e. to his uttered sentences). This intention is driven by two major convictions: (a) that the knowledge of the truth-conditions of the uttered sentences of the alien language reveals those sentences' literal meanings, and (b) that a Tarski-style truth-theory suitable to the alien language reveals the structure of that language (cf. section 2.1). The structure of the Tarski-style truth-theory is, however, the familiar structure of first-order predicate-logic. This is why the radical interpreter has a very strong motivation to impose this particular logical structure – i.e. his own logical standard – upon the alien speaker's linguistic behaviour. There is, thus, an obvious and strong reason to obey to the Principle of Coherence from within the perspective of the Davidsonian truth-conditional approach.

The other reason why the principle should be obeyed is independent of Davidson's approach. It has its roots in the fact that interpretation – whether radical or not – presupposes the *attribution of beliefs* to the speaker. We cannot, however, attribute beliefs to a speaker who seems to constantly contradict himself. We cannot, for instance, take the speaker to hold true that p and that not-p at the same time, for it is impossible to believe what is obviously logically incoherent. This does not mean that it is impossible to make mistakes or that speakers (and interpreters) are immune against occasional irrationality. A speaker (or interpreter) may, for instance, have inconsistent beliefs because he, for some reason, lacks the knowledge to recognize a mistake. He may likewise believe in entities which others do not believe to exist (phlogiston, for instance), or he may refuse to accept a thought or theory because it violates some conviction which he cherishes for whatever other reasons (as religious

---

[14] The name "Principle of Charity" is heavily misleading: it has nothing to do with charity (cf. my explanations above). I continue, however, to use the name, given that it is current in the philosophical debate (cf. Wilson 1959).

[15] My explanations regarding the Principle of Coherence as well as the Principle of Correspondence draw heavily upon Joseph 2004, pp. 62–70.

convictions, for instance, may prevent belief in the theory of evolution). The attribution of beliefs thus does not comply with what we may call "external consistency"; i.e. with what seems to be true according to contemporary science. What the attribution of beliefs complies with is rather what might be called "internal coherence"; i.e. the assumption that the subject to which the beliefs are attributed does not constantly contradict his own beliefs – *independently of how plausible or implausible these beliefs may look from an external point of view*. Such internal coherence, however, *is* prerequisite to the possibility of interpretation, for if it were impossible to discern any such logically coherent pattern in the speaker's mind, it would be impossible to attribute *any* beliefs to him: we could not, for instance, attribute to a speaker the belief that it is about to rain, if he denied at the same time that it is about to pour down, that rain is wet, that rain is water, etc (Davidson 1985, p. 196). The same is true, of course, with any other contradictory set of beliefs. Davidson thus concludes that if

> *we fail to discover a coherent and plausible pattern in the attitudes [beliefs, intentions, etc.] and actions of others we simply forego the chance of treating them as persons* (Davidson 1970, p. 221f).

But if we fail to treat others as persons we cannot treat them as speakers. *This* is why we (as well as the radical interpreter) are justified to apply our own logical standards to the (alien) speaker; i.e. why we are justified to adhere to the Principle of Coherence.

The *Principle of Correspondence*, on the other hand, asks the interpreter to transfer his *general theory of the world* into the speaker's mind. This is hardly a less surprising methodological maxim than is the Principle of Coherence at first sight. The best way to explain the importance of this second principle is to reconsider the particular situation within which Davidson's field linguist is supposed to start his interpretative task. This is, as we saw, the situation where he is supposed to share nothing with the alien speaker but a range of common perceptions about the uttered sounds, mutually perceptible entities, and, ex hypothesi, the same concept of truth. What the linguist lacks, however, is (among many other things) every single piece of knowledge about the *linguistic meanings* of the alien utterances and about the alien speaker's particular *propositional attitudes*. However, exactly these factors are decisive for the possibility of language interpretation, for the speaker (whether alien or not) typically uses his linguistic skills in order to express what he *believes*, *desires* or *intends*, and he does so by choosing exactly those sentences (phrases, words) which he takes to express his intentions, beliefs and desires.[16] It would be a great help then to the interpreter's task if he were able to put himself in a position where he would either know the particular literal meanings (= truth-conditions) of the sentences of the alien language or the particular contents of the alien speaker's propositional attitudes, since either of these pieces of knowledge could serve as a basis to decipher the other, and

---

[16] *"The methodological problem of interpretation is"*, as Davidson puts it, *"to see (...) what his [= the speaker's] beliefs are and what his words mean"* (Davidson 1975, p. 162). (Cf. also the short passage about mental holism in section 2.1, p. 16f.)

thus to advance his interpretative task. It is exactly this move which the Principle of Correspondence allows the field linguist to make (Davidson 1974, p. 152ff): the Principle of Correspondence asks the field linguist, as already said, to transfer his *general theory of the world* into the alien speaker's mind. This means that it asks the field linguist to assume the alien worldview to be (more or less) the same as his own. Doing so enables him to project himself into the alien speaker's position and to assume that what he holds (or would hold) to be true *in the alien position* is what the alien speaker actually holds (or would hold) to be true. The Principle of Correspondence thus gives the field linguist access to the alien speaker's propositional attitudes, for the field linguist knows what he, were he in the position of the alien speaker, would be holding true (or false); he thus knows what he, in the position of the alien speaker, would believe.

The Principle of Correspondence has often been criticized as resting on a grotesque overestimation of the convergence of propositional attitudes among rational beings. Jonathan Bennett, for instance, wonders why it should not be possible that *"(...) some creature's bad luck and intellectual frailty (...)"* could coincide in such a way that his beliefs are mostly wrong according to the interpreter's worldview (Bennett 1985, p. 610); and Jane Hale points out that a speaker could think and speak throughout most of his lifetime about *"alchemy, astrology or historical materialism"* (Hale 1997, p. 187); i.e. about things that we (or the radical interpreter) might hold to be empirically empty. Davidson's answer to Bennett's objection is that too much bad luck would simply undermine the possibility of interpretation:

> *The methodological advice to interpret in a way that optimizes agreement [between the interpreter's and the alien's worldview] should not be conceived as resting on a charitable assumption about human intelligence that might turn out to be false. If we cannot find a way to interpret the utterances and other behaviour of a creature as revealing a set of beliefs largely consistent and true by our own standards, we have no reason to count that creature as rational, as having beliefs, or as saying anything* (Davidson 1973, p. 137).

This is not to say that the Principle of Correspondence excludes the possibility of error altogether. A speaker may, of course, err according to the interpreter's standards (as much as the interpreter may err according to the standards of the speaker). But errors are detectable, Davidson says, only *"against a background of massive agreement"* (Davidson 1973, p. 137) between the interlocutors, for too many errors simply makes it impossible to detect a link between a speaker's utterances and the objects and events in the world to which (most of) his utterances usually refer. If it is impossible to regularly link the speaker's utterances (or parts of those utterances) with objects and/or events in the intersubjectively accessible world, it is impossible to interpret *any* of them. Davidson thus replies that whenever we want to make sense of someone's linguistic behaviour, we simply have no other choice than to suppose his worldview to be fairly similar to our own. It is because of this fact that the

interpreter (whether radical or not) is not only *justified* to apply the Principle of Correspondence, but even *forced* to do so.

Saying that the interpreter is forced to assume convergence between his and the speaker's worldview is not to say that this convergence is of the same degree with respect to all of their beliefs. It is surely very strong with respect to what is directly perceptible for if there was a considerable degree of divergence between the speaker's and the interpreter's beliefs about these things was it impossible to the interpreter to assume the speaker to have any beliefs at all. The degree of convergence decreases, however, with increasing distance between the directly perceptible entities in the interlocutor's mutual surroundings and the particular sentences which the speaker claims to hold true. This is because such sentences – i.e. sentences about the *general* or about *abstract* entities – may not conflict with the perceptible surroundings as directly as sentences about mutually perceived things or events can do. This explains why it is possible to disagree about the truth of alchemy or astrology (but likewise about astrophysics, genetics or discourse analysis). Hale's objection has thus no direct relevance with regard to the field, where the radical interpreter's task is supposed to start; i.e. with the discovery of reference-relations between sound-waves (= utterances) and the *directly perceptible* entities. Hale's objection thus neither shows radical interpretation to be impossible, nor the Principle of Correspondence to be an unconvincing overestimation of the convergence of propositional attitudes among rational beings. It only shows that divergence can be expected to increase the more general and abstract the individual's beliefs become. This is, however, no surprise, but rather in perfect correspondence with what is readily observable in everyday life.

The methodological Principles of Correspondence and Coherence constitute integral components of the field linguist's 'interpretative infrastructure': they allow the field linguist to distinguish between those T-Sentence-hypotheses that deserve to be empirically tested and those which can be ignored; they constantly report on the adequacy of the interpretation theory developed so far; they indicate when and where a revision of a part (or even the whole) interpretation theory is called for, and they show how such a revision might be carried out successfully. The Principles of Correspondence and Coherence are thus *essential* components of the field linguist's interpretative infrastructure. But given these maxims, and given the assumption that the field linguist and the alien speaker share the very same concept of truth, there is little reason to doubt that the field linguist may succeed in constructing a workable (if not adequate) interpretation theory of the alien speaker's linguistic behaviour. The thought-experiment of the field linguist thus demonstrates that it is, in principle, possible to apply Davidson's truth-conditional approach to our languages.

The case of the field linguist and the speaker of an alien language is, however, a *fictitious* case – it is nothing but the case of two *imaginary figures* in a *thought-experiment.* We still have to show then that these promising findings are, in principle, transferable to *real* cases: can we imagine a *real* interpreter to pursue the imagined field linguist's task and, in doing so, to deliver a workable (if not adequate) interpretation theory of the linguistic behaviour of a *real* alien speaker? Davidson devoted a

large part of his career to showing that the answer to this question is affirmative. Although the positive results of the thought-experiment give reasons for optimism, Davidson never denied the many and partially serious difficulties. We will return to them (and to Davidson's solutions) in the subsequent chapters (in particular in the chapters 5 and 6). First, however, let me close the present one by mentioning the first two important characteristics of the notion of *literal meaning* that already became noticeable between the lines in the past twenty or so pages.

The first characteristic is that a sentence's (phrase's, word's) literal meaning is restricted to that sentence's (phrase's, word's) *reference*. The reason for this is that the radical interpreter's verifications (and falsifications) of T-Sentence-hypotheses are verifications (and falsifications) of reference-relations between sentences (phrases, words) and things, events, etc. A sentence's (phrase's, word's) literal meaning cannot, then, comprise anything that goes beyond the respective expression's reference. Literal meaning therefore cannot have anything to do with the speaker's *communicative intentions*, the *illocutionary forces*, or the *contextual features* that typically happen to accompany the utterance of a sentence (phrase, word). This restricted understanding of the foundational notion of linguistic meaning – obviously a result of Davidson's externalism (cf. section 2.1, p. 16) – is seen by many philosophers as a serious handicap of Davidson's approach to semantics. Davidson on the other hand takes it to be one of the great virtues of his approach (Davidson 1967, p. 24). Whether it is a handicap or rather a virtue constitutes an important part of the discussion in almost all of the subsequent chapters of this study.

The second characteristic is that the reference of an expression's (= sentence's, phrase's, word's) literal meaning always remains somehow *indeterminate*. This surprising characteristic has its roots in the fact that for each of the alien expressions, there always exists more than just one true T-Sentence. Consider, for instance, the "Gavagai!"-case, and assume that the empirical evidence available to the field linguist suggests the following T-Sentence to be true of the alien's language:

(i)   "Gavagai!" is true as spoken by the alien at t if and only if there is a rabbit visible to the alien at t.

Then, the very same evidence will also suggest any of the following T-Sentences to be true of that language:

(i)   "Gavagai!" is true as spoken by the alien at t if and only if there are undetached rabbit parts visible to the alien at t.

(iii)   "Gavagai!" is true as spoken by the alien at t if and only if there is an incarnation of the rabbit god visible to the alien at t.

It is thus impossible for the field linguist to decide whether the alien speaker assents to (i), to (ii) or to (iii) (or to any other suitable T-Sentence). This phenomenon – known as the *indeterminacy of reference* – was first discovered by Quine (cf. Quine 1960, p. 73ff). He took it as proof for his sceptic conviction that there is never a *"fact of the matter"* (Quine 1980, p. 23) as to what is referred to by a particular expression (sentence, phrase, word). Davidson, on the other hand, acknowledges the existence of the phenomenon (Davidson 1983, p. 145), but rejects the Quinean conclusion: the radical interpreter's search for a Tarski-style truth-theory suitable to the alien language *de facto* amounts, he says, to the search of an accurate representation of the particular pattern of the alien speaker's propositional attitudes. This is because the field linguist's verification/falsification of T-Sentence-hypotheses reveals what the alien speaker happens to believe, what he hopes and intends, and under what conditions he does so. The interpreter's attempt to verify a particular T-Sentence-hypothesis might thus be understood as an attempt to locate a particular propositional attitude of the alien speaker – as it is expressed in such-and-such an utterance – within the speaker's entire network of propositional attitudes. The verification of a T-Sentence like (i) or (ii) or (iii) thus reveals what "Gavagai!" *refers* to: what it refers to is that the alien speaker happens to believe that "There is a rabbit (or undetached rabbit parts, or an incarnation of the rabbit god, etc.)". It is thus wrong to claim that there is *"no fact of the matter"* regarding what an expression refers to: it refers to exactly those things or events in the intersubjectively accessible world which are meant – according to the speaker's conviction – by the respective expression. The problem is only that it is impossible to the interpreter to decide which of the different possibilities – say (i), (ii), or (iii) – correspond to the speaker's conviction. It is in *this* sense that there truly is an indeterminacy of reference. It is, however, a comparatively harmless indeterminacy, for it neither renders the concept of reference void, nor does it prove there to be no semantic facts. *"[W]hat a speaker means"*, Davidson says, *"is what is invariant in all correct ways of interpreting him"* (Davidson 1999, p. 81). The indeterminacy of reference-thesis thus says that there is more than one correct interpretation; but it does not say that there is no meaning or no reference.

3 The communication-intentionalists' challenge

The most radical opponents to Davidson's truth-conditional approach are doubtless the *theorists of communication-intention*: They object not only to Davidson's approach as such, but to the underlying formalist's[17] credo alike. The structures of languages are secondary, they claim, for what is basic to the use of language is its *communicative function*. A theory of meaning and language-understanding has thus to be constructed with regard to this function and not with regard to linguistic structure. The concept of truth, or any other concept suitable to do justice to the said structures, is thus the wrong place to start. What we have to take as foundational to a promising theory of meaning and language-understanding is rather, they say, the speaker's possession of *audience-directed communicative intentions*.

The conflict between the formal semanticists and the theorists of communication-intention is not only one of the most fundamental, but also one of the most complex conflicts in the philosophy of language; it possesses, as Peter Strawson once said, a *Homeric quality* (Strawson 1969, p. 172). The present chapter's aim is not to solve this conflict, but just to show that Davidson's truth-conditional approach is capable to meet the communicationalists' challenge. It is, of course, impossible to discuss each of the numerous objections against Davidson's work, for doing so would lead too far away from the central purpose of this study. The present chapter thus discusses the communication-intentionalists' challenge in a somewhat incomplete and selective way. Neither should one misunderstand the present chapter as an attempt to prove the superiority of Davidson's approach over the communicationalists' alternatives. My intention is much more modest: I hope to convince you that the Homeric struggle between the formalists and the communicationalists is fought by equally well prepared rivals. In order to do so, I proceed in two steps: first, I show why the communication-intentionalists do not succeed in proving the inferiority of Davidson's approach (section 3.1); and subsequently, I argue that the communication-intentionalists' alleged strength over formalist approaches cannot be proven (section 3.2). The concluding section 3.3 briefly evaluates my arguments.

3.1 Proving the inferiority of Davidson's approach – not as easy as it seems

The communication-intentionalists typically present their case in three steps. These are, firstly, the definition of their foundational concept of communicative intention; secondly, the demonstration that linguistic meaning can be explained in terms of that concept; and thirdly, the proof that linguistic meaning must be explained in terms of that concept. Subsection 3.1.1 presents a rough outline of the communication-intentionalists strategy with regard to the first two steps. Subsection 3.1.2 presents

---

[17] For readability's sake, I occasionally abbreviate "formal semanticist" to "formalist" and "theorist of communication-intention" to "communicationalist" or "communication-intentionalist".

their third step along the lines of Strawson's reflections in *Meaning and Truth* (Strawson 1969), and subsection 3.1.3 explains why Strawson's argument is insufficient to discredit Davidson's formalist approach, and why it would take a lot more to do than what Strawson (and many others) apparently supposes.

3.1.1 Explaining linguistic meaning in terms of communicative intentions

The first step of the communication-intentionalists' strategy consists, as mentioned above, in the definition of the foundational concept of communicative intention. This is arguably the most challenging step, for it is characterized by various and notoriously difficult problems. One such problem concerns the fact that the foundational concept of communicative intention has to be defined in entirely *non-linguistic terms*: every attempt to do otherwise inevitably leads to a circular argument. A second problem is that the said concept has to be such that it comprises *all but only those* actions that might be reasonably considered as being carried out with a communicative intention. This poses a problem to the communicationalists, for the typical means of non-linguistic communication (say, handsigns, body movements, vocalizations, etc) are *not* specifically communicative in nature, but may occur for other reasons, or for no reason at all: a certain movement of a hand may count as a means to communicate "Good-bye", but it might likewise be a symptom of Alzheimer's disease; a cry might be uttered with the communicative intention to warn you from imminent danger, but it might just as well be the non-intentional reaction of somebody who has burnt his finger on a hotplate. A third problem is that "communicative intention" has to be defined in such a way that it allows the interpreter (or "audience", as the communicationalists prefer to say) to recognize what *particular proposition* the speaker (or rather: the utterer) aims to communicate while performing such-and-such an action: is he, by waving his hand, communicating a "Good-bye" or a "Wait a moment"? Is his using his voice in such-and-such a way a warning against a danger or rather an expression of joy? It is often held that among the various attempts to define the foundational concept of non-linguistic communicative intention, the proposal by Paul Grice is by far the most promising one (cf. Grice 1967/1987). I shall not present it at this stage, but postpone its discussion to subsection 3.2.2.

Assume that there is a suitable definition of "non-linguistic communicative intention": how does the communication-intentionalist then go on to explain linguistic meaning in terms of this concept? The central difficulty with this second step is that the communicative possibilities of a language go far beyond than what might be expressed by non-linguistic means. How, for instance, could one non-linguistically communicate the content of, say, the preceding sentence? The question is tendentious, of course, since the communication-intentionalists are not required to explain how each particular sentence of a natural language could be translated into the 'vocabulary' of non-linguistic communicative intentions. What they are required to show is 'only' that it is possible to explain the existence of a basic set of linguistic items (utterances and constituents of utterances) in terms of a

concept of non-linguistic communicative intention, and that each sentence of a natural language – independently of its particular degree of complexity – might, *in principle*, be traced back to that basic set of linguistic items.

One way to achieve this is to claim that linguistic meaning has to be conceived as governed by *rules* or *conventions*, and that the emergence of the first such rules/conventions is due to the utterer's ability to successfully and repeatedly convey a set of non-linguistic communicative intentions to his audience. These first primitive conventions/rules then constitute the first primitive linguistic items from which the more complex ones derive, until one eventually arrives at the level of a fine-grained vocabulary and a fully developed syntactic structure as they are both typical to natural languages. The decisive step, however, is the initial one: how do the communication-intentionalists explain the emergence of the first conventions/rules from the speaker's non-linguistic audience-directed communicative intentions?

One way to do so is to claim that this emergence might occur quite naturally among the members of a particular community if they experience that at such-and-such an occasion, the performance of a particular non-linguistic action is understood by the community-members in the way intended by its utterer. Suppose, for instance, that an utterer successfully communicates the proposition that p to his audience by means of making a particular movement, or by emitting a particular sound – why should he not repeat this movement or emit this sound again if he intends to communicate the same proposition again at some later occasion? The utterer's initial success obviously provides him – and the other members of his community – with a strong incentive to repeat exactly the same action when the same proposition is to be conveyed again. The case could be generalized: if it is possible to communicate that p by emitting the sound x, it is possible to communicate that q by emitting the sound y, to communicate r by emitting the sound z, etc. The first primitive *semantic* conventions/rules thus appear on the scene. They are joined by a first set of *syntactic* conventions/rules as soon as the utterer and the audience have come to a (tacit) agreement about how the already understood sounds (= utterances) x, y, and z can be decomposed and how the so derived component sounds can be rearranged in order to communicate further propositions. Such an agreement occurs as soon as an utterer succeeds to communicate a particular proposition by means of the utterance of a sound that is partly the same as, and partly different from, some already understood sounds (i.e x, y, and z). These primitive syntactic and semantic conventions/rules obviously allow for a primitive form of linguistic communication only. They explain, however, the origins of our languages' fine-grained vocabularies and full-fledged syntactic structures. They explain, in other words, the link between our everyday language use and the non-linguistic concept of communicative intention.

3.1.2 Strawson's argument against Davidson's truth-conditional approach

Let us assume that linguistic communication can be explained in terms of non-linguistic communicative intentions. Does it then *have* to be explained in this way? The most straightforward way to prove that it does is to argue that every alternative approach is necessarily bound to fail. This is, in essence, Strawson's strategy. Strawson actually cuts a few corners by claiming that the truth-conditional approach is the only approach which has *"ever (...) [been] seriously advanced (...) as providing a possible alternative to the thesis of the communication-intentionalist"* (Strawson 1969, p. 176), but this need not bother us. What he claims against the truth-conditionalists (and, thus, against Davidson) is:

(a) that our knowledge of truth-theories does not suffice to explain the intelligibility of the concept of truth, and

(b) that the intelligibility of the concept of truth is explainable only in terms of its function in communication.

With respect to (a), his observation is that a theory of truth is not telling us by itself what it is a theory *about*: a person's knowing a particular truth-theory (say, a Tarski-style truth-theory) cannot by itself explain that person's acquaintance with the concept of truth (Strawson 1969, p. 180). Strawson thus concludes that there must be something more to the concept of truth than particular (Tarski-style) truth-theories. What Strawson believes there to be more to truth is, of course, that concept's function in communication (= claim (b)).[18]

How, then, does Strawson argue in favour of (b)? Having stated that the knowledge of a particular truth-theory cannot explain why we understand the concept of truth, Strawson asks what else might do so. His answer is disappointing: the only facts which he believes to be *"uncontroversial and fairly general"* are observations such as the following:

> *One who makes a statement or assertion makes a true statement if and only if things are as, in making that statement, he states them to be. Or (...): one who expresses a supposition expresses a true supposition if and only if things are as, in expressing that supposition, he expressly supposes them to be* (Strawson 1969, p. 180).

---

[18] Notice that (a) is put forward also by authors of the formalist side of the struggle. Michael Dummett, for instance, argues that a (radical) interpreter has no idea about *"(...) the point of introducing the [truth-] predicate (...) unless (...) [he] already know[s] in advance what the point of the predicate so defined is supposed to be. But, if (...)[he] know[s] in advance the point of introducing the predicate "true" (...), [he] know[s] something about the concept of truth expressed by that predicate which is not embodied in that (...) truth definition [= truth-theory]"* (Dummett 1978, pp. xx–xxi). Where Dummett differs from Strawson is with respect to the conclusion he draws from (a): contrary to Strawson, he does not believe it to have any bearing on the alleged conceptual priority of the concept of communication (or communicative intention), but rather that it proves the concept of truth to be an insufficient foundation for a formalist approach to semantics. I will come back to Dummett's view in the chapters 4, 5 and 6.

These observations look completely harmless at first sight. Strawson argues, however, that they inevitably lead to the insight that the intelligibility of the concept of truth necessarily depends upon a prior grasp of the concept of communicative intention. He argues as follows: Davidson claims that it is the assigning of truth-conditions to a language's sentences that determines the particular meanings of that language's sentences. In Davidson's framework, truth-conditions thus function as meaning-determining *rules*.[19] Strawson accepts, for the sake of argument, the correctness of this assumption. He continues, however, that the problem for Davidson's approach arises as soon as this assumption is combined with the *"uncontroversial and fairly general"* observations about the concept of truth from the passage quoted above: the obvious result is, Strawson says, that a statement's (or assertion's, supposition's) particular linguistic meaning is taken to be determined by the rules which determine how things are stated to be (asserted to be, supposed to be). These rules – i.e. the statement's (assertion's, supposition's) truth-conditions – vary, however, from context of utterance to context of utterance, for Davidson relativizes the T-Sentences of his Tarski-style truth-theories to speakers and points of time (cf. section 2.1, p. 15f). It thus follows that a statement's (or assertion's, or supposition's) particular meaning (= truth-conditions) has to be taken as determined by what is stated (or asserted, or expressly supposed) by a particular speaker at a particular occasion of utterance. But in order to understand what it *is* that is stated (or asserted, or expressly supposed) by such-and-such a particular utterance, we are required, Strawson continues, to understand what it *is* to *state* something (or *assert* something, or *suppose* something). We are, in other words, required to understand what particular *speech act type* the utterer happens to perform with a particular utterance. The understanding of the concept of speech act type, however, is unintelligible without a prior understanding of the concept of communicative intention. It thus follows that the intelligibility of the concept of truth depends upon a prior grasp of the concept of (non-linguistic) communicative intention (Strawson 1969, p. 180f). Davidson's truth-conditional version of formal semantics thus turns out to be based on the communication-intentionalists' foundations: the Homeric struggle between the formal semanticist and the communication-intentionalist thus ends in the triumph of the latter.

3.1.3 Refuting Strawson's argument

Does Strawson's argument really show the conceptual superiority of the communication-intentionalists' approach over Davidson's truth-conditional variety of formal semantics? I believe that it does not. My counterargument falls into two parts: the first shows why Strawson's argument, as summarized above, misses its target; and the second explains why it is more difficult to argue against Davidson than Strawson (and many other communication-intentionalists) seems to think.

---

[19] Notice that the word "rule" (or "convention"), if applied to Davidsonian semantics, has to be understood in a very broad sense that does not need to coincide with any of the common usages of "rule" (or "convention"). For a detailed discussion of Davidson's views regarding the importance of rules/conventions in semantics, cf. the chapters 5 and 6.

According to Strawson, the concept of (non-linguistic) communicative intention is more fundamental to meaning and language-understanding than the concept of truth, (1) because the assignment of truth-conditions to speech acts (statements, assertions, suppositions, etc.) requires a prior understanding of those utterances' respective speech act types, and (2) because the understanding of those speech act types is unintelligible without a prior grasp of the concept of (non-linguistic) communicative intention. I agree that (1) and (2) are both correct. I do not believe, however, that Davidson's approach violates these claims. It is, of course, true that a radical interpreter cannot know what it is *in the language of an alien speaker* to perform an utterance of such-and-such a speech act type, for to be ignorant about this is part of what it means to be a *radical* interpreter. The interpreter knows, however, what it is to have a communicative intention and what it is to utter such-and-such a speech act type with respect to *his own language*. So why not simply apply *this* knowledge to the alien speaker's linguistic behaviour? Is it not exactly this what the Principle of Charity requires the interpreter to do (cf. section 2.2, p. 20ff)? – The interpreter may, of course, be mistaken about what particular speech act the alien speaker is performing on such-and-such an occasion. He may be similarly wrong with respect to the truth-conditions he assumes the alien speaker to assign to his utterances. But these are mere practical problems without any conceptual relevance to the possibility of the radical interpreter's task. Davidson's theory of Radical Interpretation thus violates neither (1) nor (2). As a consequence, Strawson's argument misses its target.

However, is this course of reasoning not merely cheating? The answer is in part "yes", and in part "no". It is "no" with respect to the fact that it shows *this particular* argument to miss its target. It is "yes", however, for it claims Davidson's radical interpreter to possess the required knowledge about the concept of truth and the concept of communicative intention *without* explaining where this knowledge comes from. It is "yes" alike for it neither explains how the interpreter became acquainted with these concepts. My so far explained argument does thus not settle any relevant issue; it merely shows that everything remains undecided so far.

Now, do we need to regard the Homeric struggle in terms of one of the rivals being the ultimate winner, and the other one the struggle's loser? The answer depends on whether the two approaches really need to regard one of the two disputed concepts – "truth" and "non-linguistic communicative intention" – to be prior to the respective other. But this is clearly not the case with respect to Davidson's project. The reason for this is that his approach to meaning and language-understanding is decidedly *descriptive* in character. Consider, for instance, the opening passage of Davidson's programmatic essay *Radical Interpretation*. It reads as follows:

> *Kurt utters the words "Es regnet" and [we know] under the right conditions (...) that he has said that it is raining. (...) How could we come to know [that] (...)? (...) [G]iven a theory that would make interpretation possible, what evidence plausibly available to a potential interpreter would support the theory to a reasonable degree?* (Davidson 1973, p. 125)

Davidson's answer to this question is, of course: the knowledge of the truth-conditions which Kurt can be taken to assign himself to his own utterance. Claiming this is, however, not claiming truth-conditions to be *identical with*, or *constitutive of*, what Kurt means by "Es regnet". Davidson only claims that the assignment of truth-conditions allows the interpreter – ignorant of Kurt's language – to *determine* what Kurt has meant by what he uttered. Remember: Davidson does not take Tarski-style truth-theories to be theories of meaning (and language-understanding); he is only claiming Tarski-style truth-theories to be *vehicles* towards theories of meaning (and language-understanding) (cf. section 2.1, p. 16). It is because of this that Davidson's approach is decidedly *descriptive* in character.

On the communication-intentionalist's side, things are somewhat different. I do not want to claim that Strawson or other communication-intentionalists *identify* meaning with communicative intentions (although such reductionist ambitions are sometimes to be found between the lines[20]). But what they clearly want to do is to clarify what is *constitutive* of linguistic meaning. This becomes evident, for instance, from the communication-intentionalists' three-step procedure I presented above (cf. subsections 3.1.1 and 3.1.2). Further evidence comes from the questions which Strawson asks in the first lines of his *Meaning and Truth*:

> *What is it for anything to have a* meaning *(...) in the way, or in the sense, in which words or sentences or signals have meaning? What is it for a particular sentence to have the meaning (...) it does have? What is it for a particular phrase, or a particular word, to have the meaning (...) it does have?* (Strawson 1969, p. 171)

These are questions about what is *constitutive* of a sentence's, word's, or signal's meaning, but not about what *describes* a sentence's, word's, or signal's meaning.

The difference between the rival approaches is important, for it allows Davidson to concede that a person's acquaintance with the concept of non-linguistic communication-intention could occur *side by side* with his acquaintance with the concept of truth. The communication-intentionalists, on the other hand, cannot do so, for allowing the concept of non-linguistic communicative intention to enter the scene *in parallel* (i.e. side by side) with the concept of truth would be to concede that the concept of truth is potentially as constitutive to meaning (and language-understanding) as the communicationalists take the concept of non-linguistic communicative intention to be. Davidson, on the other hand, can concede (and actually *does* concede; cf. chapter 4) parallel acquaintance, for he is not concerned with what *constitutes* meaning. It is, thus, irrelevant to his *descriptive* project whether acquaintance with the concept of truth occurs in parallel with acquaintance with the concept of non-linguistic communicative intention, *as long as it is possible* to describe meaning and language-understanding without making use of the latter concept.

---

[20] A reductionist ambition has been attributed to Grice's approach. Grice did not deny having been tempted to think in those terms, but neither did he explicitly endorse it. Cf. Grice 1987.

Let me return now to Strawson's argument. Strawson claims, as we saw, that (1) the assignment of truth-conditions to speech acts requires a prior understanding of those speech acts' speech act types, and (2) that the understanding of speech act types is unintelligible without a prior grasp of the concept of non-linguistic communicative intention. Both claims are, as I said, correct. Isn't this incompatible with Davidson's view? No, it is not; for neither (1) nor (2) show the concept of communicative intention to be *prior* to the concept of truth; they both allow for the possibility that the two concepts are grasped *in parallel* with each other. But this suffices, as we just saw, to save Davidsonian semantics from Strawson's attack: it suffices because Davidson does not claim anything about the constitution of the concept of meaning, but only about its being describable in terms of such-and-such a concept; namely in terms of the concept of truth.

Where does all this lead to? There are, I believe, two conclusions to be drawn: one is that the communication-intentionalists have to present an argument much stronger than the one presented by Strawson. More precisely, they are required to present an argument that prohibits Davidson from making use of the elbow-room the 'descriptive turn' gives to his truth-conditional approach (i.e. his being able to concede the parallel acquaintance of the concepts of truth and non-linguistic communicative intention). The second conclusion is that Davidson is forced to make use of that elbow-room: it is only the possibility of the parallel acquaintance of the concept of truth and the concept of communicative intention that allows him to avoid the Strawsonian conclusion.

We must thus ask: can it be shown that we acquire the two concepts in parallel to each other? The answer to this question is not only important with regard to the fate of Davidson's truth-conditional approach. It is just as important with respect to the question I asked about the Homeric struggle and its outcome. The reason is that there arises a new possible solution to the struggle if we really *do* acquire the two concepts in parallel. The new solution is that the two rival approaches may (peacefully?) *coexist* side by side to each other.

It is, however, not yet the time to discuss this issue, for I first have to complete my argument in favour of Davidson's truth-conditional approach. I thus postpone the question of how Davidson makes use of his elbow-room to the next chapter and continue the present one with a few comments on potential weaknesses of the communication-intentionalists' approaches. These weaknesses clearly make no positive claims about Davidson's approach. They are nevertheless important to our discussion in that they reveal an additional aspect of the Homeric struggle: up to now it looked as if the communication-intentionalists had a clear advantage over Davidson's truth-conditional approach. The next section shows that this is an illusion.

3.2 Are the communication-intentionalists as strong as they suggest?

It is a truism that the communication-intentionalists' primary focus is on *communicative acts*. They thus consider language *in use*. The formal semanticists, on the other hand, prefer to start at a more abstract point: they begin with general considerations about our languages' structures and postpone problems of meaning and use to a later stage (cf. section 2.1, p. 8f). The formal semanticists' strategy therefore seems to be far more technical and theoretic than the one pursued by the communication-intentionalists: it lacks the communicationalists' proximity to everyday linguistic phenomena. Does this show the communicationalists to be better off than the formal semanticists? No, it does not, for there are two unsolved problems with the communicationalists' theories which are either absent from, or irrelevant to, the formalists' theories. One concerns the question of how to make communicationalist theories compatible with the formal structures of languages, while the other concerns the definition of the foundational concept of non-linguistic communicative intention.

3.2.1 Problem I: How to make the communication-intentionalists' theories compatible with linguistic structures?

Section 2.1 showed there to be two relevant structures in natural languages: the *contextual* structure and the *compositional* structure. Communicationalist theories conform almost automatically to the contextual structure, while the compositional structure causes very serious problems. Both these facts have a common ground: the communicationalists explicit focus on communicative acts. This is particularly easy to see with respect to the contextual structure: the communication-intentionalists' theories almost automatically conform to it because their communicative focus encourages them to study words (or semantical primitives) in the context of utterances – and this obviously minimizes the temptation to regard them in isolation.[21] There is no such natural suitability, however, with respect to the compositional structure. The communication-intentionalists thus have to take an extra step to prove the compatibility of their theories with this structure. This is, however, not easy to do. Consider, as an example, what Strawson says about the issue in *Meaning and Truth*:

> *The meaning of a sentence is a syntactic function of the meanings of its parts and their arrangement. (…) [T]here is no reason in principle why a (…) utterance should not have a certain complexity – a kind of complexity which allowed an utterer, having achieved one communication-success, to achieve another by repeating one part of his utterance while varying the other part, what he means on the second occasion having something in common with, and something which differentiates it from, what he meant at the first occasion. And if he does thus achieve a second success, the way is open for a rudimentary* system *of utterance-types to become established; to become conventional within a group* (Strawson 1969, p. 175).

---

[21] Utterances such as "Gavagai!" or "Mama!" do not provide counterexamples; although containing just one word, they are obviously *complete utterances*.

There is nothing wrong with this idea – but it hardly provides more than the mere *nucleus* of an explanation of how to make a communicationalist's theory compatible with a language's compositional structure: it shows that it is possible to explain how utterances might be 'sliced' into smaller parts; i.e. into a *"rudimentary system of utterance-types"*; but it does not yet show that these utterance-types allow the introduction of a compositional structure into the theory. The reason is that it is not yet clear whether the said utterance-types are such that they 'slice' the speaker's utterances into their respective semantically and syntactically *primitive* components; and if they do not, they are hardly a help in making the theory do justice to the languages' compositional structure. Strawson, or any other conventionalist, must thus take an extra step: they have to prove that their utterance-types 'slice' the utterances in the required ways. More importantly, however, the conventionalists simply presuppose language to be organized by means of *conventions*. But conventions are, as Davidson argues, neither prerequisite to the determination of a sentence's (phrase's, word's) meaning, nor to the understanding of a particular language (cf. chapter 5). Strawson's proposal – as well as any other proposal based on conventions (or rules) – is, thus, at least problematic.

Is there an alternative to the convention-based strategy? Paul Grice has shown that there is: instead of connecting a speaker's communicative intentions by means of conventions to a language's semantic and syntactic constituents, he proposes to do so by means of so-called *resultant procedures* (Grice 1967/1987a). The difference between conventions and resultant procedures concerns their respective constitution: conventions are deduced from whole utterances (cf. the quoted passage on the previous page), and resultant procedures from those utterances' constituents. Grice thus takes the particular meaning of an utterance to be the result of the particular contents of its semantic and syntactic constituents; and those constituents to possess a particular semantic or syntactic content because of their being constituted by the respective resultant procedures. His 'dissecting' *utterances* into their respective semantic and syntactic components thus resembles the way Davidson 'dissects' *sentences* into their respective components. There are thus good reasons to assume the Gricean strategy to be equally well suited to conform to our language's compositional structure as is Davidson's formalist approach.

There is, however, an important difference between their strategies: Davidson develops his solution to the compositionality-problem over an investigation of the structures of *sentences*, while Grice starts from a consideration of the structures of *utterances*. The difference is important because it forces Grice to introduce his solution to the compositionality-problem at a stage where a language's sentences are already assumed to be used; i.e. where they are already performed as a means to communicate such-and-such a communicative intention. The obvious consequence is that he has to assume the speaker to possess a prior grasp of the concept of denotation; if he did not, it would be impossible for the speaker to understand what the Gricean resultant procedures were procedures of. Grice's problem is, however, that assuming the speaker to possess a prior knowledge of the concept of denotation is not a particularly plausible assumption: even authors sympathetic to the Gricean

approach agree that the speaker is *"unlikely to have"* that knowledge (cf. Avramides 1997, p. 79). Grice is well aware of the difficulty, but he does not present a plausible solution: a speaker understands the said concept, he says, *"in some sense, implicitly"*, but how exactly this is to be understood remains, as Grice concedes, an *"unsolved mystery"* (Grice 1967/1987a, p. 136). There is, thus, an unsolved problem with the Gricean approach: it might provide a solution to the communicationalists' compositionality problem if it was possible to eliminate the *"unsolved mystery"*; but it is exactly this lacuna which Grice does not know how to remove. Neither the Gricean nor any conventionalist approach thus succeeds in showing that communicationalists' theories are compatible with our languages' compositional structures. With regard to compositionality, Davidson's truth-conditional approach is thus in a better position.

3.2.2 Problem II: How to define the foundational concept of non-linguistic communicative intention?

The concept of non-linguistic communicative intention is to the communicationalists what the concept of truth is to Davidson: the very foundation upon which their respective approaches to meaning and language-understanding are built. Both are thus required to demonstrate that their respective foundational concepts provide an adequate basis for a theory of meaning and language-understanding. That there is a so far unresolved problem on Davidson's part has been already mentioned: Tarski-style truth-theories do not tell us by themselves what they are theories of (cf. subsection 3.1.2, p. 30). Davidson is thus forced to show that there is more to the concept of truth than what is captured by the Tarski-style truth-theories (cf. chapter 4). On the communication-intentionalists' part, however, exists an equally pressing problem with regard to how to *define* their foundational concept: is it possible to define "non-linguistic communicative intention" in such a way that it captures *all and only* those intentions that are *communicative* intentions? This is the issue at which I briefly want to look in the present section. More precisely, I want to address Grice's definition of "non-linguistic communicative intention" (or "M-intention", as Grice prefers to say). This restriction seems justified in so far as Grice's is often said to be the most promising such definition from the communication-intentionalists' camp.

Here is Grice's definition (Grice 1967/1987, p. 92): An utterer U *M-intends* by his uttering x that p:

(1) if some audience A produces r,
(2) if A thinks U intends A to produce r,
(3) and if A thinks U intends the fulfilment of (1) to be based on the fulfilment of (2).

The audience's reaction (= r) is, of course, a belief – namely, the belief that U, in uttering x, M-intends that p (in other words: that U, in uttering x, *intends* to *communicate* that p). The utterance x, on the

other hand, does not need to be the utterance of a sentence or a word, but could equally well be the waving of a hand, a particular movement of the limbs, or a non-linguistic use of the voice. The scope of the Gricean definition is thus *not* restricted to linguistic communicative intentions, but broad enough to capture non-linguistic communicative intentions as well. The definition thus fulfils an important requirement of the communication-intentionalists' approaches: it defines their foundational concept in non-linguistic terms.

But does it capture *all and only* communicative intentions? Doubts would arise if the understanding of a particular utterance turned out to be possible independently of the audience's grasp of the utterer's respective M-intention. That there are such cases has been argued by different authors. Ian Rumfitt (1995), for instance, presents the case of Galileo Galilei, who was brought to court because of his adhering to heretical non-Ptolemaic ideas: was it relevant to the verdict, Rumfitt asks, whether Galileo's judge – Bellarmine – was grasping Galileo's M-intention as he claimed to believe that the earth does not move? There is clearly no doubt that Galileo had good reasons to maintain the said intention, Rumfitt says, for the impending severe sanctions in case of a sentence did not invite him to M-intend anything else. But it seems, Rumfitt continues, that Galileo's having that particular M-intention was completely irrelevant to Bellarmine's understanding of Galileo's "The earth does not move". The reason for this is that Bellarmine was well aware of Galileo's intellectual career and thus knew that Galileo had subscribed to a non-Ptolemaic worldview for years. Bellarmine had thus good reasons to believe that Galileo's revocation was more likely motivated by his wish to escape the impending punishment than indicating that he had reconsidered his views about the organization of the solar system. It must have been clear to Galileo, however, that Bellarmine was likely to think along these lines. Galileo thus knew that it was, in principle, irrelevant to Bellarmine's further proceedings whether he (= Galileo) was able to make Bellarmine believe that he adhered to the Ptolemaic worldview by means of his uttering the suitable sentence (i.e. "The earth does not move."). This fact did not, however, impede Bellarmine's understanding of Galileo's utterance: they both spoke the same language; they both knew that Galileo's utterance meant that the earth does not move. Rumfitt thus concludes that Galileo's M-intention was irrelevant to Bellarmines understanding of Galileo's utterance (Rumfitt 1995, p. 834f).

Does the Galileo/Bellarmine-case thus show the Gricean definition (or any other such definition) to be insufficient to capture what is required for the understanding of linguistic communication? The answer is negative: Grice (or any other communicationalist) could still argue that the situation is *deviational*; i.e. that it worked as it did only thanks to Bellarmine's *prior* understanding of the relevant utterance, and that this prior understanding occurred only thanks to his grasp of the very M-intention that is typically involved in this particular utterance in such-and-such a linguistic community. The example shows, however, that a Gricean M-intention (or any otherwise defined communicative intention) is not always required for an audience's understanding of a speaker's

utterance. The Galileo/Bellarmine-case thus suggests that Grice's definition is somewhat too narrow: it does not capture *all* cases of successful (linguistic) communication.

More troublesome than the case just discussed is, however, its opposite; i.e. the case that shows Grice's definition to capture *more* than just communicative acts. That it is possible to construct such a case has been shown by Peter Strawson (Strawson 1964, p. 156ff). It goes as follows: imagine an utterer U to intend by means of a certain action to induce in some audience A the belief that p (i.e. the fulfilment of condition (1) in Grice's definition). Imagine, furthermore, that U's evidence that p is merely arranged 'evidence' that p, and imagine that U has arranged that 'evidence' at a place where A was able to observe U doing so. This rather complex situation is given, for instance, if U – while being observed by A – applies lipstick on the shirt of A's husband, and if U is doing so with the intention to make A believe something – say, that A's husband has been unfaithful.[22]

Imagine furthermore that U knows that A is watching him manufacturing the said 'evidence', and imagine that U has reasons to assume that A does not know that U knows that A is observing U. Given all this, we are confronted with a case where U

> *(…) realizes that A will not take the* arranged *'evidence' as genuine or natural evidence that p, but realizes, and indeed intends, that A will take his [= U's] arranging of it as grounds for thinking that he (…) intends to induce in A the belief that p [i.e. that A's husband has been unfaithful]* (Strawson 1964, p. 156).

This corresponds to the fulfilment of condition (2) in Grice's definition. There remains then condition (3). It is fulfilled, Strawson argues, if U knows that A has good reasons to believe that U would not intend to make A believe that A's husband has been unfaithful if it were not known to U that A's husband really has been unfaithful. But such a case is, as Strawson writes,

> *(…) not a case of attempted* communication *in the sense which (…) Grice is seeking to elucidate. A will indeed take U to be trying to bring it about that A is aware of some fact; but he [= A] will not take U as trying, in the colloquial sense, to "let him know" something (…). But unless U at least brings it about that A takes him [= U] to be trying to let him [= A] know something, (…) has [U] not succeeded in communicating with A (…). It seems a minimal further condition of his [= U's] trying to do this that he [= U] should not only intend A to recognize his intention to get A to think that p, but that he should also* intend A to recognize his intention to get A to recognize his intention *to get A to think that p* (Strawson 1964, p. 156f).[23]

---

[22] The illustration of Strawson's case is due to Schiffer 1987, p. 245.
[23] In the original passage, Strawson uses S (= Speaker) instead of U (= Utterer).

The problem is, however, that Grice cannot simply amend his definition with such a fourth condition, for the so amended definition could be challenged again with a (more complex) variety of the case of the above mentioned kind, which could in turn be met only by adding a further condition to the already amended definition, and so on ad infinitum.

Is it possible to avoid the regress? Since the Strawsonian case is built on the fact that U and A possess different knowledge about the evidence which is making A to believe that p, is it tempting to block the case by means of adding a condition to the Gricean definition which demands U and A to possess a certain kind of *mutual knowledge* about the respective evidence. The problem with this solution is that it blocks the Strawsonian case only if U's and A's mutual knowledge (almost) perfectly overlap. This is, however, an unlikely assumption: people have different interests and different biographies; they know many things others do not. But it is exactly such non-overlapping knowledge that allows for the Strawsonian case (cf. Schiffer 1987, p. 246). It is for essentially the same reason that Jonathan Bennett's alternative solution does not do much better. Bennett proposes to add a *"clause of reliance"* to the Gricean definition, which then reads as follows:

> *U intends to get A to think that p (…), he relies for this upon the Gricean mechanism [i.e. the correct working of the conditions (1)–(3) in Grice's definition], and he does not expect A to cross him with respect to his reliance on that mechanism* (Bennett 1990, p. 127).

The problem is that U's only reason to believe that A does not *"cross him with respect to his (= U's) reliance"* on the said mechanism is his assuming A to share the same (or at least nearly the same) mutual knowledge with U. But this is, as we just saw, a very unlikely assumption.

These objections do not prove Strawson's regress to be unavoidable[24], nor do they show the concept of non-linguistic communicative intention to be indefinable. But they reveal how difficult it is to formulate a satisfying definition of the communication-intentionalists' foundational concept.

3.3 Concluding remarks

In the introduction to this chapter, I announced my intention to convince you of the view that Davidson's approach to semantics is as promising as the communication-intentionalists' alternatives. I have tried to do so by showing that Davidson's approach to semantics is harder to discredit than it may appear at first sight (cf. section 3.1), and that the communication-intentionalists face their own serious problems (cf. section 3.2). Have my efforts sufficed to convince you? I think (or hope) that they make you consider my claim; I do not think, however, that I have yet convinced you. The reason is that there is still an imbalance between the two rival approaches: Grice's M-intention may capture the foundational concept of non-linguistic communicative intention only imperfectly; but having an

---

[24] This is agreed even by Stephen Schiffer who, in his later work, seems to revoke all the pro-Gricean ideas he developed at earlier stages. Cf. Schiffer 1987, p. 248f.

imperfect definition is clearly better than having no idea at all about how to explain our acquaintance with the concept of truth. Does this acquaintance, as Strawson claims, depend upon a prior acquaintance with the concept of communicative intention? Do we grasp the two concepts in parallel? Or do we grasp the concept of truth prior to the concept of communicative intention? The next chapter tries to show that these questions have to be answered in a way compatible with Davidson's truth-conditional approach. It is clear that only such a kind of explanation will remove the above mentioned imbalance between the two rival approaches. Accordingly, it is only such an explanation that may succeed in convincing you that Davidson's approach is at least as promising as the communicationalists' alternatives.

4 Davidson's understanding of the concept of truth

Is the concept of truth a suitable basis for a theory of meaning and language-understanding? The previous chapter considered this question by asking whether the concept of truth is *fundamental* enough for that role; and our provisional answer was that it is so only if it is grasped *prior to*, or *in parallel with*, the concept of communicative intention. Another, equally important aspect of the said question is, however, whether the concept of truth possesses the *right kinds of characteristics* in order to fulfil the function Davidson claims it does. This latter issue confronts us with as many doubts about Davidson's project as did the former; in fact, *all* the traditional views on truth attribute characteristics to the concept of truth which are, in one way or another, incompatible with Davidson's needs. A complete defence of Davidson's truth-conditional semantics cannot, thus, focus exclusively on the communicationalists' challenge, but has to explain also how Davidson intends to deal with the incompatibility of his approach to semantics with traditional views of the concept of truth. This is why the present chapter has a wider focus than what an answer to the communicationalists' challenge requires. It is, however, a chapter as incomplete and selective as the preceding one, for doing justice to the various traditional views on truth would go as far beyond the possibilities of this study as a complete discussion of the various communicationalists' arguments against the Davidsonian approach to semantics. This is why the sections 4.1 and 4.2 present only a simplified picture of the traditional views on truth. Davidson's own view is subsequently presented in section 4.3. Whether it suffices to dispel the doubts raised against his truth-conditional semantics and whether it allows him to meet the communication-intentionalists' challenge is discussed in the concluding section 4.4.

4.1 Is there any use for the concept of truth?

This provocative question was raised by Frank Ramsey in the course of an investigation as to what we add to a true sentence when we say that it *is* a true sentence (Ramsey 1927). Ramsey concentrated on simple sentences such as "Snow is white" and "Grass is green". The obvious way of claiming them to be true is to attach "It is true that" to them. The result of this procedure is not spectacular: "It is true that snow is white" and "It is true that grass is green" are as true as are "Snow is white" and "Grass is green". We arrive at an equally unspectacular result if we attach "It is true that" to *false* sentences, for "It is true that snow is green" and "It is true that grass is white" are as false as are "Grass is white" and "Snow is green". Attaching "It is true that" to the mentioned sentences adds, thus, *nothing* to those sentences; the mentioned prefix is, in other words, completely *redundant*. Ramsey suspected that this observation reveals a general property of the concept of truth. This is why he claimed there to be no practical use for the concept of truth.

The incompatibility of this so-called *deflationary* (or *redundancy*) *view on truth* and Davidson's approach to semantics is obvious: if the deflationist's view is correct, there is no room left for Davidson's truth-conditional approach, and if truth-conditional semantics turns out to be viable, the deflationist's view must be necessarily false. The question then is: how serious do we have to take the deflationist's view on truth?

There is at least one claim that surely deserves a closer look. It is the claim that Tarski's conception of truth, if properly interpreted, supports the deflationist's view. Such a claim has been made (in different varieties) by Putnam (1985/86), Soames (1984), Etchemendy (1988) and others. Their idea is, basically, that the absence of a *general* definition of truth – which Tarski has shown to be impossible – completely detaches the Tarskian truth-predicates from the ways languages are used (or could be used). The assigning of truth-conditions to the sentences of some L by the suitable Tarkian truth-predicate is, thus, a purely *formal act*; it has no function whatsoever besides the assignment of the truth-conditions itself. Putnam, Etchemendy and Soames try to back this claim by showing that it is possible to know Tarskian truth-predicates without knowing anything at all about the languages for which they are defined. A heavily simplified version[25] of their argument goes as follows: Consider there to be a language – a very primitive language – L_P, which exhibits no compositional devices and consists of only the following four sentences: S_1 = It is snowing.; S_2 = It is raining.; S_3 = It is sunny.; S_4 = It is cloudy. The obvious way to define L_P's Tarskian truth-predicate is as follows:

S is true-in-L_P if and only if either (S = S_1 and it is snowing) or

(S = S_2 and it is raining) or

(S = S_3 and it is sunny) or

(S = S_4 and it is cloudy).

Grasping L_P's truth-predicate thus amounts to the grasp of:

$S_1 \neq S_2 \neq S_3 \neq S_4$

Consequently, it is possible to know that:

"It is snowing" is true if and only if [either (S = S_1 and it is snowing) or (S = S_2 and it is raining) or (S = S_3 and It is sunny) or (S = S_4 and it is cloudy)],
"It is raining" is true if and only if [either (S = S_1 and it is snowing) or (S = S_2 and it is raining) or (S = S_3 and It is sunny) or (S = S_4 and it is cloudy)],
etc.

[25] It is due to Rumfitt 1995, p. 830f.

Knowing this, however, does not entail *any* knowledge about the *meanings* of L_p's sentences nor about what the speakers and interpreters of L_p could possibly intend to do by using those sentences: the sole knowledge required to grasp the Tarskian truth-predicate for L_p is the knowledge about the functioning of the logical connectors occurring in the biconditionals above (i.e. the "if and only if" and the "or"). This is why the assignment of truth-conditions to the sentences of L_p does not tell us anything about L_p's sentences themselves; it is nothing but a purely *formal act*. Soames, Etchemendy and Putnam thus conclude that Tarski's conception of truth, if properly interpreted, supports the deflationist view on truth.

Is their conclusion unavoidable? The obvious strategy to avoid it is to claim that it is not the concept of truth which is redundant, but the Tarskian conception which is inadequate. This strategy, however, is not open to Davidson, for the Tarskian conception of truth is what his approach to semantics is based upon (cf. section 2.1, p. 11ff). Davidson thus has to argue another way: he must show that the Tarskian conception does not require a deflationist interpretation. This means that he has to show there to be something more to the concept of truth than Tarski's isolated truth-predicates. The required "something more" has to be such that (a) it cannot be captured by any definition of truth, and (b) it shows the concept of truth to matter to us. While (a) saves Tarski's conception from the inadequacy charge, (b) prevents it from being given a deflationist interpretation. Is it possible for Davidson to show all this? The answer has to remain open until section 4.3, where I present his understanding of the concept of truth.

Let me ask, however, another question at this point: did Tarski regard himself as a deflationist? A look into his writings gives hardly any indication for this: Tarski's primary intention is, as he repeatedly declares, to give a formally satisfying definition of *"the intuitions which adhere to the classical Aristotelian conception of truth"* (Tarski 1944, p. 14). The Aristotelian conception of truth, in turn, reads as follows:

> *To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true* (Aristotle 1993, p. 23f).

Tarski obviously understands Aristotle in a non-deflationist way:

> *[i]f we wished to adapt (...) [the Aristotelian conception of truth] to modern philosophical terminology, we could perhaps express (...) [Aristotle's] conception by means of the (...) formula: The truth of a sentence consists in its agreement with (or correspondence to) reality* (Tarski 1944, p. 15).

Tarski thus understands the concept of truth to be a *useful* concept; i.e. as a concept that connects sentences to the world (= reality).

Does Tarski then perhaps accidentally show truth to be redundant? This seems to be what the deflationists claim; and their main piece of evidence is provided by the undisputed dramatic difference between Aristotle's and Tarski's conceptions of truth: the absence of a *general* definition of truth in Tarski's work. Tarski, however, is well aware of this difference, as the following passage from *The Semantic Conception of Truth* unmistakeably shows:

> *(...) I do not have any doubts that (...) [my] formulation does conform to the intuitive content of that of Aristotle (...). [S]ome doubts have been expressed whether (...) [my] conception does reflect the notion of truth in its common-sense and every-day usage. I clearly realize (...) that the common meaning of the word "*true*" – as that of any other word of everyday language – is to some extent vague, and that its usage more or less fluctuates. Hence (...) every solution of this problem implies necessarily a certain deviation from the practice of everyday language. In spite of all this, I happen to believe that (...) [my] conception does conform to a very considerable extent with the common-sense usage (...)* (Tarski 1944, p. 32).

Tarski's awareness of and reflections on the problem makes it rather improbable that he missed the allegedly deflationist twist to his own conception of truth. The more likely conclusion is that its deflationist interpretation is mistaken.

4.2 Is truth absolute, is it relative, or is it neither of both?

Deflationism is opposed by those who believe the concept of truth to have *some* kind of practical (i.e. non-formal) use. The members of this latter faction clearly outnumber the deflationists. It is, however, an extremely heterogeneous faction: their having a common opponent says little about their individual views about the concept of truth. Accordingly, a non-deflationist understanding of the concept does not automatically amount to an understanding compatible with the requirements of Davidson's approach to semantics.

Traditional non-deflationist views on truth might be roughly divided into two subfactions. On the one hand are those who believe the concept of truth to express what is *unconditionally* or *absolutely* true (or false); and on the other hand are those who believe it to express what is true (or false) *relative* to some criterion (or background). The former of these views (I will henceforth refer to it as the *objectivist view on truth*) entails the idea of an *objective reality*, for it is impossible for there to be absolute truths (and absolute falsities) if there were more than one reality or more than one equally correct way to describe (to analyse, to grasp) this reality. Proponents of the latter view (which I refer to as the *epistemic view on truth*) put the matter the other way round: there is, they say, no objective reality, for reality is what our sensual and intellectual abilities allow us to recognize. These epistemic abilities are, however, limited. Accordingly, it is impossible for us to exclude the possibility

that we assume there to be what there is not. Consequently, it is impossible to regard our truths (and falsities) as absolute: what we claim to be true (or false) is necessarily relative to the reach of our epistemic abilities.

Both these views are, as I said, incompatible with Davidson's needs. The *objectivist view* is incompatible because of the fact that the acceptance of absolute truths entails the acceptance of *analytic* truths. The existence of analytic truths is, however, what Quine famously denounced as an ill-founded empiricist dogma (Quine 1951). Davidson, who accepts Quine's reasoning, cannot thus allow there to be such truths. A second reason for the incompatibility is that the objectivist view allows reality to be something which is only partly about what *we* take reality to be. This is due to the fact that we are imperfect beings and cannot hope to recognize the *objective* reality in all its aspects. But if truth is about a reality that is only partly about what *we* take reality to be, the concept of truth is no suitable basis for illuminating what we believe to be true about *our* reality (i.e. about the imperfectly grasped objective reality). Consequently, it is impossible to determine the meanings of sentences that refer to *our* reality in terms of the objectivist's concept of truth.

Both these problems are inexistent with respect to the *epistemic view on truth*, for the epistemic view grants us full access to reality and, thus, to what true and false sentences happen to be about. The epistemic view is, however, as incompatible with Davidson's needs as is the objectivist view, for it allows for different varieties of *relativism*. The most extreme variety is *radical relativism*: it relativizes the concept of truth to what each singular individual believes to be true (and false). The problem with radical relativism is obvious: my epistemic abilities may diverge from yours, and yours from the abilities of a third party; and the immediate consequence of this is that what I believe to be true (or false) does not need to coincide with what you or a third party believe to be true (or false). Radical relativism has its adherents,[26] although they are rarely found among philosophers. After all, it is doubtful whether the view is tenable even in the least: to claim that the concept of truth is about what each individual believes to be true/false *for him* is to claim that each individual is the *sole* authority to judge whether such-and-such a sentence is true/false according to the respective individual's own standards. The immediate consequence of this, however, is that each individual cannot know whether she correctly or incorrectly judges such-and-such a sentence to be true/false, for there is no authority besides herself that could notice the individual's error. But saying that it is impossible for an individual to know whether such-and-such a sentence is correctly or incorrectly thought to be true/false is the same as to say that that individual possesses *no* concept of truth. The concept of "true for an individual" is thus intelligible only if it implicitly relies on a wider concept of

---

[26] The adherents of radical relativism are to be found, for instance, in pedagogics, where it often features under the label "(radical) constructivism". It is, the (radical) constructivists say, impossible to teach pupils how the world happens to be, for each individual pupil *constructs* the world in its own way. Accordingly, the truth/falsity of such-and-such a sentence depends on the particular ways such-and-such a pupil constructs the (his?) world (cf. Phillips 2000, Fosnot 2005). It is, however, not always clear whether (radical) constructivists in pedagogics really adhere to radical relativism (as their writings often strongly suggest) or whether the impression merely results from their careless formulations. (This coincides with Allen Wood's observation about the popularity of radical relativism in other fields; cf. Wood 2002, pp. 13–17.)

truth; i.e. on "true for some" or "true for all" or "absolutely true". But this is exactly what radical relativism rejects. Radical relativism is, thus, self-refuting (cf. Wood 2002, p. 3).

However, adherents of the epistemic view on truth do not need to go as far as the radical relativist. They could likewise claim that the concept of truth applies to what we all would agree to be reality under optimal conditions. One way to put this is to say that the concept of truth applies to what science claims to be reality if it continues in its progress long enough. It is obvious that this understanding avoids the problem of the radical relativist: it conceives of the concept of truth as *intersubjectively* accessible and thus guarantees a corrective for the individual who mistakenly holds such-and-such a sentence to be true (or false). This 'tamed' variety of the epistemic view on truth is, however, again as incompatible with Davidson's needs as is the radical relativist view. Why this is so is illustrated by the following thought-experiment: imagine two societies of intelligent beings, S_A and S_B, and imagine both of them to engage in scientific research. Imagine, furthermore, S_A and S_B to take their so far achieved scientific success as evidence for the assumption that their respective scientific approaches (potentially) lead to the complete knowledge intelligent beings may reasonably assume to arrive at under optimal conditions. Both these societies are thus justified in assuming their respective scientific approaches to reveal what reality (= the world) is like, and their verified hypotheses about reality to express truths about that reality. Imagine, however, S_A to maintain a *physicalist*, and S_B a *spiritualist* worldview. The obvious consequence is that S_A's reality is a completely different reality than the reality of S_B: the former is constituted by physical, the latter by spiritual objects. What the members of S_A and S_B respectively believe to be true (or false) about their realities are, thus, completely different things. Different worldviews are, of course, no problem to Davidson, for this is what his *indeterminacy of reference thesis* (cf. section 2.2, p. 25f) explicitly allows for. A problem arises, however, as soon as the existence of different worldviews entails the existence of different concepts of truth: the radical interpreter, who belongs, say, to S_A, could never arrive at an understanding of the linguistic behaviour of a member of S_B by the testing of T-Sentence-hypotheses if he could not presuppose himself and the S_B-member to share the same concept of truth. The possibility of different concepts of truth is, however, exactly what the epistemic view cannot exclude: it cannot, because excluding this requires the members of S_A and S_B to realize that their respective truths might be expressed in terms of the respective other worldview, and this in turn implies the members of S_A and S_B to realize that their concepts of truth are not bound to their respective worldviews. But that the concept of truth *is* bound to our respective worldview(s) is inherent to *every* variety of the epistemic view: it claims truth to be, as we saw, about what *we* take reality to be (cf. p. 45f). There might be, however, many "we"s; consider, for instance, tribes in remote jungles which have not yet been contacted by our civilization, or the different worldviews of our ancestors, of the generations to come, or of potentially existing intelligent beings in outer space.

The epistemic view on truth is thus as incompatible to Davidson's needs as is the objectivist view: while the objectivist view assumes the concept of truth to be about what is (partially) out of our

reach, the epistemic view assumes the concept of truth to be rather *too close* to us. Davidson must thus find a way somewhere between them: he must save the objectivist's idea of there being just *one* concept of truth while denying truth to be absolute; and must save the epistemist's idea of truth to be about what is (in principle) accessible to us while showing it to be independent of varying worldviews. He has, in other words, to show the concept of truth to be *universal* but *not absolute.* Is it possible to do so? This (among other things) is what I set out to show in the next section.

4.3 Davidson's understanding of the concept of truth

Davidson claims the concept of truth to be one of our most fundamental concepts. It plays, he says, a crucial role in our interactions with each other and with the world, for it is the concept of truth that allows us to determine what our interlocutor's (uttered) sentences mean. The present section explains Davidson's view on truth in two steps: subsection 4.3.1 describes how Davidson explains our grasp of the concept of truth, and subsection 4.3.2 illuminates what characteristics the so-understood concept has.

4.3.1 The role of the concept of truth in subject/object interactions

One thing is certain: if Davidson correctly claims the concept of truth to be foundational in linguistic subject/object interactions, then it has to be grasped *prior to,* or *in parallel with*, the other concepts prerequisite to such interactions. In other words: the concept of truth has to be grasped prior to, or in parallel with, the concept of *reference* and the concept of *communication* (or *communicative intention*). Subject/object interactions which make use of these concepts are, however, not the only kind of such interactions, for there are likewise interactions which do not entail the grasp of *any* concept at all. Consider, for instance, the case of arational beings who react to their environment: such beings – cats, dogs, and the like – do not possess the said concepts, but they obviously interact with certain objects (say, with their prey). In order to fully understand subject/object interactions and the roles the concepts play in them, it is thus important to take a step back and to ask first what is prerequisite for the existence of such interactions *in general*: do subject/object interactions require, for instance, just *one* subject, or do they require more than one?

A brief reflection reveals that there must be at least *two* subjects for there to be a subject/object interaction. The reason is that a singular and isolated subject could not judge whether the object it believes to interact with is an object *in the world* or rather an object it merely *imagines* there to be. It cannot, in other words, clearly distinguish between itself and what it is interacting with (or imagining to interact with). This amounts, however, to saying that the subject and the 'object' it interacts with are essentially the *same*. But a world where the subject and the object(s) are the same obviously has no need for interaction, for if objects are indistinguishable from the subject, there is simply nothing to interact with. Subject/object interactions thus require a second subject; i.e. an

independent instance which is able to judge whether the other subject correctly or incorrectly claims such-and-such an object to be *in the world*. Davidson takes this *triangular* situation – two subjects, one object – very serious and considers all the different types of subject/object interactions with regard to this situation.[27] He distinguishes between three types of subject/object interaction: (i) primitive interactions, which do not involve any linguistic skills, (ii) interactions where language comes into play, and (iii) interactions between subjects with fully developed linguistic skills.

Interactions of type (i) comprise (among other things) interactions where no grasp of any concept is involved. A good example of such an interaction is provided by two animals, say, two cats, chasing a prey (= the object). Typically, a cat's chasing a prey involves a certain kind of behaviour; for instance, a particular way of staring at the prey, slow and smooth movements, prudent approaching, etc. A cat's behaviour generally changes, however, if it notices[28] that its prey is also chased by another animal (i.e. by the second cat): besides the chasing-behaviour, both cats will now exhibit a behaviour that allows the conclusion that they are reacting to each other; for instance, by staring and growling at each other, or by trying to chase the respective other cat away. What the change in their behaviour suggests is that they *are aware of* each other, and that they are aware of the fact that they are chasing the very *same* prey. This, however, is the same as to say that they are aware of their *sharing (a slice of) the same intersubjectively accessible world*. Davidson claims this observation to be of utmost importance: It does not only tell us something about cats and their prey, but it tells us something about subject/object interactions *in general*; namely, that we are entitled to assume a vis-à-vis – whether a cat, a human being, or a Martian – to be aware of its sharing its world with us as soon as the behaviour of this vis-à-vis alters with respect to our behaviour toward them, and towards the mutually triangulated object. What the primitive interaction-situation teaches us, then, is that we *share the same world with whatever subject we are able to interact with* (Davidson 1997, p. 128f).

An example of an interaction-situation of type (ii) is the situation in which one of the two subjects acquires its first language. This triangular constellation comprises the following 'angles': (a) an infant who just begins to acquire his first language, (b) another person (say, his mother) from whom the infant is learning the language, and (c) any suitable object (say, a baby's bottle, a teddy bear, the infant's father). We may assume each of the two subjects to interact with the object and with each other as soon as they both react to each other and to the object – this is what we learned from the interaction-situation of type (i). Interaction thus obtains, for instance, when the infant drops the teddy bear, realizes that the mother picks it up, and reaches his hands out to his mother in order to grab the teddy again. But what does the infant have to master in order to interact with the teddy bear and its mother by making use of a language? The answer is that it has to understand that such-and-such sound

---

[27] This is the basic idea behind Davidson's *model of triangulation*. Although it is implicit in his writings as early as in Davidson 1973 and Davidson 1975, it had to await elaboration until the nineties (cf. Davidson 1992, p. 120ff; Davidson 1997, p. 128f, and Davidson 1997a, p. 138ff).

[28] My wording of the cats' *noticing* and *being aware of* each other is an attempt to avoid psychological vocabulary. I apologize for failing to do so on a few occasions: my sentences would have become too complicated and clumsy if I hadn't allowed for a few compromises.

pattern serves to interact with objects, that these sound pattern/object associations are not random, but determined associations, and that he, the infant, has to produce these sound patterns whenever he aims to interact with the associated objects.[29]

A crucial part in the infant's language-acquisition is played by his imitating the sound patterns he recognizes the mother to produce, and the reaction he receives from the mother to these imitations. The reactions might be reasonably divided into two categories: *positive* reactions and *negative* reactions. The mother is likely to react positively to the infant's imitations if she interprets them as somehow resembling the objects which are being triangulated, say, the baby's bottle, the infant's teddy bear, or the infant's father. She is, however, likely to react negatively if she interprets the infant's sound patterns as having no resemblance whatsoever with the triangulated objects.

The situation presents itself rather different from the infant's perspective: lacking any idea about reference, he cannot yet take the mother's positive (negative) reactions as a result of his having uttered the adequate (inadequate) sound pattern on such-and-such an occasion. We may assume, however, that he perceives the mother's positive reactions as in a certain way rewarding reactions, and the negative reactions as unrewarding (or not so rewarding) reactions: positive reactions are likely to be agreeable; they comprise the hearing of the mother's joyful and happy voice, the feeling of the mother's hand stroking his cheek, or even the receiving of the bottle after the utterance of something that sounds more or less similar to "bottle". Negative reactions, on the other hand, are likely to be disagreeable (or less agreeable); say, the mother's slightly harsher and perhaps disappointed voice, fewer strokes and a reduced chance of receiving bottles or teddy bears on such-and-such occasions.

The decisive question, however, is how the infant takes the step from perceiving the mother's reactions as rewarding/unrewarding to the knowledge that such-and-such a particular sound pattern refers to such-and-such an object in the world. A key role in this process is arguably played by the fact that the infant is likely to prefer positive reactions over negative ones, for people generally prefer the agreeable over the disagreeable (Davidson 1997a, p. 140). We may thus assume that the infant is inclined to maximize the rewarding reactions and to minimize the unrewarding ones. But how is the infant supposed to do so? The intuitive answer is (as already mentioned) that it has to learn to utter the right sound patterns on the right occasions; i.e. that it has to utter "teddy bear" (or "bottle", or "Papa") on those and only those occasions where it is appropriate to do so. There are, of course, different ways of learning this. The paradigmatic one takes place, however, in the triangular interaction situation, for this is where the infant is able to perceive *all* the relevant 'ingredients' to his learning process: (a) the

---

[29] This answer is heavily simplified for there are many other things the infant has to learn. Among these are, for instance, the ability to control his larynx and his vocal cords, the ability to generalize, and the ability to understand the persistence of objects. I ignore them here, for they are either irrelevant to conceptual considerations or not exclusive to the acquisition of a first language. Into the former category fall the physiological requirements and into the latter the ability to generalize and to understand the persistence of objects. Both of them are already involved in interactions of type (i): cats are aware of their prey's persistence through time, for they do not abandon their chases as soon as they loose sight of their preys; they rather relocate themselves in order to regain an undisturbed sight. Likewise are they capable to generalize: painful experiences with hedgehogs, for instance, make them avoid these species at later chases; while a successful chase of a mouse encourages further chases of other mice.

mother's uttering of a particular sound pattern (i.e. "teddy bear" "bottle", Papa"), (b) the object referred to by these sound patterns (i.e. the teddy bear, the bottle, the father), and (c) the immediate rewarding/unrewarding responses to the infant's own utterances of these patterns.[30] Assuming that the infant develops as most infants do, he will sooner or later realize that there are particular well-defined sound pattern/object associations. He will thus begin to utter such-and-such a sound pattern ("bottle", "teddy bear") whenever the suitable object is triangulated by the infant and the mother; and he will soon thereafter produce that particular sound pattern in such a way that it adopts the quality of a *communicative act*: the child will, in other words, begin to utter "bottle" or "teddy bear" when he wants his mother to perform the related rewarding response; for instance, giving him the bottle or the teddy bear. However, if the infant begins to use particular sound patterns (or let us say now: one-word sentences) in order to *communicate* such-and-such an *intention* to his mother, he has understood that those one-word sentences *refer* to the respective objects: it is his using that particular one-word sentence that allows him to communicate with his mother about the mutually triangulated object (i.e. the bottle, the teddy bear). Saying that the infant has understood "bottle" to refer to the baby's bottle and "teddy bear" to the teddy bear implies also that he understands that "bottle" does *not* refer to the teddy and "teddy bear" *not* to the bottle. We thus have to assume the infant to understand that it is an *error* to refer to the bottle by uttering "teddy bear", or to refer to the teddy bear by uttering "bottle". But to say that the infant understands that it is an error to do so is to say that he understands that it is *false* to do so, and that it is *not false* (i.e. *correct*) to act the other way round. The infant's grasp of the first correct (and incorrect) reference relations (= sound pattern/object associations) thus implies, Davidson says, the infant's *grasp of the concept of truth*, and since those reference relations are grasped through the infant's attempt to linguistically interact – i.e. to *communicate* – with his mother, we are entitled to assume that the concepts of *truth*, *reference* and *communication* (or *communicative intention*) are all acquired *in parallel* with each other (cf. Davidson 1984, p. 104f and Davidson 1997a, p. 141).

4.3.2 The characteristics of the concept of truth according to Davidson

Let us assume for a moment that we acquire the concept of truth in a way along the lines Davidson imagines the infant to do: what characteristics do we have to attribute to the concept? The most obvious characteristic is that it is a *useful* concept: it is a useful (or, in fact: an *indispensable*) concept, for it allows us to *refer* to the world (= reality). Does this show the deflationist view on truth to be wrong? No, for the deflationist can still argue that Davidson's concept – whatever else it might be – is

---

[30] This is not to say that the infant cannot learn from situations where the object referred to is not triangulated by him and his mother (or any other person). It may do so, however, only if there is evidence available to the mother that the infant aims to refer to something, rather than to merely produce sounds; for if she lacks such evidence, it is impossible for her to decide whether her reaction should be positive or negative. But without reactions that are clearly discernible as rewarding or unrewarding, it is impossible for the infant to learn anything about the reference-relation of the uttered sound pattern. This is why the triangulation-situation has to be regarded as the paradigmatic situation in first language acquisition.

not the concept of truth. This is so, the deflationist might claim, because it emerges during Davidson's considerations of the different kinds of subject/object interactions; subject/object interactions, however, are simply not the place where we should go looking for the concept of truth; we should not do so because truth is a purely formal concept and thus not to be found where interaction takes place.

However, is this a plausible manoeuvre? I think that it is not: it does not contribute anything new to the deflationist's claim, but merely repeats what we already knew; i.e. that the deflationist believes the concept of truth to have no practical use, to be redundant. But why fight so hard against the intuition that truth *is* an important concept? This is, let me remind you, not only Davidson's view; it is the common sense view of almost every reasonable human being. The deflationist's insistence on there being a difference between the *real* concept of truth and the concept which most of us intuitively take to be that concept is thus hardly founded on facts, but rather on how they want to look at those facts.

A second important characteristic of Davidson's conception of truth is that it is not *absolute* in the sense specified in section 4.2. It is not absolute, for it is the reality *perceived by the infant* (or by any other linguistically gifted being) that makes sentences true or false. The concept of truth thus relates the infant to what he takes reality to be, and not to some reality which is (at least partly) behind or beyond what is perceived. This is not to say that the infant (or other linguistically gifted beings) could *de facto* know every truth or falsity about *that* reality (i.e. *that* world), for there might be things in that world which are inaccessible to the sensual and intellectual abilities the child is going to develop during his life. But it says that these truths and falsities are truths and falsities about *his* world and, thus, *in principle* accessible to the infant (to the linguistically gifted being).

Davidson's understanding of the concept of truth is also not *relativistic* in the sense of section 4.2, for we share it with everybody who may *teach* us, or from whom we may *learn,* some belief to be erroneous. But this ability is clearly not restricted to individuals who are as close to us as is a mother to his infant, or as you are to your family members or friends. It is an ability possessed by all individuals with whom we are able to *communicate*.[31] It is, thus, a very general ability which we share with foreigners, with members of alien tribes and even with visitors from outer space (given, of course, that they all possess the concept of communicative intention). This is not to say that these exotic interlocutors need to share our worldview (say, the physicalistic worldview) – they may believe the world to be constituted by incarnations of gods, by spirits and forces or by the good and the evil. Accordingly, they may claim beliefs to be erroneous and, hence, false which are true according to *our* judgment. Such cases, however, merely indicate a different opinion about what *is* true, but they do not concern what the concept of truth *is*. It is exactly because of this that Davidson believes the concept of truth to be a *universal* concept; i.e. a concept to which we are *all* acquainted to, but which is independent of any particular worldview.

---

[31] This is due to the fact that *teaching* and *learning* happen to be activities that take place in *communicative* acts. I shall return to this point immediately.

I just said that the concept of truth – according to Davidson's understanding – allows us to *communicate* with each other; and I should add that it allows us so with respect to linguistic *as well as* non-linguistic communication. The reason is that the interaction-situation is exactly the same in both cases: whether the infant has to learn such-and-such a sound pattern to refer to such-and-such an object, or whether he has to learn a particular movement of the hand or body to do so does not change anything substantial: what he has to learn is that a particular sound pattern or a particular movement *does* refer to such-and-such an object, while it does *not* refer to any other object. The learning of non-linguistic communication thus involves the grasp of the concept of error, and thus the concept of truth, as much as the learning of linguistic communication does. This is why Davidson takes the concept of truth to be a *foundational* concept: there is, he says, *no* communication without the grasp of the concept of truth, and – as we saw at the end of subsection 4.3.1 – no grasp of the concept of truth without the grasp of the concept of communicative intention. The concept of truth and the concept of communicative intention are thus acquired *in parallel with* each other.

Let me come back now to the central question of *part I* of this study: is Davidson's truth-conditional approach to semantics capable to meet the communication-intentionalists challenge? The finding of the present section suggest that he is, for what is required to meet this challenge is that we acquire the concept of truth prior to, or at least *in parallel with*, the concept of communicative intention (cf. chapter 3.1.3, p. 34). The subject/object interaction of the language-acquiring infant just discussed suggests that this is the case: the infant acquires the two concepts *in parallel with* each other. We may thus close *part I* with the conclusion that Davidson's truth-conditional approach to semantics is capable to meet the communication-intentionalists challenge.

4.4 Is Davidson's understanding of the concept of truth plausible?

One way to answer this question is to consider whether it is possible to imagine an individual to acquire one of the mentioned concepts – truth, communicative intention, or reference – without the others. Is it, for instance, possible to claim that we acquire the concept of communicative intention without the parallel acquisition of the concept of truth? I can imagine only one way to do so: one might try to widen the understanding of the concept of communicative intention in such a way that it covers subject/object interactions at a level where the concept of truth is not yet in play. The obvious place to look for such interactions are the interaction-situations of type (i); i.e. where sentient beings become aware that they share the same (slice of the) intersubjectively accessible world (cf. p. 49).

But should we really count the interactions of these beings – say, the cats' meowing and growling at each other – as communicative interactions? This is hardly a question the communicationalists are tempted to answer affirmatively, since it immediately blurs the distinction between what is performed *intentionally* and what is not. This distinction, however, is indispensable to the communicationalists, for of course they agree that communicative 'doings' are intentional

'doings'; i.e. *acts*. The cats' meows and growls are, however, no acts, for cats are arational beings and, thus, incapable of *intending* something. The idea of widening the concept of communicative intention in the just described sense is, thus, not open to the communicationalists: instead of giving them an advantage over Davidson's truth-conditional approach, it rather undermines their own.

The concepts of communicative intention, truth and reference seem thus to be intimately linked to each other. It is impossible to grasp one of them without grasping the others at the same time:

(a) Grasp of the concept of truth presupposes grasp of the concept of reference and communicative intention, for otherwise, it is impossible to make/to avoid errors. The making/avoiding of errors is, however, what allows us to grasp the concept of truth.

(b) Grasp of the concept of communicative-intention requires, in turn, grasp of the concept of truth and of the concept of reference, for it is impossible to linguistically/non-linguistically communicate something to someone if one does not understand that such-and-such a linguistic/non-linguistic means refers to such-and-such an object but not to any other object.

(c) Finally, grasp of the concept of reference requires grasp of the concept of communicative intention, for if there is no communicative intention, there is no need for any reference and, thus, no grasp of that concept. Likewise is the concept of reference incomprehensible without grasp of the concept of truth: if you do not understand that it is correct to refer to such-and-such an object by doing so-and-so, and that it is wrong to refer to this object by doing something else, then it will be impossible for you to understand what the concept of reference actually is about.

I do not claim that Davidson's reflections on the triangular interaction-situations resolve all the conceptual mysteries surrounding the grasp of the concepts of truth, reference, and communicative intention. Neither do I claim that his model of triangulation clarifies his understanding of the concept of truth in all its details. Davidson just sketches a general idea with a few brush strokes rather than explaining its subtleties.[32] This is, I have to confess, frustrating and unsatisfactory. I nonetheless believe that his thesis about the parallel grasp of the concepts of communication, truth, and reference is hard to reject (cf. (a)–(c) above). Consequently, I believe that we are *justified* in claiming Davidson's truth-conditional approach to semantics to meet the communication-intentionalists challenge.

---

[32] Davidson's comments on the interaction-situations and his model of triangulation hardly ever extend over more than two pages. He even confesses (occasionally) that his explanations are no more than *"hasty remarks"* (Davidson 1992, p. 121) or *"some suggestions"* (Davidson 1997, p. 128).

Part II:

Davidson's foundational notion of meaning – its characteristics and explanatory power

5 Literal meaning vs. conventional meaning

Now that we have gone through a defence of Davidson's truth-conditional approach in *part I*, it is time to proceed to that approach's very core: its foundational notion of meaning (i.e. the notion of *literal meaning* [33]). What are the characteristics of this notion? And: is literal meaning really a suitable notion to describe the possibility of language-understanding? – The intuitive answer is that it is *not*, for the notion of literal meaning excludes from meaning what our everyday experience suggests is requisite to successful linguistic communication: the speaker's audience-directed communicative intentions. But is this intuitive answer correct? It probably is with regard to our everyday linguistic exchanges, for it is obvious that we rather try to grasp a speaker's communicative intentions than to discern what he might plausibly hold to be true. Davidson's interest, however, is *not* in the question of what *de facto* enables understanding, but in the question of what enables understanding *in principle*. The central question of *part II* is, thus, whether the mentioned intuition is also valid with respect to Davidson's philosophical question. In other words: is Davidson correct in claiming that it is, *in principle*, possible to interpret (and, thus, to understand) a speaker's linguistic behaviour by making use of a notion of meaning that lacks any relation to that speaker's communicative intentions?

Communicative intentions are usually said to determine linguistic meaning thanks to *conventions* and/or *rules*. The present chapter thus begins by scrutinizing the question just mentioned by considering whether it is possible to interpret a speaker's linguistic behaviour without making use of any notion of conventional and/or rule-governed meaning. Chapter 6 subsequently completes the picture by considering the same question with respect to notions of meaning which are determined by the speaker's communicative intentions in non-conventional and/or non-rule governed ways.

On the view that linguistic meaning is the result of conventions/rules between communicative intentions and particular patterns of soundwaves (or marks), our understanding of them is the result of the speaker's conforming to the relevant conventions/rules. This is why the first section of the present chapter (section 5.1) begins by asking what it might plausibly mean to *follow rules* and/or *conventions* in linguistic contexts. Section 5.2 then turns to the most general linguistic conventions/rules which are said by some to govern the understanding of language: the conventions relating an uttered sentence's *illocutionary force* to that sentence's *grammatical mood*. Section 5.3 is a short interlude that explains how Davidson proposes to deal with the grammatical moods – an issue which remains unsolved in section 5.2. Section 5.4 comes back to the conventions/rules and considers whether there are indispensable conventions/rules relating the speaker's communicative intentions to the meanings of *sentences* and to the meanings of *subsentential items* (i.e. words, phrases). The concluding section 5.5

---

[33] I already mentioned (cf. p. 5) that Davidson, in the mid-eighties, replaced the notion of literal meaning by the notion of first meaning. Since it is not before the next chapter that I explain the reasons for this replacement, I continue to use the name "literal meaning" in the present chapter. As will become clear in the next chapter, is the said replacement without any relevance whatsoever to the here discussed issues.

summarizes our findings about the characteristics and the explanatory power of Davidson's notion of literal meaning.

5.1 Rule-/convention-following in language: the analogy of language-understanding and the playing/winning of games

I already mentioned that the speaker's communicative intentions are usually said to determine linguistic meaning thanks to *conventions* and/or *rules*. But is it not an oversimplified picture not to distinguish between *conventional meaning* and *rule-governed meaning*? Aren't conventions something completely different than rules? Consider, for instance, the following list of often mentioned distinguishing criteria:

(1) A regularity constitutes a rule iff it is possible to state that regularity explicitly, whereas no such requirement exists for the constitution of a convention.

(2) A regularity constitutes a rule iff its violation inevitably leads to an immediate end of the activity it governs, whereas no such automatism exists with activites governed by conventions.

(3) A regularity constitutes a rule iff the existence of some activity depends upon its being governed by *this* very rule (or *this* very set of rules), whereas some activity may continue to exist when its governing (set of) convention(s) is substituted by some sufficiently similar other (set of) convention(s).

(4) A rule distinguishes itself from a convention in so far as the replacement of a particular rule by an equally well suited other rule does not need to change the rule-follower's behaviour, while the replacement of a convention by an equally well suited other convention always changes the convention-follower's behaviour.

The difference between rules and conventions is undisputed. But philosophers tend to ignore this as soon as it comes to issues surrounding meaning and language-understanding. Ludwig Wittgenstein, for instance, explicitly states that in language use, rule-following has to be understood as a specific *practical ability*, i.e. as a certain kind of *knowing-how*, which might be performed without any awareness at all. *"We follow a rule blindly"* (Wittgenstein 1952/2009, § 219), he claims, and if he considers this to be a general characteristic of what it is to follow a rule, it is difficult to see why he is not violating the first of the above mentioned distinguishing criteria. The same holds for many other authors: Michael Dummett (1978a), for instance, uses "rule" and "convention" interchangeably while discussing the relations between the illocutionary forces and the grammatical moods (cf. section 5.2). Peter Strawson, in *Meaning and Truth* (Strawson 1969), does the same in the course of making his case that the concept of linguistic meaning presupposes the concept of non-linguistic communicative

intention (cf. subsections 3.1.2 and 3.2.1). And finally, there is Davidson: he does not care about the differences between rules and conventions, for he believes both of them to be equally superfluous to the description of the conceptual preconditions of language-understanding (cf. Davidson 1984). Distinguishing between conventions and rules is, thus, an ill-advised idea with respect to our debate: it rather confuses the issue than clarifies anything.

What does it mean, then, to say that language-understanding (or more general: language use) depends on rule/convention-following? A remarkably frequent answer is that following rules/conventions in language is somehow similar to following rules/conventions in *the playing and winning of games*. The comparison is generally understood as an *analogy*. It needs, however, not to be taken in this way, but could be understood also in a more *literal* sense. The difference between the two interpretations is important, since the *literal view* does not offer the convenient elbow-room the *analogy view* provides: while the *literal view* regards rule-following in language as *the same* as the following of rules in games, the *analogy view* merely claims there to be a more or less pronounced similarity between the two activities.

Davidson (1992, p. 113f) suggests that the late Wittgenstein adhered to the *literal view*; and there are at least some passages in the *Philosophical Investigations* which seem to support this. Consider, for instance, the following lines:

> *I send someone shopping. I give him a slip marked "five red apples". He takes the slip to the shopkeeper, who opens the drawer marked "apples"; then he looks up the word "red" in a table and finds a colour sample opposite it; then he says the series of cardinal numbers – I assume that he knows them by heart – up to the word "five" and for each number he takes an apple of the same colour as the sample out the drawer* (Wittgenstein 1952/2009, § 1).

The passage suggest that Wittgenstein took the *literal view* because he describes the understanding of "five red apples" as an activity as rigid and mechanical as, for instance, the adding or subtracting of numbers. But Davidson's judgment is hardly fair, for there are many passages in Wittgenstein's work which suggest that he actually rather tended towards the *analogy view*. And even the above quoted passage is not absolutely supportive to the *literal view*, since Wittgenstein immediately adds to the quoted sentences that *"it is in this* and similar *ways that one operates with words"*.

But let us ask: is the *literal view*, independently of what Wittgenstein thought, a plausible view? It is hardly possible to give an affirmative answer, since language use seems to be far more complex than the comparing of colour samples or the adding and subtracting of numbers. There is, for instance, just *one* way to add or subtract two numbers, but there are *many* ways to assert something: that Joan is wearing her purple hat again might be asserted, for instance, by the utterance of the suitable sentence in the indicative mood (i.e. "Joan is wearing here purple hat again."), but it might be asserted just as well by the speaker's making use of the interrogative mood (= "Did you notice that

Joan is wearing here purple hat again?") or the imperative mood (= "Notice that Joan is wearing here purple hat again!").[34] The *literal view* is, thus, too crude to mirror the ways language is used and understood by competent speaker-interpreters.

The 'liberal' counterpart, the *analogy view*, lacks the just mentioned shortcoming, for its operating with analogies allows adapting it to whatever our language uses require it to conform to. This strength, however, is also a potential weakness, for analogies always rest on similarity relations. But how similar do two analogues have to be in order to reasonably count as such? A general answer to this question is difficult to give; but it is clear that there are *some* obvious similarities between language use and the activity to which it is often said to be analogous; i.e. the playing and winning of games: both activities are social activities, they both involve the achievement of well-defined practical ends, and they both presuppose the possession of a certain degree of means-end rationality. But is there also a similarity with respect to the analogy's core; that is, with regard to the two activities' *rules/conventions*?

A closer look reveals that there are at least two importance *differences*. The first one concerns the *authority* of the respective rules: a game's rules (say, the rules of chess) determine whether a player is *allowed* to make a move and whether he is *obliged* to do so. He is *allowed* to if the last move was made by the player's adversary, and he is not if the last move was made by himself. At the same time, he is *obliged* to make a move if it is his turn to do so, for if he refuses to comply with the game's rules, he immediately provokes an end or a standstill to the game. Nothing parallel can be observed, however, on the part of the rules of language: there is, firstly, the fact that in most cases, we cannot discern anything that influences our free decision to utter something or to keep quiet. There is, thus, in most cases nothing like an authority that allows or obliges moves in linguistic games. And even in cases where there *is* such an authority, it is *not* the authority of our language's rules, but rather the authority of the other *player(s)* which allows/obliges us to make or abstain from a move: if you are asked by the police officer whether you disregarded the red traffic light, it is not the asking of a question that forces you to answer, but the authority of the speaker *qua police officer*. A related difference between the two types of rules concerns the fact that in language use, there is not always a possibility to 'end a game' by means of refusing to comply with the rules. Consider, for instance, the question "Have you been unfaithful?": if you have *not*, it is irrelevant whether you answer that question or not, for your refusal to answer is very likely to be understood as a confession. There is, thus, no possibility here to 'end the game' by means of ignoring its rules.

A second major difference concerns the fact that it is, in principle, *always* an option to learn a game's rules by studying the suitable rule manual, while this is *not* always an option with respect to the learning of a language. This fact implies a difference in the nature of the respective rules: it allows, in the case of games, to strictly distinguish between the rules and their application; or rather, between the rules and the respective empirical circumstances that accompany their application. Sure: a chess

---

[34] The example is from Davidson 1979, p. 110.

player needs to know some empirical facts about what is going on during a chess match, for otherwise, he does not know whether it is his or his opponent's turn to make a move. But this knowledge can be clearly separated from the game's rules; it is something that plays a role *in addition* to the rules. Consequently, it is possible to explain a player's making a move at t_1 and his abstaining from doing so at t_2 by pointing at his following the game's rules and his *additional* knowledge that such-and-such circumstances happen to obtain.

The situation looks different with the learning of a language, for there are cases where it is impossible to sharply distinguish between a language's rules and the empirical circumstances that accompany their application. This is particularly so with respect to the situation of *first* language acquisition, where the learner is required to grasp, say, that "teddy bear" refers to the teddy bear, "bottle" to the bottle, and "Papa" to the father. These 'reference rules' are acquired, as explained in subsection 4.3.1, thanks to the learner's observing a second subject applying them whenever the suitable objects (= the teddy bear, the bottle, the father) are salient to both of them. It is, then, impossible for the language learner to sharply distinguish between the rules and the empirical circumstances that accompany their application. From the perspective of the language learner, empirical circumstances are thus nothing *additional* to the rules, but rather *an integral part of* them.

Do these differences suffice to shake confidence in the *analogy view*? Hardly so, for claiming language use and the playing/winning of games to be *analogous* is not to claim them to be the *same*; it only amounts to claiming them to be *similar* in relevant respects. However, if the *analogy view* really captures what is essential to language use, there have to be *some* aspects where the analogy unconditionally holds. The parallels needed hardly concern the activities' *rules* themselves, for we just saw them to be different in at least two fundamental respects. The parallels might be found, however, in the ways those dissimilar rules happen to *govern* the respective activities. I think that at least the following two characteristics of the playing/winning of games have a pronounced analogue on the part of the use of language:

(1) People who play a game usually intend to win.
(2) The question "What counts as winning in a game?" cannot arise if all players agree on the rules of the game.

The language's analogue with respect to (1) is, I think, that a speaker, by carrying out such-and-such a speech act, typically intends to arrive at a specific end. He intends, for instance, to receive a suitable answer to his question, or to see a specific action being executed as a result of his issuing a particular command. So why not count the speaker's intentions to arrive at these ends as the analogue to the player's intention to win? Regarding the matter from this perspective obviously involves the idea that the acquaintance with a language's rules allows the playing and winning of *many* linguistic games. This is doubtless an important difference to the rules of non-linguistic games, which typically enable the playing and winning of *one* particular game only. This difference is, however, irrelevant with

respect to (1), for (1) does not claim an analogy between the *number* of games a set of rules allows to play, but an analogy between the *ends* of those games.

Assuming that you accept what I claim to be the linguistic analogue to (1), the analogue to (2) is that a speaker and his interpreter hardly ever disagree about the particular end of such-and-such a linguistic game: they typically both know that the asking of a question requires a suitable answer, the issuing of a command the execution of a particular action, etc. Questions such as "Did you mean so-and-so as a question?" or "Is so-and-so a command?" do not undermine the analogy, for they do not indicate a disagreement with respect to the *end* of a particular linguistic game, but an uncertainty whether such-and-such a speech act is part of the playing of *this* or rather of *that* particular linguistic game.

We have finally found, then, at least two aspects of language use which are unconditionally analogous to the playing/winning of games. Can we thus conclude that the game analogy really holds? I think that we should, for the two mentioned strict analogues are not the only existing similarities. There are also the pronounced similarities with respect to the compared activities' relatively *general* characteristics (i.e. their being both essentially social, involving practical ends, presupposing the possession of a certain degree of means-end rationality, etc.) and the far less pronounced similarities with respect to language's 'rules'. Following rules/conventions in language, we may therefore conclude, is, in some rather loose ways, like the playing and winning of games.

However, what does this finding tell us with regard to Davidson's application of truth-conditional semantics to natural language? And what does it tell us about his notion of literal meaning? Not much yet, for it only tells us that it is our *everyday language use* which resembles in more or less important ways the playing and winning of games. Language's rules/conventions may thus be said to *de facto* govern our linguistic activities, but whether they do so *in principle* is still an open question. It is therefore time to come to a close with our general reflections and to scrutinize whether there are any *particular* kinds of rules/conventions which are, *in principle*, indispensable to the use and understanding of a language. The next section begins to do so with the consideration of a very general kind of linguistic conventions/rules. More specific ones will be considered later on in section 5.4.

5.2 The Dummett/Davidson controversy about conventions relating illocutionary forces to the grammatical moods

Mutual obedience to the same linguistic conventions/rules obviously results in *standardized* language use. It is, thus, intuitive to say that obedience to the same linguistic conventions/rules results in the *sharing of a language*. But which of the many linguistic conventions/rules are indispensable to make speaker-interpreters share a language? Michael Dummett says that conventions relating the *illocutionary forces* of uttered sentences with these sentences' *grammatical moods* are particularly important. This is, he continues, unfortunate for Davidson's theory of Radical Interpretation, for there

are good reasons to believe that force/mood conventions are indiscernible to the radical interpreter (cf. subsection 5.2.1). Davidson agrees that they are indiscernible, but they are so, he claims, due to their inexistence (cf. subsection 5.2.2). Who is right? Dummett does not accept Davidson's objection but claims him to have misunderstood his view. Subsection 5.2.3 explains why Dummett is probably right but why this has no relevance for the radical interpreter's interpretative task, and hence for Davidson's theory of Radical Interpretation.

5.2.1 Dummett's argument in favour of the indispensability of force/mood conventions

Force/mood conventions are indispensable for language-understanding, Dummett claims, and he backs this conviction in at least three different ways. He points, firstly, to the obvious fact that there are strong regularities between the expression of forces and the use of moods: commands, for instance, are typically uttered in the imperative mood, questions in the optative mood and assertions in the indicative mood. These regularities allow for exceptions (cf. section 5.1, p. 58), but they are very strong and very general – a fact which Dummett regards as an indication for their importance in language use. Dummett mentions, secondly, the criterion which Frege set out to judge the adequacy of theories about the semantics of natural languages. Such theories are adequate, Frege says, iff they explain (or describe) what a language's sentences *refer* to, what particular *sense* they have and what the speaker adds to those sentences while uttering them with such-and-such a *force* (cf. Frege 1918, p. 34ff). Frege also suggests that it might be the moods that mirror force on the sentential side, for moods are doubtless a sentential feature, but have no visible importance whatsoever to the determination of the sentence's reference and sense.

Dummett's third point rests on a thought-experiment. Unsurprisingly, it is a thought-experiment concerning the analogy between language-understanding and the playing/winning of games. Dummett considers, more precisely, the situation of a Martian who visits our planet and who is eager to learn what he experiences humans to engage in (Dummett 1978a, p. 94f). Dummett imagines the Martian first to try to learn a game; namely, chess. He asks how the Martian might plausibly proceed. There is, Dummett observes, just one possible answer: the Martian has to learn the game by means of a laborious process of *learning-by-doing*. He has, in other words, nothing to begin with than the observations of the chess players' behaviours and his faculty to draw inductive conclusions from these observations. Given enough perseverance and intelligence on the side of the Martian, his efforts lead, Dummett continues, to the development of a theory, which may be powerful enough to allow him to engage in the playing of chess. However, should we really regard the Martian's performances as a *playing* of chess? Dummett's answer is negative: we should not regard the Martian to be playing chess, he says, for the Martian's chess theory is built on facts that reveal only the *application* of the game's rules, but do not tell him anything about the rules themselves. The Martian's theory is, thus, only telling him how moves in chess are executed and how they are not, but it does not tell him what distinguishes a *lawful* move from an *unlawful* move. It is, thus, impossible to the Martian to

understand what it is to *win* and to *lose* in chess, for *this* is what the game's rules determine. Dummett thus concludes that the Martian's chess theory is necessarily incomplete: it does not tell him what is *essential* to the game; it does not tell him why people engage in the playing of chess; i.e. what gives their playing *"its point"* (Dummett 1978a, 94f).

The situation is essentially the same, Dummett continues, if we imagine the Martian to construct a theory about his hosts' linguistic behaviour: he proceeds along the same lines as he does in the case of learning chess; i.e. he observes the relevant human behaviour and tries to draw the right inductive conclusions from it. Given favourable circumstances, this procedure will again lead to a theory that allows him to engage in communicative exchanges and to produce, in such-and-such situations, appropriate sounds. But again, the Martian's sound productions cannot count as linguistic acts, Dummett claims, for he has no idea about the newly acquired language's rules (or conventions, as Dummett prefers to say). The situation is analogous to the chess case: the Martian lacks an understanding of the reasons and motives (= *"the point"*) of a speaker's making an utterance; he does not know, in other words, what the speaker *is doing*; i.e. whether he happens to command something, or whether he rather states, or promises, or asks something (Dummett 1994a, p. 91). Lacking knowledge about these conventions thus makes the Martian's sound productions as much *imitations* of linguistic acts as his moving of chess pieces is an imitation of playing chess.

The linguistic conventions which the Martian misses, according to Dummett, are the conventions that govern the use of the illocutionary forces. He has good reason to claim this, for it is thanks to the forces that the utterance of a particular sentence counts, on one occasion, as an assertion, on another as a command, and on yet another as a question. Dummett has, furthermore, good reasons to assume that the *forces*, but likewise the *moods*, and even more the *force/mood relations*, should raise considerable difficulties for Davidson's truth-conditional approach. The problem with respect to the *forces* is that the speaker's making use of them is driven by *communicative intentions*; namely, by the intention to indicate to the interpreters that such-and-such a sentence is uttered as an assertion, command, question, etc. Accepting force as a part of semantics implies, thus, to accept communicative intentions *in* semantics – and this is, as we know from chapter 2, an absolute *no-go* within Davidson's approach. The *moods*, on the other hand, are a potential source of trouble to Davidson's truth-conditional semantics in that some of them – the *non-indicative* ones – allow for sentences that apparently lack truth-conditions and, thus, a Davidsonian literal meaning. (We will come back to this problem in section 5.3.) Dummett observes, however, that there is, with regard to both topics, one exception each: with regard to force, the *assertive force*, and with regard to the moods, the *indicative mood*. Assertive force is compatible with Davidson's truth-conditional approach, Dummett observes, since asserting a sentence is holding that sentence to be true, and the holding true of a sentence is just what Davidson's radical interpreter relies on. The indicative mood, on the other hand, presents an exception to Davidson's 'mood problem' in so far as indicative sentences *usually* possess truth-conditions and, thus, Davidsonian literal meanings.

However, do these exceptions save Davidson's radical interpreter from the Martian's fate? Dummett claims that they do not: the problem lies, he says, in Davidson's acceptance of Tarski's thesis that it is impossible to define a general definition of truth (i.e. "is true" for variable L) (cf. section 4.1). Lacking such a general definition, the radical interpreter has no other choice than to interpret the alien speaker on the basis of the assumption that the Tarski-style truth-predicate suitable to his own language is also suitable to the language of the alien speaker. This works, as Dummett concedes, perfectly well with respect to the determination of the senses and references of the alien uttered sentences. It necessarily fails, however, with respect to the assertive force/indicative mood convention. The reason is not, he emphasizes, that this convention has no connection to the concept of truth, for he explicitly claims there to be such a connection (cf. subsection 5.2.2); but the connection, if seen from the Davidson/Tarski perspective, is necessarily a connection with respect to the individual Tarski-style truth-predicates for particular languages. The immediate consequence is, Dummett continues, that the only thing which the interpreter may learn about assertive force is that the alien utterances – if correctly assumed to be uttered in the indicative mood – happen to be assertions *according to the conventions of the interpreter's own language*. He does not learn, however, anything about what it is to express assertive force according to the conventions of *the alien language*. The problem is insuperable, for it would require a general definition of truth, and this is, as just mentioned, what Tarski and Davidson explicitly deny there to be. Davidson's radical interpreter is, then, in the very same situation as the Martian in the thought-experiment summarized above: he may arrive at perfect *imitations* of the alien's language uses, but he may never understand the reasons and motives (= *"the point"*) of those uses. Dummett concludes that Davidson's truth-conditional approach to semantics fails to meet Frege's adequacy criterion: it fails to tell us what the moods contribute to a speaker's utterance of a sentence.

5.2.2 Davidson's objection: the existence of force/mood conventions is impossible

Davidson denies, of course, that the just presented argument puts the matter the right way, and his strategy against Dummett is essentially a strategy of turning the tables: he claims that Dummett's force/mood conventions are unavailable and that therefore the insistence on their importance must be an error; an error which shows that in the end, it is not him, but rather Dummett who fails to meet the Fregean adequacy criterion.

Davidson's objection concentrates on Dummett's formulation of the already mentioned assertive force/indicative mood convention. This formulation reads as follows:

> *The utterance of a [indicative] sentence serves not only to express a thought, and to refer to a truth-value, but also to assert something, namely that the thought expressed is true (...). (...) there is a general convention whereby the utterance of a [indicative] sentence, except in*

special contexts, is understood as being carried out with the intention of uttering a true sentence (Dummett 1974, p. 298).

Davidson interprets Dummett to maintain here that there is, firstly, a conventional relation between the utterance of a sentence in the indicative mood and the speaker's intention to use it assertorically (except in special contexts) and, secondly, a conceptional (and maybe conventional) relation between an assertion and the intention to say what is true (Davidson 1979, p. 112f; Davidson 1984, p. 266). Davidson focuses mainly on the first of the two relations. He notices a first problem with respect to the fact that it is difficult to say where the said relation is supposed to manifest itself. Contexts of utterances where indicative sentences are understood as assertions are obviously the place to look at, but the mere fact that such contexts show a striking regularity between the utterance of indicative sentences and their being understood as assertions is too unspecific to show there to be any of the searched conventions. There is, Davidson says, an equally striking regularity between the observations of horses and their having four legs; but nobody would ever honestly claim the horses' having four legs to be a convention (Davidson 1984, p. 269). Saying that the alleged conventional relation between assertive force and indicative mood manifests itself in suitable contexts of utterance is, thus, too unspecific to be truly informative.

The required specification might be claimed to consist in the speaker's intention to indicate to his interpreter that a sentence, if uttered on such-and-such an occasion, should be interpreted as an assertion. The assertive force/indicative mood convention might then be said to exist if it is possible to show that it is exactly *this* intention which is governed by the said convention. Davidson argues, however, that this manoeuvre cannot lead to the desired result. Consider, he says, the case of an actor who is supposed to play the role of a figure who warns the audience from a fire: the most convincing way to do so is obviously to scream as persuasively as possible "There is a fire!" or "Hey, I mean it! There *really* is a fire!". Now assume that at the same time, there occurs an accident which sets the stage on fire. What further means does the actor then have *in addition* to the ones already used to warn the audience that *this* time the fire *is* a fire? Obviously none, Davidson says, for otherwise he would have used them already before so as to make his acting more persuasive and realistic. The same is true, Davidson holds, with Dummettian mood/force conventions, with the Fregean *assertion sign* (⊢), and with whatever further conventional devices we may imagine to indicate a particular sentence to be an assertoric speech act. Davidson's objection is, thus, that conventional force indicators may always become a victim of a speaker's further purposes. Such a purpose might be, as the example shows, the actor's intention to act as persuasively as he can, but it could just as well be the everyday speaker's intention to pretend a communicative aim which he does not really have. Davidson thus concludes that a speaker's overt communicative intentions cannot serve as reliable force indicators; they are, he says, ill-suited to anchor the conventions in question (Davidson 1984, p. 270).

Is there a possibility to avoid this conclusion? The only strategy open to the conventionalist is, Davidson continues, to restrict the application of the said conventions to those cases where speakers are *sincere*. Such a sincerity constraint would have to consist, however, in a further convention; i.e. in a convention that restricts the application of the force/mood conventions to those cases where speakers are sincere. Such a *convention of sincerity* cannot exist, however, for there is nothing that could prevent it from being undermined in the same ways as it is possible to undermine the force/mood conventions (Davidson 1984, p. 270). Davidson thus concludes that the Dummettian force/mood conventions fail to explain what they are designed to explain: they cannot explain the forces' contributions to the 'total meaning' of an uttered sentence, for they are too weak to oblige a speaker to do what they ask him to do. Dummett's insisting on the importance of those conventions leads, thus, inevitably to his failure to conform to the requirements of the Fregean adequacy criterion for theories about our languages' semantics. It is, thus, rather Dummett than himself who is guilty of delivering an insufficient approach to the preconditions of language-understanding.

5.2.3 Is Davidson's objection the result of a misunderstanding?

Dummett does not accept Davidson's objection and argues that Davidson misunderstands his aim. He explains that in order to understand his view, we have to distinguish between the following two questions:

Question 1:   *What is it to attach each of the distinguishable kinds (= types) of force to an utterance? (...)*

Question 2:   *Under what conditions is it correct to say that someone has made an utterance with a specific type of force (...)? (...) How do we recognize an utterance as carrying a certain force (...)?* (Dummett 1993a, p. 202)

Asking *Question 1*, Dummett says, is to ask in a very *general* way what the significance is of attaching a certain force to certain utterances, or, as Dummett puts it for the case of assertion, *"(...) what it is to make an assertion, or to take someone else as having made one (...)"* (Dummett 1993a, p. 202). To ask *Question 2,* on the other hand, is to ask with respect to *particular cases of utterance* how an interpreter can recognize a speaker as having made a speech act of such-and-such a type. Dummett claims that Davidson has misinterpreted his argument because he took him to discuss *Question 2* while he has never been interested in *that* question. What he was up to is, Dummett says, to give an answer to *Question 1*.

Dummett concedes that Davidson is right with respect to *Question 2*. He concedes, in particular, that it is enormously difficult to specify the conditions under which we recognize a speaker as having made an utterance with such-and-such a specific force, and that conventions alone may not suffice to allow us to do so. He does not accept, however, the actor example as conclusive. It is, he

says, insufficient for Davidson's argument, for the cases of acted speech acts cannot be treated as speech acts in 'real' life. Unlike Frege (1918, p. 36), Dummett does not claim acted speech acts to have no illocutionary force. He rather argues that in acting, we have to assume there to be a further *special convention* which applies *"(…) to all that happens on stage from the first to the final curtain"*. This special convention is such, Dummett continues, that it does not cancel the usual assertoric force an utterance would have off-stage, but that it transposes it in such a way that the audience is able to understand that the actor *"is not making an assertion, but acting the making of one"* (Dummett 1993a, p. 212). It is, however, unclear what Dummett believes the merit of this special convention to be, for it is obviously based on the assumption that there exists a sharp and reliable distinction between acted speech acts and real ones. But there is no such distinction: think, for instance, of all those situations where an actor is asked to play himself on stage. Hollywood has made use of this constellation for many funny scenes, and the German stage director Christoph Schlingensief provoked a major scandal as he replaced professional actors by neo-nazis in his *Hamlet* production at Zürich's Gessnerallee. The scandal was obviously due to the fact that he let it open whether his 'actors' were playing their figures or whether he allowed them to spread their views in one of the country's most distinguished cultural institutions. Now, if there is no sharp distinction between acted and real speech acts, as Hollywood and Schlingensief's *Hamlet* production suggest, it is completely unclear how Dummett's special convention could serve to distinguish one from the other.

Dummett's interest is, however, primarily with *Question 1* and not with *Question 2*. Let us turn then to *Question 1*. This question is, as mentioned above, more general than *Question 2* insofar as it asks what it is to attach each of the distinguishable types of force to such-and-such a sentence *in general*. The "in general" has, of course, to be understood as "in general *with respect to a certain language*", for Dummett claims, as subsection 5.2.1 revealed, that there is a specific set of such conventions for each particular language. But is Dummett correct in claiming that Davidson does *not* address *Question 1*? The answer is: obviously yes, for it is *impossible* for Davidson to do otherwise. It is impossible, since the radical interpreter's 'interpretative equipment' only allows him to determine the 'rules' which determine the references and senses of the sentences of the alien language; they do not allow him, however, to determine any linguistic rules that might be said to exist in addition to those 'rules'. However, is this fact really as dramatic as Dummett regards it to be?

I do not think so, and my reason is quite simple: Dummett concedes, as we saw, that Davidson's radical interpreter may, if ingenious enough, arrive at the determination of the senses and references of the sentences of the alien language. Doing so requires him to discern regularities in the alien linguistic behaviour (cf. section 2.2, p. 19f). But if it is possible for the radical interpreter to discern the regularities required to determine the senses and references, it is hardly plausible that this should be impossible with respect to the regularities that determine how the alien attaches such-and-such a particular force to such-and-such an uttered sentence. To be sure, the force-regularities are undetectable to the radical interpreter's *testing of T-Sentence hypotheses* (this is what Dummett's

argument shows). But this does not imply that these regularities are undetectable to the radical interpreter *as such*. This is particularly obvious if the forces are said to manifest themselves in the grammatical moods, for moods are sentential features and, thus, intersubjectively accessible. There is, thus, no reason to assume that the radical interpreter should not be capable to recognize these regularities.

The saliency of these force-determining regularities should not make us forget, however, that they may occasionally be flouted, for it is obvious that all of them are *conventional* means and, thus, potential victims of a speaker's insincerity. The interpreter's awareness of these regularities is, thus, no *sufficient* device for him to determine the force of the speaker's uttered sentences. What the interpreter – whether radical or not – is required to possess in addition is *intuition*, *luck*, and whatever further *non-linguistic knowledge* may help him to push his interpretation in the right direction. But saying this amounts to claiming that the determination of an utterance's particular force cannot rely on any *general* reflections (i.e. upon answers to *Question 1*), but has to be considered for each singular case anew. Dummett's *Question 1* is, thus, not as important to the radical interpreter as Dummett claims it to be: *Question 1* informs him only about general rules/conventions in the linguistic behaviour of the alien community, but remains mute with respect to the question whether the alien speaker, while making such-and-such an utterance, is conforming to one of these rules/conventions or not. This shows *Question 1* to be irrelevant to the radical interpreter's needs and, thus, to be irrelevant to Davidson's theory of Radical Interpretation. Accordingly, it is irrelevant whether Davidson does or does not misunderstand Dummett's argument in favour of the force/mood conventions.

Let me close this section with the answer to the following question: why do Dummett and Davidson consider force-indicators from two so entirely different points of view? The reason has been briefly mentioned at the beginning of section 5.2: Dummett believes that language-understanding presupposes the sharing of a language, and that the sharing of a language is ensured by the sharing of that language's constitutive conventions/rules. *Question 1* is, thus, the perfectly correct question to ask from his perspective, for it asks about a *general* rule/convention within a particular *language*. *Question 2*, on the other hand, is irrelevant for him, since it cannot unveil any general rule/convention, but only far less reliable *regularities*. The situation presents itself completely different from Davidson's point of view: *Question 1* is, as we just saw, irrelevant for him, for it remains on a too general level to be useful to the radical interpreter's task. *Question 2,* on the other hand, is important to Davidson, for it is *this* question which poses itself to the radical interpreter.

In view of these different perspectives, it is no surprise that the Dummett/Davidson controversy about force/mood conventions found a sequel when Davidson announced his disbelief in the philosophical importance of the notion of a *shared language* (Davidson 1986). But before turning to this issue in the subsequent chapter, there are two other issues to be addressed first. One of them concerns the fact that conventions relating forces and moods are not the only conventions which might be said to be constitutive for language-understanding. This is why section 5.4 considers whether

language-understanding presupposes knowledge of conventions which relate the speaker's communicative intentions with the meanings of sentences, phrases, and words. The other issue concerns the moods: the present section has shown that they are no reliable force indicators, but it has not shown how Davidson removes the danger which the non-indicative moods pose to his truth-conditional approach. The next section explains how he does.

5.3 Interlude: The paratactic analysis of sentences in the non-indicative moods

In order to show the non-indicative moods to do no harm to Davidsonian semantics, one must show that, contrary to appearance, it is possible to analyze sentences in these moods as having truth-conditions and, thus, as having clearly determined literal meanings. There are, in principle, two paths leading to this result. The first one is to claim that the non-indicative moods are reducible to the indicative mood and are thus of no semantic relevance whatsoever. The second path abstains from such reductive temptations: it concedes that the moods add something to a sentence's meaning, but claims that this does not make it impossible to analyze them as having truth-conditions. Davidson opts for the second path: we cannot, he says, simply reduce the non-indicative moods to the indicative mood, since speakers do not randomly choose the moods; in general, they have some *reason* why they are using, say, the imperative mood at t_1, but the optative mood at t_2. Pursuing the reductive strategy annihilates, however, the relations between the moods and their uses by speakers.[35] The reductive path thus ignores, rather than explains, the speakers' reasons for using the moods (Davidson 1979, p. 114f).

According to Davidson, an adequate approach to the grammatical moods has to fulfil the following three conditions:

(0) It has to show that a sentence uttered in the indicative mood (say, "The moon is full.") shares a *"common element"* with its uses in the non-indicative moods ("Is the moon full?"; "I wish the moon were full.", etc.).

(1) It has to *"assign an element of meaning"* to utterances of a particular sentence in a given mood that is absent in utterances of the same sentence in another mood.

(2) It has to be *"semantically tractable"*; i.e. it has to be compatible with Davidson's truth-conditional approach (Davidson 1979, p. 115f).

How does Davidson aim to fulfil these conditions? His strategy is to apply his so-called *paratactic analysis*. This analysis (which he originally developed in order to analyze quotations (Davidson 1979a) and utterances in indirect speech (Davidson 1968)) treats non-indicative sentences as involving *two* sentences in parallel: one sentence that states the content of the non-indicative sentence in the indicative mood, and another sentence that states the particular illocutionary force with which the non-indicative sentence is uttered by the speaker. Consider, as an example, Peter's command:

---

[35] For an elaboration of the reductive path see Lewis 1973.

(1) Put off your shoes!

Paratactically analyzed, (1) turns into:

(2) Peter commands this. You put off your shoes.

The second sentence in (2) states the content of (1) in the indicative mood. It is therefore called the *content sentence*. The first sentence in (2), on the other hand, states the particular illocutionary force with which the content sentence is uttered. Because illocutionary forces are typically uttered by the use of the suitable mood, this sentence is called the *mood-setter*. Both, the mood-setter as well as the content sentence, are in the indicative mood. It is, thus, possible to assign truth-conditions to both of them. This is why the paratactic treatment shows that contrary to appearance, it is possible to analyze non-indicative sentences as having truth-conditions, and, therefore, as having Davidsonian literal meanings.

So far, so good; but how convincing is this analysis? At first sight, (2) appears to be miles away from (1) and, thus, to be an unlikely candidate for an adequate analysis of (1). This impression should not be taken too seriously, however, since it is at least partly due to our unlucky tendency to read a *temporal chronology* into (2). The paratactic analysis suggests, however, no such chronology, since (2) is *not* understood as an alternative formulation of what the speaker is *uttering*; (2) is 'only' an *analysis* of the literal meaning of (1). (2) tells us something about the literal meaning of (1), but *nothing* about how the speaker proceeds in uttering (1). This is why there is absolutely no temporal element involved in (2). For the very same reason does the paratactic analysis not suppose the speaker to utter *conjunctions* of mood-setters and content sentences: Peter is supposed to say "Put off your shoes!" (i.e. (1)) but not – neither simultaneously nor in succession – "I command this" *and* "You put off your shoes".

The mood-setter and the content sentence have to be understood, furthermore, as being *neutral* with respect to force: if they were not, it would immediately follow that the particular illocutionary force of (1) must be the result (or the effect) of the illocutionary force of the content sentence or the mood-setter of (2). The content sentence, however, is an unlikely candidate, for it is not uttered by anybody (cf. above). The illocutionary force of (1) would thus have to be the result (or the effect) of the illocutionary force of the mood-setter. But consider what would be the case if the mood-setter had such a force: its most likely force would be the assertive force, since what the mood-setter appears to do is to *assert* that the content-sentence is uttered by Peter with imperative force. However, if the mood-setter itself had assertive force (or – in principle – *any* illocutionary force), it would follow that it needs its own mood-setter which specifies its assertive (or whatever other) force. The mood-setter's mood-setter, in turn, would again need its own mood-setter, and so on ad infinitum. Assuming the

mood setter and/or the content sentence to possess an illocutionary force thus leads into an infinite regress. Does this show Davidson's paratactic analysis to be insufficient to solve his 'mood-problem'? No, for the regress-problem exists only if one assumes a paratactically analyzed sentence to be *uttered*: only uttered sentences are bearers of forces! But as we saw, this would be a misunderstanding. To repeat: Peter utters (1) and *not* (2); (2) only analyzes the literal meaning of (1).

Assuming mood-setters and content sentences to be neutral with respect to force and to be in the indicative mood makes paratactically treated non-indicative sentences *"semantically tractable"* in Davidson's sense. The paratactic analysis of non-indicative sentences thus fulfils condition (3) above. But does it also fulfil the other two conditions? It does fulfil condition (1), i.e. the condition that a sentence uttered in the indicative mood shares a *"common element"* with utterances of the same sentence in the other moods: the required common element is provided by the content sentence, since the content sentence remains unaltered in whatever mood the speaker decides to use when uttering it on such-and-such an occasion and with such-and-such a force. Finally, there is condition (2), which demands of the moods to add an *"element of meaning"* to a sentence which is absent in the other moods. This condition too is fulfilled, since the varying mood-setters provide exactly what (2) is asking for. The paratactic analysis thus complies with the demands Davidson puts on an adequate approach to the grammatical moods.

But does this prove the paratactic analysis to be an adequate approach to the non-indicative moods? That there are some difficulties with the analysis is uncontested. One such problem is that the paratactic treatment does not always result in satisfying paraphrases of the original sentences. This problem emerges, for instance, with the so-called "wh-questions"; i.e. with questions of the form "What…?", "Where…?", "Why…?", "Who….?", and "How…?". Consider the following example:

(3) How many planets are there in the solar system?

The natural result, if paratactically analyzed, is:

(4) This is a question. There are planets in the solar system.

This is obviously no adequate paraphrase of (3), since what is asked with (3) is the *number* of planets in the solar system and not if there are any planets at all. Davidson proposes to solve the problem by treating wh-questions as open sentences (Davidson 1979, p. 115). In this case, the paratactic analysis of (3) results in:

(5) This is a question. The number of planets in the solar system = x.

Contrary to (4), (5) does provide an adequate paraphrase. The 'paraphrasing-problem' is thus solvable with respect to wh-questions. It resurfaces, however, elsewhere. Consider (6):

(6) If you stop by the post office, could you buy some stamps?

Since (6) might be read equally well as a question or as a politely formulated command, there are two equally plausible mood-setters. Which of them we have to choose depends (as always) on the specific context of the utterance. But whatever option one chooses, is it clear that paraphrases such as:

(7) This is a command/question. Buy stamps.

are inadequate. What is missing in (7) is the conditional-clause "If you stop at the post office". Exactly this clause, however, seems to introduce the directive element which gives (6) its imperative air. (6) might thus be better paraphrased as follows:

(8) This is a directive. If you stop by the post office, buy stamps.

The problem with (8), in turn, is that it has lost the interrogative element which is undeniably observable in (6). So maybe we get a better result if we locate the conditional-clause in the mood-setter rather than in the content sentence. The result of this manoeuvre is:

(9) If you stop by the post office, this is a directive. Buy stamps.

(9) seems to be slightly better than (8), since it somewhat softens the imperative aspect. But does it really carry the interrogative element which we are looking for? The question is difficult to answer, since the interrogative element in (6) might be said to be nothing more than a rhetorical device to soften the sentence's essentially imperative force. But even if one accepts this course of reasoning and takes (9) to be an adequate paraphrase of (6), the problem remains that in (9), much more is moved into the mood-setter than Davidson does when presenting his paratactic analysis as a suitable means to solve the moods issue (cf. Davidson 1979). Since he never considers sentences of the complexity of (6), he never addresses this problem; and commentators on his work disagree on the issue (cf. Lepore/Ludwig 2007, p. 273ff).

A completely different criticism of the paratactic treatment arises from the fact that Davidson does not tailor this analysis exclusively to accounting for sentences in non-indicative moods, but uses it (in variants) to handle many (and partially very different) recalcitrant linguistic phenomena. It is a consequence of this fact, some critics say, that the strategy fits well with some of the phenomena in question, but badly with others. According to Robert Harnish, it does particularly badly in the just

discussed case; i.e. the case of non-indicative sentences. It is, he says, an *"ad hoc"* solution (Harnish 1994, p. 420). What bothers Harnish is that Davidson proposes to treat the non-indicative sentences in exactly the same way he proposes to treat performatives. The latter, however, are typically uttered in the indicative mood; they thus lack the defining characteristic of non-indicative sentences. How, Harnish thus asks, could it be that the paratactic analysis resolves, in one case, the problem with non-indicative moods, while in the other case, where there are no such moods, it resolves another problem? Davidson's reply is that contrary to appearance, there are not *two different* problems, but *one common* problem to solve. His reasoning goes as follows: in order to utter a sentence as a command (or as a question, or as a wish), the speaker has to attach the suitable illocutionary force to that sentence. Typically, this is done by means of the speaker's using a sentence in the imperative (or interrogative, or optative) mood. Analyzing these sentences along the lines of the paratactic treatment thus provides not only a way of showing how these sentences might be made compatible with truth-conditional semantics, but likewise a way of separating the sentences from the attached illocutionary forces. This, however, is exactly what the paratactic analysis is supposed to do in the case of the performatives. Consider, for instance, the following assertion:

(10)   It is raining.

Paratactically analyzed, this turns out as:

(11)   This is an assertion. It is raining.

The content sentence in (11) reports the semantic content of (10), while the mood-setter states the particular illocutionary force with which the content sentence is uttered in (10). The situation is thus exactly analogous to the case of the non-indicative sentences. This shows, Davidson says, that there is no significant difference between the performatives and non-indicative sentences: there are not *two* problems to solve, but only *one*; namely, the treatment of the illocutionary forces.[36]

---

[36] Why does Harnish miss the analogy between performatives and the use of non-indicative moods? The problem is hardly that he fails to recognize the two phenomena as analogous, but rather that he misunderstands the particular functioning of the paratactic treatment which Davidson proposes in the two cases. Although I have no sound proof for this, there is some evidence in Harnish's essay that he takes Davidson to apply the paratactic analysis, in the case of performatives, along the lines of the *conjunction view*, while he does not do so in the case of non-indicative sentences. This suspicion feeds on passages such as *"Sometimes the utterance of two declaratives is equivalent to the utterance of a conjunction of them; why is this not true in the case of mood?"* (Harnish 1994 p. 420). There is, however, no evidence in Davidson's *Moods and Performances* (i.e. Davidson 1979) that he reads his paratactic treatment differently in the case of performatives than he does in the case of non-indicative moods. (Probably, this is not the only misunderstanding in Harnish's interpretation. Consider, for instance, the wording *"(...) the utterance of two declaratives (...)"* in the above quoted passage. Davidson makes it absolutely clear that the paratactically analyzed sentences – whether in the case of performatives nor in the case of non-indicative sentences – should be understood as *paraphrases* of what is *uttered* by a speaker, and *not* as the speaker's utterances themselves.)

Does this rejection of Harnish's argument show that he wrongly claims the paratactic analysis to be *"ad hoc"* if applied to non-indicative sentences? The answer to this question depends on your particular point of view: if you look at Davidson's proposal from a non-Davidsonian perspective (as Harnish seems to do), you are probably inclined to give a negative answer. But this is hardly due to the particular treatment Davidson proposes for the case of the non-indicative sentences. It is rather due to the fact that the introduction of special tools into his framework will always look like being ad hoc to you. This is because of the fact that it is always possible from a non-Davidsonian perspective to ask why Davidson prefers to introduce such-and-such a tool to solve such-and-such a problem rather than to acknowledge the insufficiency of his truth-conditional approach to semantics. If you take, however, the perspective from within the Davidsonian framework, the situation is completely different, for then, everything is welcome – and *not* ad-hoc – that leads to the desired result and stays within the confines of what is accessible to the truth-conditional semanticist. This is not to say that the paratactic analysis *could not* turn out to be ad hoc. It only says that it *is not* ad hoc as long as we have not shown it to fail with whatever problem it is supposed to solve. The last few pages have shown that it is rather the other way round: that the paratactic analysis seems to be a very powerful tool.

5.4 More about conventions and rules

Let me come back now to the conventions/rules which are supposedly constitutive to the understanding of language. We already considered convention relating the different types of forces with the respective grammatical moods. Other potentially indispensable conventions might obtain, however, between communicative intentions and the meanings of sentences, and between communicative intentions and the meanings of subsentential items (words and phrases). Both of these convention-types are discussed in the present section.

5.4.1 Conventions relating communicative intentions to the meanings of sentences

It is a truism that competent speakers are capable to use a language's sentences in a remarkable variety of ways. A particular sentence is on most occasions used in order to refer to the same or to relevantly similar events, actions, processes, or things. We may say that in these instances, speakers use that sentence in the *conventional* way. Accordingly, they intend the so-used sentence to mean what it *conventionally* means. It happens, however, that a speaker uses a sentence in some *non-conventional* way. This is the case, for instance, when he uses it *ironically* or *figuratively.* In the former case, the uttered sentence is intended to mean the opposite of its conventional meaning, and in the latter case there seems to be no discernible relation to the conventional meaning at all.[37]

The possibility to use sentences either in conventional or non-conventional ways poses a problem to those who take conventions to determine a sentence's meaning(s). There are basically two

---

[37] For more on figurative meaning, cf. section 6.3.2, p. 101f.

ways to deal with this fact: one can either claim (a) that there are as many conventions/rules as there are uses of the totality of a language's sentences, or (b) that there is *one* (or a *small* number of) uses for each sentence which constitute the basis on which the other uses conceptually depend. Option (a) is actually not really an option, since it inevitably leads to the conclusion that a language comprises an infinite multitude of meaning-determining conventions. The consequence would be that humans, qua finite beings, were *in*competent language users; an obviously absurd conclusion.

What about option (b)? It clearly does not disqualify as immediately as option (a). There are, however, some difficulties with (b) too: Davidson points out that it is unclear which use of a sentence could establish its allegedly indispensable conventional meaning. The intuitive first answer is that it is a sentence's *standard use*, i.e. the use which most speakers adhere to on most occasions of utterance. To rely on this intuition is, however, a dangerous move, Davidson says, since it is very likely that this intuition is an intuition about that sentence's very *meaning*. The obvious result is, thus, a circular argumentation (Davidson 1984, p. 271f). The only way to avoid this danger is to anchor the relevant sentence uses in something explicitly *non-linguistic*. The natural choice falls on the speaker's communicative intentions, for they are, on the one hand, clearly non-linguistic and, on the other hand, close enough to the language users' linguistic actions in order to be plausibly regarded as potentially informative with respect to the issue under consideration. The problem is, however, that a speaker's communicative intentions do not necessarily coincide with the speaker's ulterior aims and purposes: they do so if the speaker is honest, but they do not if he is not. Conventions, however, may attach only to what is made publicly accessible by the speaker, and this is what the speaker presents, honestly or not, as being his communicative intentions. Conventions may thus tell us only what an uttered sentence means according to the communicative intentions the speaker wants us to believe him to have, but not according to the communicative intentions that he really has. Davidson thus concludes that we should treat the sentential conventions in the same way as he proposes to treat the force/mood regularities: they often are, but need not to be, an aid to language-interpretation. Consequently, sentential conventions are as irrelevant to the conceptual preconditions of the possibility of language-understanding as are force/mood regularities.

5.4.2 Conventions relating communicative intentions to the meanings of sub-sentential items

The omnipresence of linguistic conventions is nowhere as obvious as it is with respect to sub-sentential items (words, phrases): we call a table "table" and a chair "chair", but it would not impede communication if it were common usage to do the other way round. Or think about the verbs "to chat" and "to skype": the former had a different meaning before the invention of the internet, and the latter simply had no meaning at all. Words (and phrases) may thus gain or lose their meaning(s). That they refer to what we take them to refer to is, thus, arbitrary and obviously the result of conventions.

But how do we acquire the conventions constitutive for word and phrase meaning if we cannot presuppose there to be any linguistic knowledge available? A promising attempt to solve this problem has been presented by David Lewis (Lewis 1968 and Lewis 1969). His definition reads as follows:

> *[A] regularity R, in action or in action and belief, is a* convention *in a population [say, a linguistic community] P if and only if, within P, the following six conditions hold. (…)*
>
> (1)   *Everyone conforms to R*
>
> (2)   *Everyone believes that the others conform to R*
>
> (3)   *This belief that the others conform to R gives everyone a good and decisive reason to conform to R himself. (…)*
>
> (4)   *There is a general preference for general conformity to R rather than slightly-less-than-general conformity – in particular, rather than conformity by all but any one. (…)*
>
> (5)   *R is not the only possible regularity meeting the last two conditions. (…)*
>
> (6)   *Finally, the various facts listed in conditions (1) to (5) are matters of* common *(or* mutual) *knowledge: they are known to everyone, it is known to everyone that they are known to everyone, and so on (…).* (Lewis 1968, p. 164f).

The definition is promising for two reasons: first, it avoids making use of linguistic means, and second, it does not allow Davidson to apply his argument concerning the speakers' occasional insincerities (cf. p. 65f and p. 75) since Lewis anchors the conventions in the speaker-interpreters' *common (or mutual) knowledge*, and not in the *communicative intentions* of individual speakers.[38] Is it subsentential conventions, then, which explain the understanding of a language? Davidson claims that no, since, as he argues, the understanding of words and phrases does not depend on our discerning of what is indispensable to every kind of convention (or rule), i.e. the presence of a *regularity* (Davidson 1984, p. 276f).

The best way to illustrate Davidson's claim is to consider the understanding of *malapropisms*. Malapropisms occur where speakers intentionally or non-intentionally confuse words (or phrases) of entirely different conventional meanings because of their sharing some particularly salient feature; for instance, a pronounced similarity in sound or rhythm. Representative examples of malaprop word-uses are, for instance, the confusion of the word "extensive" with "excessive" in a sentence like "The police used … force against the peaceful demonstrators"; or the confusion of "precede" and "proceed" in "Lead the way and we'll …".

But why do malapropisms provide a counterexample to the conventionalists? They do, Davidson claims, because it is impossible to relate them in any reasonable way to the meanings of

---

[38] Tyler Burge has argued that this only appears to be a merit. The reason is that a linguistic community could believe the origins of linguistic meanings to lie rather in divine mercy than in Lewisian conventions – consider, for instance, what *Genesis I* suggests. Lewis' definition does not consider this case. Cf. Burge 1979.

conventionally used words. The conventionalists might, of course, try to make use of the only tie malaprop word uses seem to provide; namely, their sharing a *salient feature* to some suitable conventionally used words. This relation is, however, ill-suited to serve the conventionalists aim, for it could simply replace a malaprop-used word by such-and-such a better suited conventionally used one. A malapropism need not, however, be an accidental event, but might be uttered *intentionally*; for instance, in order to make a joke. The replacement of the malaprop-used word would, thus, amount to the complete opposite of what a joking speaker intends to communicate and, thus, rather block than allow for understanding. The salient-feature relation could serve, thus, the conventionalist's aim only if it were such that it replaces the malaprop-used word only on those occasions where this is required. However, whether a replacement is required or not is recognizable to the interpreter only if he already knows what the speaker wants to communicate. The salient-feature relation cannot, therefore, contribute anything to the understanding of the malaprop-used word.

A further problem concerns the fact that the shared salient feature need not be a linguistic feature. Consider, for instance, the situation of a car driver who asks his passenger: "Could you pass me the water, please?" while intending him to pass the Coke. It is not unlikely that he receives what he wants if there are sufficient salient features to the passenger that indicate the driver's intention. Such features cannot concern, however, any similarities between the *words*, for "Coke" and "water" sound as dissimilar as it gets.[39] The required salient features rather concern what the two words denote; i.e. drinkable liquids. But if the salient features do not concern the words themselves, they are clearly the wrong tie for the conventionalists to rely on. The understanding of malapropisms thus shows, Davidson claims, that the knowledge of a word's/phrase's conventional meaning is not what allows for the understanding of language.[40]

5.5 The characteristics of the notion of literal meaning

The preceding sections explained why Davidson believes linguistic conventions/rules to be, in principle, irrelevant to the understanding of language. His two main arguments – the inexistence of a convention of sincerity and the occasional absence of suitable regularities – are, I think, convincing. But what do they tell us about the characteristics of his notion of literal meaning?

---

[39] There are authors who insist that a malapropism *by definitionem* resembles in sounds and/or rhythm the suitable conventionally-used word (cf. Hacking 1989, p. 449f). This is fine with me, for it is obviously merely a terminological manoeuvre: its only consequence is that the conventionalists not only have to solve a malapropism-problem, but also the problem I am describing in this paragraph.

[40] A rather bizarre argument to avoid this conclusion has been presented by Aarre Laakso (2004). He, too, operates with the similarity relation, but puts malaprop word uses not in relation to conventional word uses, but rather in proximity to the imitation of *non-linguistic* sounds (say, the imitation of the sounds of an angry dog.) An obvious consequence of this move is that malaprop-used words become as void of non-natural meaning (in the Gricean sense) as a dog's barking is. Laakso is fully aware of this consequence: *"Convention is the difference between mere verbal behaviour and true language"* (p. 3). I do not discuss his argument here in detail, since I doubt that anyone is willing to agree with Laakso that malaprop used words are only *"mere verbal [in contrast to linguistic] behaviour"*.

Disappointingly little; namely, the triviality that literal meaning is *not* a conventional notion of meaning and that it consequently rarely coincides with what our everyday intuition suggests a word's, phrase's, or sentence's meaning to be. Is the notion of literal meaning therefore a thoroughly uninteresting notion? Assuming this would be clearly premature, for the disappointing findings have little to do with Davidson's foundational notion of meaning itself, but a lot with the chapter's specific perspective. This perspective aimed at calling into question what others believe to be prerequisite to language-understanding, rather than at telling us something *positive* about why Davidson believes his own notion of meaning to do a better job. Davidson's own notion of meaning will thus loose its weak profile as soon as we say something positive about it in the next chapter.

6 Literal meaning, first meaning, and the interpretation of non-conventional language uses and linguistic idiosyncrasies

In daily life, speaker-interpreters generally express themselves in roughly the same ways as their families, friends, and colleagues do. The practical advantage of this behaviour is obvious: speaking and interpreting in accordance with the ways the others do increases the probability of understanding and being understood. It is, however, a matter of fact that occasionally, we come across an utterance where this pattern is not followed. This is the case, for instance, with ironic or figurative language use, with malapropisms, with slips of the tongue, or with speakers who suffer from physical malfunctions that do not allow for conventional pronunciation.

Are non-conventional language uses a challenge for Davidson's theory of Radical Interpretation? A look at the *paradigmatic situation* suggests that they are not, for it is obviously irrelevant to the field linguist whether the alien utterance (i.e. "Gavagai!") happens to conform to the respective language's rules/conventions or not. But this is true only with respect to the very *onset* of the field linguist's (= radical interpreter's) task, for as soon as his testing of T-Sentence-hypotheses allows him to match utterances to the entities (supposedly) referred to, non-conventional language uses become a serious problem: at no stage do they conform to the current interpretation theory, and thus they undermine, if frequent, the field linguist's attempt to construct such a theory.

Davidson reacts to this threat with a thorough modification of the theory of Radical Interpretation. This modification – first presented in his classic essay *A Nice Derangement of Epitaphs* (Davidson 1986) – provoked one of the most heated and emotional debates in the philosophy of language. It is thus the modification as well as the subsequent debate that are at the centre of the present chapter. Section 6.1 explains the modification's central aspects and mentions its most important consequences. Sections 6.2 and 6.3 are devoted to objections and to Davidson's replies: section 6.2 considers the dispute between Dummett, Hacking and Davidson, which is doubtless one of the most confusing due to the unusual amount of misunderstandings, misinterpretations, and dubious manoeuvres on all sides. Section 6.3 concerns further objections, which will allow me to clarify where my explanations have so far remained vague. Section 6.4 finally summarizes the newly emerged aspects and characteristics of Davidson's foundational notion of meaning.

Let me add a terminological remark: in order to avoid lengthy and complicated formulations, I will henceforth abbreviate the essay title *A Nice Derangement of Epitaphs* as *NDE*. The *NDE-modification* is, consequently, the modification which Davidson proposes to the theory of Radical Interpretation as presented in chapter 2. Furthermore, I will from here on refer to the theory of Radical Interpretation *before NDE* as the *pre-NDE theory of Radical Interpretation*, and to the theory of Radical Interpretation *after* the NDE-modification as the *NDE-modified theory of Radical Interpretation*.

6.1 *A Nice Derangement of Epitaphs***: modifying the theory of Radical Interpretation**

I already mentioned that the pre-NDE theory of Radical Interpretation is badly suited for describing our ability to interpret non-conventional language uses. But what exactly is the reason for this? The reason is that it assumes the radical interpreter to interpret a speaker's linguistic behaviour by making use of a *single* Tarski-style truth-theory (cf. section 2.1). This assumption presupposes the speaker's linguistic behaviour to remain very *constant* over time, since if it were not, the interpreter could not rely on one *single* such truth-theory. Non-conventional language uses are, however, *inconstancies*. Accordingly, the constancy of the Tarski-style truth-theories is not matched by an equal constancy in the ways languages are typically used. The problem is not restricted, furthermore, to non-conventional language uses, but might occur in conventional uses alike: consider, for example, the case of a child who learns that a particular word possesses more than just *the one* conventional meaning it has so far become familiar with. If the child later starts using the word with the newly acquired further conventional meanings, he breaks with the constancy previously observable in his linguistic behaviour. Changes in language use are, thus, a ubiquitous phenomenon. They are also, however, something that the pre-NDE theory of Radical Interpretation is incapable to deal with. This shortcoming is, then, much more severe than a superficial first look might suggest.

6.1.1 The NDE-modification: a first outline
The assumed constancy in the speaker's linguistic behaviour led Davidson, as just explained, to the formulation of a *static* theory. The speaker's *inconstant* linguistic behaviour requires, however, a *flexible* theory. Davidson proposes to add the required flexibility to his theory of Radical Interpretation by replacing the radical interpreter's *single* Tarski-style truth-theory by *two* such theories which he supposes him to use *in parallel*. One of those truth-theories is what Davidson calls the interpreter's *prior theory*. This is the theory the interpreter relies on *prior to the onset* of a particular interpretative task. The other truth-theory is called the interpreter's *passing theory*. This is the theory the interpreter relies on *during the process* of interpretation.

The interpreter's prior and passing theories either *coincide* with each other or *diverge* from each other to different degrees. The two theories *coincide* whenever the speaker continues to use his words (phrases, sentences) in exactly the ways the interpreter had supposed him to do at the onset of their exchange. The two theories *diverge* from each other, however, if the speaker, at some point during the exchange, happens to use a particular word(s) (phrase(s), sentence(s)) in a way that does not match the interpreter's initial expectations. Davidson illustrates this case with the linguistic behaviour of *Mrs Malaprop*, a literary figure from Richard Sheridan's novel *The Rivals*. Mrs. Malaprop suffers from confusion of words and astonishes her environment by occasional cryptic utterances. This is the case, for instance, as she comments on the beauty of some verses of a poetry book by stating that "This is a nice derangement of epitaphs". According to Standard English, her utterance suggests that she is expressing some rather bizarre thought about tombstone inscriptions. Given, however, the

mentioned context of utterance, it is very unlikely that this is what she intends to do. The interpreter must thus put aside his so far successful prior theory (which happens to be the Tarski-style truth-theory for Standard English) and find another theory which delivers a better interpretation for Mrs Malaprop's weird utterance. This is the moment when the interpreter's passing theory begins to deviate from his prior theory.

The interpreter's attempt to find a new and better-suited passing theory for Mrs Malaprop's "This is a nice derangement of epitaphs" consists in the formulation of *ad-hoc hypotheses* about the particular meanings of those words which he believes to cause the implausible interpretation. In the present case, it is not too difficult to imagine how the interpreter might proceed: simply take him to suppose Mrs Malaprop to use "derangement" and "epitaph" to respectively mean what "arrangement" and "epithet" mean in Standard English. The troublesome utterance might then be interpreted as meaning *that this is a nice arrangement of epithets* – and this is just perfectly plausible, since Mrs Malaprop is, as we saw, acknowledging the beauty of some verses in her poetry book.

Such an ad-hoc introduced passing theory may, however, turn out to be wrong. This is the case if it fails to deliver a plausible interpretation for the utterance which prompted its introduction. (One could say that in such a situation, the ad-hoc introduced passing theory diverges from the so far successful prior theory in a wrong way). The interpreter will then have to abandon this theory and replace it by some better fitting alternative. Such a situation occurs if Mrs Malaprop, on a later occasion, intends her "This is a nice derangement of epitaphs" to mean that such-and-such a thing is *a nice derangement of epitaphs*. She might do so, for instance, to remind her interpreter of the funny occasion when she mistakenly confused "epitaph" with "epithet" and "derangement" with "arrangement". This new situation obviously makes the interpreter's earlier passing theory an inadequate one, for it leads him to interpret her utterance as meaning that such-and-such a thing is a nice arrangement of epithets. The interpreter may not notice the inadequacy of his actual passing theory immediately; the necessity of a change suggests itself to the interpreter, however, as soon as he cannot make sense of the speaker's utterances anymore. An interpreter's passing theory thus has to be understood as an *ever changing* theory: it changes shape whenever it fails to deliver what the interpreter takes to be the most plausible interpretation of a speaker's utterance.

Another important aspect of the interpreter's prior and passing theories is that they are *intimately linked* to each other. The passing theory is linked to the prior theory in that it is the prior theory that serves as the basis for the formulation of passing theories. The prior theory, on the other hand, is linked to the passing theory in that it incorporates the passing theory if successful. Consider, again, the example of Mrs Malaprop: if she continues to use "epitaph" as meaning what "epithet" means in Standard English, her interpreter begins to *expect* her to use this word in this particular way. The new ad-hoc interpretation of "epitaph" thus *migrates* from the interpreter's passing theory into his prior theory. The obvious consequence of this linkage is that *theory change* is not only an issue with respect to the interpreter's passing theory, but with respect to his prior theory alike. Sure: the prior

theory does not change as rapidly as the passing theory, for a new meaning migrates from the passing theory to the prior theory only if it allows for successful interpretation over a longer period of time. The prior theory's evolution is, thus, slow in comparison to the evolution of the interpreter's passing theory. Nevertheless, there clearly *is* an evolution on both sides, and it is exactly this two-fold evolution which adds the required flexibility to Davidson's theory of Radical Interpretation: it allows, on the one hand, to react to whatever linguistic idiosyncrasies the interpreter may come across and, on the other hand, to continually adjust his expectations about the general lines of the speaker's future linguistic behaviour.

But where in this picture do we have to look for the theory's foundational notion of meaning? The answer to this question is not difficult to find, for this notion has obviously to be where interpretation actually takes place. The foundational notion of meaning is, thus, generated by the interpreter's *passing theory*. Davidson abandons, however, to refer to this notion by the label "literal meaning", for he takes it to be *"(...) too incrusted with philosophical and other extras (...)"* (Davidson 1986, p. 91). As an alternative, he proposes the name "first meaning". The renaming is, however, more than a merely terminological matter, since the notion of first meaning clearly differs from the notion of literal meaning. The notion of first meaning is, of course, as much restricted to a sentence's truth-conditions as is "literal meaning", but while a sentence's literal meaning is supposed to be *constant* over time, is its first meaning highly *volatile*. The difference between the two notions emerges because "literal meaning" is generated by *one single* Tarski-style truth-theory, while "first meaning" is generated by a permanent succession of such theories. The volatility of the notion of first meaning illustrates thus perfectly well how far-reaching the NDE-modification happens to be.

But what exactly is the impact of this modification within the larger theoretical context? This is what the next two subsections will clarify. Subsection 6.1.2 evaluates the NDE-modification's impact on the theory of Radical Interpretation and on truth-conditional semantics, and subsection 6.1.3 asks what the said modification changes with respect to the question of what it means to interpret and to understand a language.

6.1.2 The NDE-modification and its impact on the theory of Radical Interpretation and on truth-conditional semantics

To consider the NDE-modification's impact on the theory of Radical Interpretation and on truth-conditional semantics is to consider whether it changes anything *foundational* in Davidson's so far developed description of the possibility of language-understanding: does the modification change the theory of Radical Interpretation in such a way that it can no longer be taken as an attempt to show Davidson's truth-conditional semantics to be applicable to the natural languages? Or does the modification leave the theory more or less unaltered, but forces us to rethink what a truth-conditional approach to semantics actually is?

An intuitive first step towards an answer to these questions is to ask which of the central elements in Davidson's approach to language-understanding remain unaltered by the modification and which do not. Among the elements that remain unaltered are at least the following three. Firstly, the *data* which are subject to the radical interpreter's interest: they have always consisted in the individual's linguistic behaviour and go on being so. There is, secondly, the *structure* of the investigated data: Davidson has always claimed linguistic behaviour to be compositional, and he repeats this thesis, with the same arguments, and on various occasions, in *A Nice Derangement of Epitaphs* (Davidson 1986, pp. 93ff and p. 107). Finally, there is the semantically foundational *notion of meaning*: it remains, in essence, the *same* it used to be. It has, of course, altered in *some* aspects – consider the above mentioned volatility of the notion of first meaning in contrast to the constancy of the abandoned notion of literal meaning. The foundational notion of meaning remains, however, *essentially* the same, since the notion of first meaning is restricted as much to the sentences' truth-conditions as the good old notion of literal meaning was.

While the data and the foundational notion of meaning thus remain unaltered, there are important changes with respect to what *links* the two to each other. The linkage is still provided by Tarski-style truth-theories; what the NDE-modification changes, however, is the *number* of Tarski-style truth-theories required for this linkage: it is no longer just *one* such theory – and neither is it just a *pair*, but rather a *permanently changing duo* of such theories. This fact has an important consequence: it forces Davidson to abandon the idea that a Tarski-style truth-theory, or a pair of such theories, suffices to mirror a speaker's *entire* linguistic competence, for any such theory, or pair of theories, mirrors only the speaker's competence with respect to *a limited period of time* of his existence (Davidson 1986, p. 104ff). This might be illustrated with the case of a Tarski-style truth-theory which has been successfully used as passing theory, but which has to be substituted by another theory at some time t due to a change in the speaker's use of a particular word (phrase, sentence): replacing the interpreter's passing theory along the lines described in subsection 6.1.1 obviously results in the new theory's incapability to deliver plausible interpretations to all (actual and potential) utterances verbalized *prior* to the substitution. On the other hand, the passing theory prior to t is incapable to deliver plausible interpretations to all (actual and potential) utterances verbalized *after* the substitution. Passing theories thus adequately mirror a speaker's linguistic behaviour only for limited periods of time, but never over their entire life-span. (With the prior theories, the situation is basically the same: although they possess some prognostic value with respect to a speaker's future linguistic behaviour, they are subject to continual change too. It thus follows that no prior theory, at any point in time, is capable of mirroring the speaker's entire linguistic behaviour.)

Do these changes prove the theory of Radical Interpretation to have a fundamentally different status than it had prior to the modification? No, for the theory still *is* the application of truth-conditional semantics to the natural languages. It does so even better than it did before, since the introduction of the notion of first meaning allows the interpreter to deal with a speaker's non-

conventional and idiosyncratic uses of his language. It allows, in other words, to accommodate the *individualized* ways of speaking, which are so typical for human linguistic behaviour. The effect of the NDE-modification is, thus, not a change in the theory's function, but an improved fulfilment of this function.

Claiming the theory of Radical Interpretation to retain its function does not mean, however, that the NDE-modification does not have any relevant impact on Davidson's work as it presents itself at this stage. On the contrary, it is due to this modification that Davidson postulates the speaker's *idiolects* (relativized to periods of time), rather than the notion of a *shared language*, to be the semantically foundational notion of "language". His reflections can once again be illustrated by a variant of the Mrs Malaprop example: imagine her to use the word "epitaph" in the sense of what "epithet" means in Standard English, and imagine her furthermore as having done so for a long period of time (or indeed ever since). In this case, the radical interpreter's prior theory is likely to account for this particular idiosyncrasy; i.e. the interpreter *expects* her to use "epitaph" as meaning "epithet". But what happens if Mrs Malaprop suddenly uses "epitaph" as meaning "epitaph"? The answer is that she deviates from her hitherto observable linguistic behaviour. The interpreter thus has to adapt his so far successful passing theory. He does so, as subsection 6.1.1 explained, by assuming Mrs Malaprop's unusual use of "epitaph" to mean what "epitaph" means according to the conventions of Standard English; namely *epitaph*. What counts as a conventional use of "epitaph" and what does not is thus turned upside down. Davidson concludes from this that from the perspective of the radical interpreter, the conventions of shared languages cannot have any priority over whatever other regularities he might discern in a speaker's linguistic behaviour. Understanding cannot take place, then, thanks to the interpreter's knowledge of such conventions. It rather takes place thanks to the interpreter's ability to interpret *individual* linguistic behaviour that changes continually. It is, Davidson concludes, rather *idiolect*s (relativized to periods of time) than shared languages which constitute the philosophically relevant notion of "language".

Davidson even goes a step further and claims that shared languages are not only theoretically superfluous, but that they are, in the strict sense of the word, *inexistent*:

> (...) 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 structure which language-users acquire and then apply to cases (Davidson 1986, p. 107).

It is no surprise that this passage (which I henceforth refer to as the *No-Language Postulate*) has provoked vehement and partly emotional reactions. A part of these reactions is, however, rather due to the postulate's ambiguous formulation than its content, for a closer look reveals that it is not as revolutionary as it appears on first sight. The ambiguity is caused, on the one hand, by the vague

formulation that there is no such thing as *"what many philosophers and linguists have supposed"* and, on the other, due to Davidson's claim that he had maintained the now denounced notion of "language" himself at earlier times (cf. Davidson 1986, p. 95; Davidson 1994, p. 110). (Ironically, it is his long-term opponent Michael Dummett who defends him later against this self-accusation (cf. Dummett 1994b, p. 257).) What notion of "language" exactly it is that Davidson attacks will become clearer in sections 6.2 and 6.3. It remains, however, a novelty in the philosophy of language to claim that it is the speaker's idiolect, relaivized to periods of time, that provides the philosophically relevant notion of "language". It is also *this* fact which makes for the NDE-modification's most important change to Davidson's theory of Radical Interpretation and his truth-conditional approach to semantics.

6.1.3 The NDE-modification and its impact on the understanding of what it means "to interpret a speaker" and "to know a language"

The NDE-modification leads to *four* major changes in the understanding of what it means "to interpret a speaker" and "to know a language (or idiolect)" as elaborated so far. The *first* such change is explicitly stated in the passage from *A Nice Derangement of Epitaphs* which I quoted a few lines ago:

> *(…) There is (…) no such thing to be learned, mastered, or born with. We must give up the idea of a clearly defined structure which language-users acquire and then apply to cases.*

This passage clearly renounces the idea of a speaker's language as a *sharply delimited corpus of knowledge*, as it is constituted, for instance, by the contents of grammar books and dictionaries. Exactly this understanding of the nature of "language", however, was central to the pre-NDE theory of Radical Interpretation, since it was precisely such a corpus which Davidson hoped a particular Tarski-style truth-theory would mirror (cf. 6.1.1, p. 80). His concession that the nature of a speaker's linguistic behaviour is too complex to be mirrored by one single such theory, but rather requires many of them, thus amounts to the concession that "to interpret a language (or idiolect)" and "to know a language (idiolect)" should not be understood as a finite and clearly delimited task that is learned once and from then onwards deliberately mastered, but rather as an *ever-ongoing process* which comes to an end only with the decease of that language's (idiolect's) speaker. We may thus say that with the NDE-modification, Davidson starts to understand the interpreting and knowing of a language (idiolect) as something very lively, creative, and flexible, whereas his pre-NDE understanding of the mentioned activities was rather austere, static, and mechanical.

The *second* important change concerns the emphasis on the *communicative aspect* of language use. Although Davidson never denied this aspect to be important, it previously was of rather peripheral interest. The reason is that the pre-NDE theory of Radical Interpretation tacitly assumed the speaker's linguistic behaviour to be constant over time, and that it was thus natural to assume the speaker to be the interpreter's *passive* partner. This oversimplified picture changes completely with the NDE-

modification: Davidson now regards successful interpretation as the result of the mutual efforts of the interpreter *and the speaker*. This becomes particularly clear from the case where Mrs Malaprop unexpectedly intends her "This is a nice derangement of epitaphs" to mean that *this is a nice derangement of epitaphs*: although it is the exclusive decision of the *interpreter* whether *this* particular linguistic behaviour requires him to adjust his passing theory, it is Mrs Malaprop (i.e. the *speaker*) who indicates what such an adjustment must look like; the interpreter cannot arrive at a suitable adjustment without a careful scrutiny of what he takes Mrs Malaprop's communicative intentions to be.[41] It thus follows that successful interpretation is no longer taken to depend exclusively on the radical interpreter's ability to find the suitable interpretative theory, but likewise on the speaker's ability to let the interpreter know about what she aims to convey to him in such-and-such a particular situation. Abandoning the idea of strict constancy in the speaker's linguistic behaviour thus emancipates the speaker from her so far passive role; and this is obviously the same as to say that the NDE-modified theory of Radical Interpretation puts much more weight on the communicative aspect of language use than the pre-NDE theory did.

This adjusted and more complete picture of language-interpretation and understanding of course leaves its traces in the formulation of the NDE-modified theory of Radical Interpretation. Although I have spoken so far only about the *interpreter's* passing and prior theories, Davidson takes the *speaker* to have such theories too. The speaker's prior and passing theories fulfil, however, different functions than the interpreter's theories do: while the speaker's prior theory reflects how the speaker assumes his interpreter to interpret particular sentences prior to the onset of their linguistic exchange, his passing theory reflects how he intends his interpreter to interpret a particular sentence on a particular occasion of utterance (Davidson 1986, p. 101). It is obvious that the speaker's prior and passing theories change their appearance over time, that they may coincide and deviate from each other, and that they influence each other along the same lines as we saw the interpreter's theories do (cf. subsection 6.1.1). Furthermore (and in analogy to the interpreter's side), it is the speaker's passing theory which is decisive for the successful conveyance of communicative intentions (and, therefore, for the possibility of successful interpretation). It follows that according to the NDE-modified theory of Radical Interpretation, the conceptual precondition for successful interpretation (and understanding) is the *coincidence* of the speaker's and the interpreter's respective passing theories. Davidson puts it as follows:

> *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 [passing] theory the speaker intends the interpreter to use. Only if these [theories] coincide is understanding complete. (...) The passing theory is where, accident aside, agreement is greatest* (Davidson 1986, p. 102).

---

[41] Contrary to appearance, this fact does *not* introduce an intensional element (= the speaker's *communicative intentions*) into Davidson's truth-conditional approach to semantics; cf. my explanations on p. 87.

The *third* important change concerns the particular *nature of the extra-semantic evidence* involved in the processes of language-interpretation. Such evidence is central throughout the development of the theory of Radical Interpretation, since the assignment of particular truth-conditions to a speaker's uttered sentences obviously requires the radical interpreter to possess evidence which is *independent* of the truth-conditions-assigning theory itself. The pre-NDE theory of Radical Interpretation thus assumed there to be such evidence. What it did not assume, however, is that the speaker is *actively creating* such evidence. This assumption obviously requires a revision in the light of the above explained explicit communicative focus of the NDE-modified theory: it is, as we saw, the speaker's communicative intentions which allow the interpreter to construct a successful interpretation theory. Davidson is well aware of this fact, and welcomes it enthusiastically:

> *What matters to successful linguistic communication is the intention of the speaker to be interpreted in a certain way, on the one hand, and the actual interpretation of the speaker's words along the intended lines through the interpreter's recognition of the speaker's intentions, on the other* (Davidson 2005b, p. 51).

Although this passage has an unusual Gricean touch, it should not be misunderstood as a deviation from Davidson's so far pursued path: communicative intentions surely gain a so far unprecedented importance (or rather: prominence) in his picture, but they clearly remain *outside* of what he has always considered to be *semantics*. This becomes clear from the fact that Davidson speaks in the quoted passage about the importance of communicative intentions for *successful communication*, but not for the *constitution of an utterance's first meaning*.

The explanation of the *fourth* (and last) important change requires drawing attention to two already noticed characteristics of the NDE-modified theory of Radical Interpretation:

(a) Successful communication takes place if the speaker's and the interpreter's passing theories coincide, and

(b) the philosophically relevant notion of "language" is the speaker's idiolect relativized to periods of time (= the *No-Language Postulate*)

If we consider (a) and (b) in combination, it is unclear at first sight where the evidence for the interpreter's construction of his passing theory may come from. A first attempt to solve the problem is to claim that it comes, on the one hand, from the observable evidence in the speaker's and the interpreter's mutual environment, and, on the other hand, from the interpreter's so far developed prior theory for the speaker under consideration. The construction of a prior theory requires, however, its own empirical evidence. The question is, thus, where *this* evidence comes from. An intuitive resource

is the conventions/rules of the speaker's linguistic community. Convention-governed language use, is, however, supposed to be conceptually *posterior* to the speaker's individual language use (this is what is claimed by the *No-Language Postulate*). The construction of the interpreter's interpretation theories cannot rely, thus, upon any basis *distinct* from the particular ways the speaker uses his words, phrases, and sentences. What the interpreter must rely on instead is, firstly, the experience collected so far about the linguistic behaviour of the speaker under consideration, secondly, the linguistic conventions/rules he supposes the speaker to follow (in general), and, thirdly, a good portion of *"wit, luck, and wisdom"* (Davidson 1986, p. 107). The NDE-modified theory of Radical Interpretation thus supposes the interpreter to make use of *whatever empirical evidence* he has at his disposition. The consequence of this is that his construction of interpretation theories (i.e. of prior and passing theories) no longer relies on any evidence exclusively designed for the construction of such theories, but on evidence which may serve in the construction of *any* empirical theory. Davidson thus concludes that:

> *(…) [w]e should realise that we have abandoned not only the (…) notion of a language, but (…) erased the boundary between knowing a language and knowing our way around the world generally* (Davidson 1986, p. 107).

It is not difficult to see that this consequence is no less revolutionary than the No-*Language Postulate*. Davidson has been criticized for erasing this boundary as often as he has been criticized for the claim that *"there is no such thing as a language"*. Among the most vehement critics of both these claims are Michael Dummett and Ian Hacking. The subsequent section begins to evaluate the NDE-modification by considering their objections.

6.2 The Dummett/Hacking/Davidson controversy on *A Nice Derangement of Epitaphs*

It is the aggressive polemic that first strikes the reader's eye when reading Dummett's and Hacking's replies on *A Nice Derangement of Epitaphs*. Hacking, for instance, qualifies the *No-Language Postulate* as *"downright astonishing"*, and rhetorically asks whether *"(…) there [is] no longer language for there to be philosophy of?"* (Hacking 1989, p. 447). Dummett, on the other hand, wonders whether it is a collective illusion that brings suppressed people to claim their language to be *"the soul of their culture"*, and their oppressors to *"punish [those people's] children for speaking (…) [these languages] in the playground"* (Dummett 1989a, p. 465). No doubt: such statements are far off the point and have hardly anything to do with what Davidson is up to. We should not ignore, however, that Davidson himself invites such responses: the claim *"(…) that there is no such thing as a language if a language is anything like what many philosophers and linguists have supposed"* is an obvious provocation, since it implies that Davidson's colleagues (and allegedly he himself at earlier times) have been wrong about what the whole philosophy of language is about. No wonder that many of them are more than just a little upset by it!

Besides the considerable degree of polemic and provocation in the Dummett/ Hacking/Davidson controversy, it is beyond doubt that all involved philosophers are driven by serious aims and that they believe their respective opponents to commit far-reaching errors. But what *is* the controversy among them? As far as I can see, the disputed issues are the following two:

(5) Is it a shared language or rather the speakers' idiolects (relativized to periods of time) that constitute the philosophically relevant notion of "language"?
(6) Is a conventionally shared language prerequisite to the uncontested view that language use is an essentially social activity?

Both (1) and (2) obviously stand in close relation to the *No-Language Postulate* and its consequences. But before discussing the opponents' views with respect to these issues, we have to get rid of a few misunderstandings about the content of Davidson's *A Nice Derangement of Epitaphs*.

6.2.1 Eliminating misunderstandings

The first important misunderstanding concerns the precise notion of "language" Davidson attacks. Consider, for instance, Hacking's and Dummett's above mentioned polemic remarks: Hacking seems to fear that Davidson deprives the philosophy of language of its subject altogether. This is, of course, nonsensical, for idiolects are as *linguistic* as any other notion of "language". It thus follows that in Davidson's NDE-modified theory of Radical Interpretation, there clearly *is* a defined linguistic entity for there to be philosophy of. The notion of "language" which Davidson attacks thus has to be something else than what Hacking supposes it to be. Then there is the dubious statement by Michael Dummett: as far as I can see, Dummett takes Davidson to attack the notion of "language" which refers to such things as what the Kurdish or the Roma people speak (i.e. "the Kurdish language", "the Roma language", respectively). This everyday notion of "language" is, however, clearly *not* the notion which Davidson attacks, since he explicitly agrees that idiolects often resemble each other to high degrees, and that aggregates of such idiolects might be referred to as "the Kurdish language", "the Roma language", etc (Davidson 1994, p. 111). It is, then, wrong to say that Davidson denies the Kurds, the Roma or any other people to posses a language, or that these people (and their oppressors) fight for (or against) a chimera. But what, then, is the precise sense of "language" which the *No-Language Postulate* claims to not exist? Davidson clarifies the exact content of his postulate in his response to Hacking's and Dummett's objections:

> *(...) the concept of language I opposed [in "A Nice Derangement of Epitaphs"] (...) was this: in learning a language, a person acquires the ability to operate in accord with a precise and specifiable set of syntactic and semantic rules; verbal communication depends on speaker and [h]earer sharing such an ability, and it requires no more than this. I argued that sharing such*

a previously mastered ability was neither necessary nor sufficient for successful linguistic communication (Davidson 1994, p. 110).

The notion of "language" presented in *this* passage is obviously a very constrained and limited one and hardly what Hacking and Dummett take Davidson to attack. Had we perhaps better qualify their objections as mere rhetoric without any philosophically relevant substance? This is hardly a plausible idea: why should they have made their bizarre claims if they knew them to be completely wrong? Sure: there *is* an element of rhetoric in their statements (as much as there is rhetoric in Davidson's), but it is implausible to claim them to be *nothing but* rhetoric. At least Dummett's later concession that Davidson was attacking his constrained notion of language for the *"right reasons"* (Dummett 1994b, p. 257) suggests that at the time of writing, he was not yet aware of what exact notion of "language" Davidson's *No-Language Postulate* happens to attack.

Even more important than the quarrels about the attacked notion of "language" are the misunderstandings about the exact workings of Davidson's prior and passing theory-construction. This is particularly true with Hacking, who mistakenly assumes the radical interpreter to apply his passing theory *in succession*, rather than *in parallel*, to his prior theory. The result is that he takes the radical interpreter, at every instance of theory-change, to begin *from scratch* instead of assuming him to introduce the new passing theory on the basis of the continuously developed prior theory. This wrong understanding has two misleading consequences: it makes (a) the Tarski-style truth-theories which constitute the interpreter's prior- and passing theories look as completely detached from each other, and suggests (b) that the respective Ls of these Tarski-style truth-theories seem to lack any resemblance to the everyday notion of "language". The second observation is correct insofar as the respective L's of each of the interpreter's prior and passing theories correspond only to a relatively small slice of the total linguistic competence of the speaker's whole life. It is wrong, however, if these L's are understood as having nothing to do with his linguistic competence at all. But it is exactly this conclusion which Hacking's wrong understanding of the prior and passing theory-construction suggests, for it looks from his perspective like nothing more than a lucky accident that those disconnected truth-theories assign exactly those truth-conditions to the speaker's utterance which look, from the interpreter's perspective, as the truth-conditions which the speaker himself is assigning to them. It is thus no surprise that Hacking claims these theories' Ls to be *"mere formalisms"* (Hacking 1989, p. 454). If he had realized, however, that the prior and passing theories are intimately linked to each other, he would hardly have made this mistake; he would have seen their connection to the Ls of antecendent prior and passing theories, and thus understood their intimate connection to the speaker's linguistic behaviour; i.e. to his idiolect.[42]

The very same mistake is at work when Hacking writes that he is

---

[42] The mistake partly explains why Hacking believes the NDE-modification to deprive the philosophy of language of its subject matter (cf. the quote on p. 88). It explains it, however, only *partly*, for it remains unclear why Hacking fails to realize that idiolects are as linguistic as any other notion of "language".

(…) calling in question (…) [the] idea that we construct a "total" theory [of interpretation] of the other. (…) The more one thinks of conversation as evolving passing theories, the more open one may be to a lot of different passing theories about different aspects of a person's life (Hacking 1989, p. 457).

Hacking is right, of course, that there is no *"total theory"* and that there need to be *"a lot of different passing theories"*. But this is not what Davidson objects to. On the contrary: the permanently developing prior and passing theories provide, at no stage of their development, a *"total theory"* in Hacking's sense. They rather provide what Hacking is asking for, namely *"(…) a whole bunch of tricks for seeing what connects with what"* (i.e. *"a lot of different passing theories"*) (Hacking 1989, p. 456).

It was Dummett who first realized that Hacking must have misunderstood the exact functioning of the prior and passing theory construction. He complains, however, that the misunderstanding is primarily due the confusing labels Davidson attaches to his theory-duo: the word "prior" in "prior theory" suggests, he says, that this theory has to be understood as a theory which is *prior* to some other, subsequent theory. Hacking's misunderstanding could thus have been prevented, Dummett says, if Davidson had named his theories in another way, say, as *long-range theory* and *short-range theory*, respectively (Dummett 1989a, p. 460). Whether or not this is correct is difficult to say: as far as I know, Hacking has never confirmed nor disconfirmed Dummett's supposition.

An ambiguous wording in Davidson's *A Nice Derangement of Epitaphs* is, however, the cause of Dummett's own misunderstanding of how the prior and passing-theory construction works. The problematic passage reads as follows:

For the hearer, the prior theory expresses how he is prepared in advance [i.e. prior to the onset of a communicative exchange] 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* (Davidson 1986, p. 101).

The problem is with *"For the speaker, the prior theory is what he* believes *the interpreter's prior theory to be (…)."*. Dummett reads this sentence quite literally and understands Davidson as claiming that the speaker's prior theory happens to be a theory about another theory; namely, about the interpreter's prior theory. This reading is, if correct, absolutely lethal to Davidson's NDE-modified theory of Radical Interpretation; for if the speaker's prior theory had to be understood as a second-order theory, it would immediately follow that the interpreter's prior theory is a second-order theory too. But if both the speaker's and the interpreter's prior theories were second-order theories, it would

immediately follow that neither of them could be relevant to successful interpretation. In Dummett's words:

> *We cannot say both that the speaker's theory is a theory about what the hearer's theory is, and that the hearer's theory is a theory about what the speaker's theory is, without falling into an infinite regress: there must be some first-order theory* (Dummett 1989a, p. 467).

To suppose that Davidson has made a final and lethal mistake here is, however, premature: it is true, of course, that Dummett's literal reading of the above quoted passage suggests that Davidson takes the speaker's and the interpreter's prior theories to be second-order theories. But it is very unlikely that Davidson really intends to be understood in this way. It is unlikely, since Davidson has *always* emphasized that he does not want the theory of Radical Interpretation to be understood as an attempt to explain what the interpreter (or the speaker) explicitly knows, or explicitly needs to know, in order to be in a position to interpret (or to be understood). The theory of Radical Interpretation – whether NDE-modified or not – is not intended to be about what is *de facto* going on in an interpreter's (or speaker's) head. It is rather a theory which is intended to *describe* what kind of knowledge a third and uninvolved party (for instance: the armchair philosopher) has to suppose the interpreter (and speaker) to possess in order to arrive at a satisfactory understanding of what it is that enables two (or more) linguistically gifted individuals, in principle, to use these abilities in order to successfully communicate with each other. Davidson repeats this genuinely *descriptive perspective* of the theory of Radical Interpretation in *A Nice Derangement of Epitaphs* as follows:

> *To say that an explicit theory for interpreting a speaker is a model of the interpreter's linguistic competence is not to suggest that the interpreter knows any such theory. (…) In any case, claims about what would constitute a satisfactory theory [concerning the question of what counts as an adequate model of the interpreter's linguistic competence] are not, as I said, claims about the propositional knowledge of an interpreter (…). They are rather claims about what must be said to give a satisfactory description of the competence of the interpreter* (Davidson 1986, p. 95f).

Neither the speaker's nor the interpreter's prior theory should thus be treated as being known explicitly by the speaker or the interpreter, respectively. As a consequence of this, they should neither be treated as second-order theories, since second-order theories are necessarily theories which are *explicitly known*: it would not make any sense to say that one entertains a theory about a theory (i.e. a second-order theory) if one were incapable to say about what particular theory one is entertaining that theory. But if neither the speaker's nor the interpreter's prior theories are second-order theories, it is ill-advised to put much weight on the badly chosen wording in Davidson's troublesome sentence (i.e.

"For the speaker, the prior theory is what he believes *the interpreter's prior theory to be (…)."*). It is far more in the spirit of Davidson's approach to read that sentence as meaning that the speaker's prior theory is what the speaker *is disposed* to believe the interpreter's prior theory to be. Although this proposal is not Davidson's own, but Ernest Lepore and Kirk Ludwig's (cf. Lepore/Ludwig 2005, p. 273), it is unlikely that he would have disapproved of it. The reason is not only that it avoids the second-order theory problem. It is likewise because of the fact that the proposed reading prevents the introduction of *intentional* notions into the theory (namely beliefs), which would be obviously lethal to the idea that the theory of Radical Interpretation is the application of Davidson's truth-conditional approach to the semantics of natural languages. It is, thus, pretty safe to suppose that Davidson would have preferred Lepore/Ludwig's reading of the ill-formulated passage to the very literal one proposed by Michael Dummett.

6.2.2 What remains of the controversy

Is there anything left of the controversy after having cleared the misunderstandings? Yes, for neither of them changes anything about the disputed issues mentioned at the beginning of the present section. These issues were, to repeat, whether it is rather a speaker's idiolect or a conventionally shared language which has to be understood as the philosophically relevant notion of "language", and whether a conventionally shared language is prerequisite to the uncontested view that language use is an essentially social activity.

But what do Dummett and Hacking exactly claim about these two issues? They claim, according to Davidson:

(1)   that his (i.e. Davidson's) emphasis on the theoretical possibility of convention-/rule-free language use is a piece of pointless speculation, since it is hardly plausible that anything like it ever should occur in practice,

(2)   that the speakers of a linguistic community are responsible to the relevant social norms of their community (i.e. the community's linguistic conventions/rules), and that it is therefore misguided to discuss non-conventional language uses without considering them in the light of the said responsibility, and

(3)   that he, Davidson, without taking into account the social norms of linguistic communication, is incapable to explain the difference between "using a word correctly" and "thinking that one is using it correctly"; i.e. that he, Davidson, is incapable to explain what makes language use an essentially social activity (Davidson 1994, p. 116).

With respect to (1), Hacking and Dummett argue as follows: the NDE-modified theory of Radical Interpretation is, against Davidson's intention, not about linguistic *communication*, but rather about

monologue interpretation (Dummett 1989a, p. 462) and about the one-way linguistic exchange between two individuals who are both frozen in their roles as interpreter and speaker, respectively (Hacking 1989, p. 548). These claims, Dummett and Hacking explain, are based on the fact that Davidson ignores that linguistic communication is a process where the roles of speaker and interpreter change constantly. It is his focus on the role of the interpreter, they say, that leads Davidson to ignore the importance of conventions/rules in communication. Only if Davidson would start to recognize the speaker and hearer as constantly switching their roles would it be possible for him to see how far removed from practice his reflections about convention-/rule-free communication happen to be. His speculations about the possibility of convention-/rule-free communication are, thus, pointless.

The argument is bad: Davidson at no point ignores that the speaker and the interpreter switch roles constantly. Sure: his reflections almost always take the interpreter's perspective. But this does not say that he does not suppose the speaker and interpreter to exchange their roles. It merely emphasizes that the theory of Radical Interpretation is a theory about language-*understanding*: such a theory simply cannot be formulated from another perspective than from the interpreter's perspective.

Neither does Davidson ever deny the importance of conventions in daily linguistic affairs. On the contrary: he agrees throughout that in everyday language use, conventional usage of words, phrases and sentences is rather the rule than the exception (cf. Davidson 1984, p. 265f; Davidson 1994, p. 119). The theory of Radical Interpretation is, however, not about what *de facto* allows us to communicate with each other, but about what is required *in principle* to do so (cf. section 5.1, p. 61). The de facto importance of conventions/rules in linguistic communication is, therefore, simply irrelevant to the issue the theory of Radical Interpretation is about. Objection (1) therefore misses its target.

Objection (2) is about the speaker's responsibility towards the relevant social norms of his (and maybe his interpreter's) linguistic community. Davidson agrees with Dummett that in everyday situations, there is such a responsibility. He even agrees that this responsibility might be strongly felt by an individual, and that violations can be harshly punished by others. Davidson considers cases such as the situation of a teenager who happens to use particular words in an unfashionable, out-dated way, or a child who curses in the presence of her parents. The parents are, he says, likely to condemn the infant's cursing, and the teenager's peers may mock him. Impending punishment is, then, likely to make both the child and the teenager feel obliged to the respective relevant linguistic norms. Davidson claims, however, that this obligation has more to do with the particular social status of the respective individual within a group (the family, the clique) than with the understanding of what the individual intends to communicate. *"Using a word in a nonstandard way (…)"*, he says,

> may be a faux pas *in the same way that using the wrong fork at a dinner party is, and it has as little to do with communication as using the wrong fork has to do with nourishing oneself, given that the word is understood and the fork works* (Davidson 1994, p. 117).

Obeying the social norms and conventions of one's language is, according to Davidson, therefore no precondition for communication. The only responsibility that he grants there to be is the speaker's responsibility to express his communicative intentions in a way that maximizes the likeliness of being interpreted as intended. The best way of doing so is in many cases to comply with the communities' social norms (conventions, rules). Such conformity, however, is neither a necessary nor a sufficient condition for successful interpretation: our generally effortless understanding of malapropisms or slips of the tongue proves the opposite; and the fact that it is sometimes easier and more straightforward to convey a communicative intention by making a joke or by using a sentence ironically or figuratively does so too. To use language in conformity to the relevant social norms is, thus, only a *contingent* requirement for the successful conveyance of communicative intentions (cf. chapter 5).

Dummett's opinion on Davidson's reply is not entirely clear; the reason is that either Dummett or Davidson misunderstand what their respective opponent argues for. According to Dummett, Davidson misunderstands him as claiming that he, Dummett, takes responsibility to social norms to be conceptually prerequisite to the possibility of linguistic communication. But this, he says, has never been his intention: what he argues for is rather, he explains, that the speakers of a particular linguistic community are required to obey their shared language's norms in order to preserve this language's particular expressive powers, its distinctive vocabulary and particular syntactic characteristics (Dummett 1994b, p. 266). Assuming that Dummett really wants to be understood in this way, it is of course correct to emphasize the importance of historically-grown linguistic norms, since repeated violations of them doubtless harm the linguistic heritage. This heritage, however, is completely irrelevant to Davidson's aim, since the interpreter's ignorance of the historic roots of a speaker's language cannot impede his ability to introduce a new passing theory whenever he judges this to be required in order to arrive at an adequate interpretation of a speaker's utterance. The obvious irrelevance of Dummett's argument thus makes it difficult to see what Dummett might have in mind when replying in the way he does. Since he does not mention any further argument with respect to (2), it is unclear what he thinks about Davidson's rejection.

Let us proceed then to (3), i.e. to Wittgenstein's question about how to distinguish between "using a word correctly" and "thinking that one is using a word correctly". Being in a position to draw this distinction obviously requires there to be at least *two* linguistically skilled individuals: one who thinks that he is using such-and-such a word correctly, and another who judges whether he indeed does so or not. A theory capable of drawing this distinction is, then, a theory that regards language use to be an essentially *social* activity. Objection (3) consequently claims the *opposite*; i.e. that Davidson's NDE-modified theory of Radical Interpretation allows the use of a language to be an entirely *private* matter because of its alleged inability to distinguish between correct and incorrect uses of words. Private languages, however, are not only useless in communication, but have been shown by Wittgenstein to be conceptually impossible (Wittgenstein 1952/2009, §§ 256f).

According to Wittgenstein (and Dummett), the only way to prevent a theory from allowing for private languages is to think about language use in terms of *rule-* (*convention-*) *following* (cf. section 5.1). A word is, according to this view, correctly used if it is used in accordance with the relevant linguistic rule (convention), while it is wrongly used if it is not used in accordance with that rule (convention). The Wittgenstein/Dummett way of reasoning is obviously not open to Davidson, for his claiming language-understanding to be independent of a speaker's rule/convention-following makes it impossible to claim that it is thanks to such rules/conventions that there is a distinction between correct and incorrect uses of words. Is Dummett, then, correct in claiming Davidson's NDE-modified theory of Radical Interpretation to allow for private languages?

Davidson says that he is not, for rejecting the view that languages are governed by rules/conventions is not to claim that there is no *norm* that distinguishes between correct and incorrect uses. On the contrary: a speaker who intends to communicate such-and-such a proposition cannot do so, Davidson says, without simultaneously intending *to be understood as communicating this particular proposition*. This fact forces him, however, to use his linguistic abilities in specific ways; namely, in those ways that maximize the probability of being understood as he wants to be understood. The speaker cannot, thus, use words (phrases, sentences) in whatever ways that happen to please him, but has to consider how the interpreter, according to his (= the speaker's) best knowledge, is most likely to interpret them. In order to be understood as he wants to be understood, the speaker must thus care about how words (phrases, sentences) are used *by others*. The mentioned intention thus provides an intersubjectively accessible *norm* with respect to how words (phrases, sentences) should or should not be used (Davidson 1994, p. 120). This alternative norm is, of course, a different norm than the Wittgenstein/Dummett norm, for it provides a criterion to distinguish between what *fosters* and what *hinders successful communication*, whereas the Wittgenstein/Dummett norm explains what distinguishes *correct* and *incorrect* uses of words with respect to *rule/convention-governed languages*. This difference is, however, irrelevant with respect to the question whether Davidson's norm allows him to regard the use of language as an essentially social activity: it obviously does, for claiming language use to depend on making oneself interpretable to a *second* person obviously *is* to say that language use *is* essentially social.

Dummett's reaction to Davidson's alternative norm is unfortunately completely unclear: he grants that making oneself interpretable is an important aspect for successful communication, but continues to maintain that the sharing of a language is so too. His thought is that two people who initially possess no knowledge about each other's language, but who come to a mutual understanding at a later point in time, do so thanks to their mutually acquiring the rules/conventions which govern their respective languages (Dummett 1994b, p. 263). Whether this is correct or not is, however, irrelevant to the present issue, for Dummett's communicators become sharers of their respective languages *in the course of* their communicative exchanges with each other. Whether Davidson's above mentioned norm really provides an alternative to the Wittgenstein/Dummett norm depends, however,

on the question whether the sharing of a rule/convention-governed language is required *prior* to the onset of such exchanges. Dummett's conclusion that it remains *"somewhat obscure (…) how far apart Davidson and I really are on the strictly philosophical issues"* (Dummett 1994b, p. 265) thus gains an ironical significance: it is doubtless true, although hardly because of the reasons Dummett is thinking of. It is clearly unsatisfactory that Dummett's objections against the NDE-modification cannot be clarified, but rather diffuse in dubious ways. Nevertheless, none of them shows the NDE-modified theory of Radical Interpretation to be untenable. It is, then, time to leave the Dummett/Hacking/Davidson controversy and to turn to further objections against the NDE-modification.

6.3 Further objections and replies

Davidson's *A Nice Derangement of Epitaphs* provoked an astonishing number of critical reactions. Since it is impossible to discuss all of them, I focus here on three particularly interesting ones. Subsection 3.6.1 presents an argument by Dorit Bar-On and Mark Risjord (Bar-On/Risjord 1992), who are both highly sympathetic to Davidson's truth-conditional approach, but who believe it to be undermined by the *No-Language Postulate.* Subsection 3.6.2 considers an objection by Karen Green (Green 2001), who criticizes that Davidson does not treat all kinds of non-conventional language uses in the same way, and argues that the removal of this incoherence avoids the *No-Language Postulate.* Subsection 3.6.3 concludes this chapter by considering Catherine Talmage's thesis that Davidson's notion of first meaning does not replace the notion of literal meaning, but rather runs parallel with it (cf. Talmage 1994 and Talmage 1996).

6.3.1 Bar-On/Risjord's attempt to save Davidsonian semantics by means of rejecting the No-Language Postulate

Dorit Bar-On and Mark Risjord are, as I already said, sympathetic to Davidsonian truth-conditional semantics. They believe, however, that the *No-Language Postulate* puts it in severe jeopardy and needs to be withdrawn in order to save it. The danger is, they say, that the *No-Language Postulate* deprives the interpreter's interpretation theories (i.e. his passing and prior theories) of all empirical content and Davidson's truth-conditional approach of its entire subject matter. This is a very serious verdict; but what course of reasoning leads Bar-On/Risjord to this conclusion?

The starting point for their reflections is the (correct) observation that the Tarski-style truth-theories which constitute the interpreter's prior and passing theories can only serve as a basis for the interpretation of linguistic behaviour if these theories' allocations of truth-conditions are confirmed or disconfirmed by evidence completely independent of the theories themselves. We already came across this fact in subsection 6.1.3 (cf. p. 87): its importance is uncontested and emphasized by Davidson on various occasions. In *The Structure and Content of Truth*, he puts it as follows:

(...) the question whether a theory of truth is true of a given language (that is, of a speaker or a group of speakers) makes sense only if the sentences of that language have a meaning that is independent of the theory (otherwise the theory is not a theory in the ordinary sense, but a description of a possible language) (...). [I]f the question can be raised (...) the language must have life independent of the definition (...) (Davidson 1990, p. 300f).

According to Bar-On/Risjord there is, however, no such evidence – i.e. nothing that makes *"a theory of truth (...) true of a given language"* – if Davidson is right that *"there is no such thing as a [shared] language (...)"*. Their argument comprises two points: firstly, that in order for there to be the required evidence, an intersubjectively accessible *criterion* is needed which distinguishes between what counts as such evidence and what does not; and secondly, that it is only thanks to language-constituting *conventions* (or *rules*) that there is such a criterion (Bar-On/Risjord 1992, p. 187).

The first point is certainly correct. But what about the second? The second point is, above all, extremely puzzling: it is puzzling because Davidson has *always* claimed linguistic conventions/rules to be superfluous to the possibility of language-understanding. So why should their absence pose a problem to the NDE-modified theory of Radical Interpretation if it did not do so, as Bar-On/Risjord agree, to the pre-NDE theory? The second point is also puzzling because the said conventions/rules obviously provide *linguistic* evidence for the confirmation/disconfirmation of a truth-theory (= passing theory). The theory of Radical Interpretation requires, however, *extra-linguistic* evidence to confirm/disconfirm truth-theories, for in the paradigmatic situation of radical interpretation, extra-linguistic evidence is the only evidence available to the radical interpreter (cf. section 2.2).

Part of the puzzle seems to be due to Bar-On/Risjord's mistaken fear that the *No-Language Postulate* slices the speaker's language in fragments too little to allow for a determination of their truth-conditions (and, thus, their literal/first meanings). Consider the following passage from their essay:

On the picture that emerges from the later Davidson [i.e. the NDE-Davidson] (...), the axioms and structural pattern laid out by an interpreter's truth theory [= passing theory] must not be regarded as representing a set of rules or conventions antecedently mastered by speakers and discovered by an interpreter. (...) This is how we become liberated from the notion of a language. But then the only thing that could be said to have "a life independent of the [truth] definition" [cf. the Davidson-quote above] would be individual utterances uttered by speakers, considered in isolation (Bar-On/Risjord 1992, p. 187).

The quote reveals a misunderstanding: it does not follow from the NDE-modification that the only remaining notion of "language" is *"individual utterances (...) considered in isolation"*, for the NDE-

modification assumes an interpreter's passing theory for such-and-such a speaker to be intimately linked to a prior theory for this speaker (cf. subsection 6.1.1). The passing theory's particular shape depends, thus, on the shape of the prior theory, whereas the shape of the prior theory is the result of successful interpretations of *preceding* utterances of that particular speaker. What 'gives life to' a passing theory for such-and-such a speaker is, then, the successful interpretation of a *succession of utterances* rather than of *"individual utterances (...) considered in isolation"*. The *No-Language Postulate* thus does not fragment the notion of "language" (or better: "idiolect") in the sense Bar-On/Risjord seem to assume. – However, Bar-On/Risjord's misunderstanding might be only a (minor) *part* of the puzzle, since it does not explain the puzzle's *core*; i.e. why Bar-On/Risjord insist that it is a *convention-/rule-governed* notion of "language" that is required for there to be a criterion which distinguishes between evidence that confirms a particular passing theory and evidence that does not. So what might explain the puzzle's core?

The answer is, I think, a rather unphilosophical fact; namely, that Bar-On/Risjord uncritically rely on the correctness of Bjørn Ramberg's reading of Davidson (cf. Ramberg 1989) instead of consulting Davidson's own essays. The consequence is that they hardly ever criticise Davidson's own arguments, but almost always what Ramberg claims Davidson to say.[43] The problem is, however, that there is a far-reaching mistake in Ramberg's interpretation, and it is exactly this mistake that must have lead Bar-On/Risjord to their puzzling claim. Ramberg's mistake (or *thesis* I should rather say) is not restricted to the NDE-modification, but concerns Davidson's entire philosophy of language. It is the thesis that Davidson is driven by the intention to *"exorcise"* (Ramberg 1989, p. 2) the very concepts that are typically considered as fundamental in theories of meaning and language-understanding; namely, the concepts of "reference", "meaning" and – with the NDE-modification – of the concept of "language". Ramberg's thesis is, however, untenable. Consider the concept of "language": what Ramberg has in mind when he claims that Davidson wants to exorcise it is apparently the famous dictum that *"there is no such thing as a language (...)"*. But does this phrase prove that Davidson intends to *exorcise* the *complete* concept of "language"? Certainly not, since this claim, i.e. the *No-Language Postulate*, is not aimed at the concept of "language" *in general*, but only at a very specific understanding of it; namely, that a language consists in a *"precise and specifiable set of syntactic and semantic rules"* (cf. Davidson 1994, p. 110, as well as subsection 6.2.1, p. 89f). This, however, says nothing about any *other* notion of "language"; for instance, about the notion of "idiolect, relativized to periods of time". But if the *No-Language Postulate* does not tell us anything about idiolects, relativized to periods of time: why should they fall victim to Davidson's alleged exorcist appetite? There is obviously no reason why they should. It thus follows that Ramberg's thesis is wrong.

Bar-On/Risjord's uncritical acceptance of Ramberg's thesis explains, however, our puzzlement over their argument, for assuming Ramberg's thesis were correct, it would obviously be

---

[43] (Bar-On/Risjord 1992) comprises 27 pages; 22 of them are devoted to rejecting what Ramberg says Davidson says, and only 5 deal with Davidson directly.

true that the only thing that would remain from the exorcized concept of "language" would be fragmented *"individual utterances"* which an interpreter cannot but *"consider (…) in isolation"*. But if these utterances can be considered *in isolation* only, an empirical determination of their truth-conditions is impossible. This, in turn, implies that it is impossible to determine what particular evidence happens to confirm or disconfirm a radical interpreter's passing theory. Bar-On/Risjord are thus concluding that the *No-Language Postulate* is *isolating* the radical interpreter from the passing theory-confirming evidence. This is, given their wrong assumption, a correct conclusion. It is, furthermore, obviously this conclusion that makes them claim the *No-Language Postulate* to deprive the interpretation theories (i.e. the passing- and prior theories) of all empirical content and Davidson's truth-conditional approach of its entire subject matter.

But why do Bar-On/Risjord claim conventions/rules to be prerequisite for the existence of a criterion that distinguishes between evidence that confirms and disconfirms a truth-theory (= passing theory)? The answer is that on their mistaken understanding of the NDE-modification, the linguistic conventions/rules are the only means to reestablish the connection between the speaker's supposedly isolated utterances. Such a connection allows the radical interpreter to get in touch again with the non-linguistic evidence which makes such-and-such T-Sentence-hypothesis true (or false); and this, in turn, allows the radical interpreter to confirm or disconfirm his passing theories, and thus to interpret a speaker's linguistic behaviour along the lines Davidson's truth-conditional approach suggests.

However, Davidson does *not* exorcize the concept of "language". Accordingly, Davidson is *not* forced to regard his radical interpreter as being as isolated from the confirming/disconfirming evidence for his passing theory as Bar-On/Risjord fear him to be. As a consequence, it is wrong to suppose that the *No-Language Postulate* disconnects the radical interpreter from the evidence which confirms/disconfirms his particular passing theory. The harm which Bar-On/Risjord fear the *No-Language Postulate* to do to Davidson's semantics is, thus, illusionary. It follows that it is not necessary to withdraw it in order to save the theory of Radical Interpretation and the truth-conditional approach to semantics.

6.3.2 Karen Green's strategy of rejecting the No-Language Postulate by proposing an alternative approach to the interpretation of malapropisms

Karen Green agrees with Bar-On/Risjord that the *No-Language Postulate* is fundamentally wrong. She disagrees, however, that this is due to the allegedly resulting inexistence of the concept of "language". The problem is rather, she says, that the postulate is the result of the *superfluous* and *arbitrary* NDE-modification. The modification is *superfluous*, according to Green, for Davidson's theory of metaphor (cf. Davidson 1978) provides a sufficient means for treating non-conventional language uses within the pre-NDE theory of Radical Interpretation. It is, moreover, *arbitrary*, since Davidson never explains why *some* non-conventional language uses (namely: malapropisms) should *not* be treated as his theory of metaphor suggests. The remedy to these aberrations lies, she continues, in the withdrawal

of the NDE-modification. But to do so leads, as Green correctly observes, not only to the withdrawal of the *No-Language Postulate*; it shows at the same time that the only notion of "meaning" required for the interpretation of non-conventional language uses is the notion of *conventional* meaning.

Of course, understanding and evaluating Green's reasoning requires an understanding of Davidson's theory of metaphor. Let me begin, therefore, with a (very) brief recapitulation of its essential theses. These are, roughly, the following three: (a) that in order for there to be a metaphor, there must be a *relation* between the metaphorically meant sentence and that sentence's normal (conventional, literal) meaning; (b) that the said relation cannot be a relation between two types of *meaning*, and (c) that metaphor thus has to be an *extra-semantic* phenomenon (cf. Davidson 1978).

Thesis (a) follows from the following two facts: firstly, that a sentence is *recognizable* as a metaphor only if it is possible for the interpreter to realize that the speaker intends him to grasp something that *deviates* from that sentence's normal (conventional, literal) meaning. The metaphorically meant sentence has, secondly, to deviate in such a way from its normal (conventional, literal) meaning that it is still recognizable to the interpreter as a sentence of the *same* language; for if it is not, it is more likely to be understood as a normally (conventionally, literally) meant sentence of a hitherto unknown *other* language. The relation between a sentence's normal (conventional, literal) meaning and its metaphorical meaning (or aspect) is thus indispensable for there to be a metaphor (Davidson 1978, p. 249f).

Thesis (b) is where Davidson deviates from the traditional view on metaphor: the metaphor-constituting relations cannot be relations between two types of *meaning*, he says, for the only discernible types of metaphor-constituting relations are *similarity relations*. Things are, however, similar in endless different ways. It thus follows that an interpreter could detect the required metaphorical meaning of such-and-such a sentence only if there were a prior consensus between him and the speaker about what particular similarity relation he must be concerned about. The metaphor's metaphorical meaning, if there was any, would thus have to be *prior* to the similarity relation (Davidson 1978, 254). But since it *is* similarity relations that constitute metaphors, the notion of metaphorical meaning is an impossibility.

Thesis (c) is a consequence of thesis (b): if metaphor is not the result of a special notion of *meaning,* it has to be the result of something *extra-semantic.* But what exactly could this mean? It means that metaphors do not express propositions or state facts, but that they are rather, as Davidson says,*"like a picture or a bump on the head [that] makes us appreciate some fact (...)"* (Davidson 1978, p. 262). They are a kind of subjective personal experience, vague to a certain degree, and impossible to be conclusively captured or paraphrased. What makes us experience them as a *"bump on the head"* is, of course, their obvious falsity: "John is a pig" or "Sally is a block of ice" simply *cannot* be meant literally by a reasonable speaker, given that John and Sally are humans and, thus, neither pigs nor blocks of ice. What the speaker wants the interpreter to appreciate is, thus, something

that goes *beyond* or *above* the respective utterance's semantics; namely, something that the interpreter experiences as being somehow similar to pigs and blocks of ice.

Bearing these few facts about Davidson's theory of metaphor in mind, it is not difficult to see why Green considers it as an attractive alternative to the introduction of the NDE-modification: malapropisms (as well as every other non-conventional language use), if considered along the same lines as metaphors, are something *extra-semantic*, they are like a *"bump on the head"* that make us appreciate that they cannot be meant as what they literally, i.e. conventionally, mean.[44] But just as the interpretation (or better: appreciation) of a metaphor does not require a metaphorical meaning, the interpretation (appreciation) of a malapropism does not require first meaning. Thus turns the NDE-modification out as superfluous.

But is Green's idea convincing? It certainly is with respect to intentional malapropisms such as "Lead the way and we'll precede" or "The plane is landing momentarily", for interpreting them as meaning what they conventionally/literally mean obviously leads to extremely bizarre results that cannot be intended by a reasonable speaker. It is, thus, plausible to claim that the interpreter is supposed to look for an interpretation that goes beyond those sentences' conventional/literal meaning. His trying to find such a non-semantic interpretation will allow him to appreciate the speaker's humorous intention, and to understand that what the speaker intended to convey is something that has nothing to do with the uttered sentences' semantics, but rather with something else. However, Green's idea is not only convincing with regard to the interpretation of intentional malapropisms, but also with respect to other non-conventional language uses. The best of her own examples is, I think, the case of her son's speaking about Hitler's *Nasty Party* while summarizing the history of World War II to his schoolmates: being too young to know the accurate details of Nazi Germany and National Socialism, it is clearly implausible to suppose that he refers to the Nazi Party while uttering the words "Nasty Party"; it is far more plausible to suppose that he somehow (non-semantically) means the said party to be a very, very nasty party instead of reinterpreting his "Nasty Party" as first-meaning what the term "Nazi Party" means according to its conventional/literal meaning (cf. Green 2001, p. 242f and p. 250f).

There is, then, a certain amount of evidence supporting Green's thesis. So why not accept it as a sound objection to the NDE-modification? The reason is that Green seems to be mistaken about the motivation of Davidson's introduction of the modification and, as a consequence, concentrates on the wrong target: Green seems to believe that Davidson introduced the NDE-modification in order to

---

[44] My equating *conventional meaning* with *literal meaning* here sounds like a contradiction to everything that has been said in chapter 5. This impression is, however, wrong: a metaphor is recognizable as such only if the interpreter is already well acquainted with the speaker's linguistic behaviour. It follows that the interpreter is well acquainted with the linguistic conventions the speaker usually obeys. Accordingly, the literal meaning of a speaker's utterance is very likely to coincide with that utterance's conventional meaning. – What is important, however, is that this literal/conventional meaning coincidence is only *likely* to happen, but no conceptual necessity. Karen Green's thesis is thus not fully correct: replacing the NDE-modification with Davidson's theory of metaphor does not lead to an interpretation of malapropisms in terms of *conventional meaning*, but rather in terms of *literal meaning-that-is-likely-to-coincide-with-conventional meaning*. The mistake is, however, of minor importance and irrelevant to my refutation of her argument.

defend his anti-conventionalist convictions. Consequently, it is her first aim to show that all kinds of non-conventional language uses can be interpreted in terms of conventional meaning. This, however, is a far too one-sided perspective on the matter: for Davidson, the introduction of the NDE-modification is – if at all – only in a minor part an issue of backing his anti-conventionalist convictions. (He presented his reasons for this view already a few years earlier in *Communication and Convention* (cf. Davidson 1984).) Davidson's main aim is rather to show that the interpretation of non-conventional language uses and individual linguistic idiosyncrasies does not force him to abandon his theory of Radical Interpretation as a *compositional* theory; i.e. as a theory that fulfils what he himself has postulated as a criterion for the adequacy of a theory of natural language semantics at the beginning of his career (cf. chapter 2.1, p. 8f). As a consequence of this, it is not so central for Davidson whether this or that kind of non-conventional language use could be (or even should be) interpreted in the way he proposes to interpret metaphors. It would thus not undermine the NDE-modification if Karen Green were right that all malapropisms have to be interpreted as metaphors are.[45] As a consequence, her argument does not undermine the NDE-modified theory of Radical Interpretation. What it rather shows is, perhaps, that Davidson locates (intentional) malapropisms too far away from figurative language uses, and that he is wrong about where the boundary between figurative language use and non-figurative-but-non-conventional language use has to be drawn. But this is not substantial for Davidson's general aims. It shows neither the NDE-modification to be superfluous or arbitrary, nor the No-*Language Postulate* to be wrong.

6.3.3 Catherine Talmage's doubts about the replacement of the notion of literal meaning by the notion of first meaning

Catherine Talmage's objection (Talmage 1994 and Talmage 1996) differs from the arguments considered so far because it is not aimed at the *No-Language Postulate*, but rather at the *status* of the newly introduced notion of first meaning. Her objection is, more precisely, that (a) the notion of first meaning *coexists* rather than replaces the notion of literal meaning, and (b) that it is the notion of literal meaning rather than the notion of first meaning that has to be taken as semantically foundational to Davidson's NDE-modified theory of Radical Interpretation. Talmage's claims are obviously incompatible with Davidson's views. But what course of reasoning brings her to her conclusions?

The centrepiece of her reasoning consists in a pondering of the situation where Mrs Malaprop's "This is a nice derangement of epitaphs" is interpreted as first-meaning that such-and-such a piece of text is a nice arrangement of epithets. – This is obviously the case if both Mrs Malaprop and

---

[45] It is not clear if this is really correct, for Green concentrates exclusively on *intentional* malapropisms and completely ignores the *non-intentional* ones. The latter, however, do not seem to fit her thesis. The reason is that non-intentional malapropisms are not uttered with the intention to convey anything beyond or above the malapropism's literal/conventional meaning. Exactly such an extra-semantic intention is, however, essential for there to be something analogous to a metaphor, since if there is none, there is nothing that the speaker could intend the interpreter to appreciate besides the literal/conventional meaning of his uttered sentence. Cf. Glüer 1995, p. 82f.

her interpreter are well acquainted to this particular piece of Mrs Malaprop's idiosyncratic language use. But what happens, Talmage asks, if Mrs Malaprop, without the interpreter's knowing, is driven by misanthropist intentions and likes to satisfy her hatred by willingly misleading her interpreter? The natural answer is, she says, that Mrs Malaprop is likely to intend her utterance to mean what her interpreter does *not* assume her to mean. A particular nice occasion to do so arises if she utters the said malapropism whenever she comes across a piece of text which reminds her of a *nice derangement of epitaphs*. The obvious result of her doing so is that she intends her utterance to mean what it *literally means*, while her interpreter, ignorant of her misanthropist character, assumes her to mean what is her utterance's *first meaning* (i.e. that the relevant piece of text is a nice arrangement of epithets). – Talmage concludes from these considerations that Mrs Malaprops utterance might have two meanings: a literal meaning *and* a first meaning (Talmage 1994, p. 219f).

But does the example really provide compelling evidence? Talmage considers the possible objection that the absence of a communicative intention on Mrs Malaprop's part shows that there is *nothing* to be understood, and that this might be taken as an argument that her utterance involves no linguistic meaning at all. This objection is intuitive, Talmage agrees, but it is clearly wrong according to Davidson's own standards: Mrs Malaprop's utterance possesses a *literal meaning*, she says, for all that is required for there to be such a meaning is that Mrs Malaprop *holds it to be true* that such-and-such a piece of text is (like) a nice derangement of epitaphs. Exactly this condition is, however, fulfilled in the case at hand. Mrs Malaprop's utterance possesses, furthermore, a *first meaning*, for the interpreter supposes her to *intend to communicate* that such-and-such a piece of text is a nice arrangement of epithets. The latter intention is not present, Talmage concedes, but this is undetectable to the interpreter; accordingly, it is impossible for him to realize that the utterance's first meaning is only a *fake* meaning. Talmage thus concludes that the 'misanthropist Mrs Malaprop case' *is* compelling: her utterance has, as the example suggests, a literal meaning *and* a first meaning (Talmage 1994, p. 221). – However, it is obvious that Davidson's NDE-modified theory of Radical Interpretation cannot have *two* equally foundational notions of meaning. So which of the two is the truly foundational one? The answer depends, Talmage says, on the question whether Davidson still understands his theory of Radical Interpretation as the application of truth-conditional semantics to natural languages. If he does so, he has no real choice, for it is evident that he must choose that notion of meaning which is constituted by nothing but truth-conditions. This, however, is not the notion of first meaning, Talmage says, but the notion of literal meaning (Talmage 1994, p. 225).

Is Talmage's argument convincing? Hardly so. It rather appears to be the result of two independent errors; namely (a) an ill-conceived conclusion derived from the 'misanthropist Mrs Malaprop case', and (b) a misunderstanding regarding the nature of the two notions of meaning under consideration. Let me begin with (a): Talmage concludes from her thought-experiment that it is possible to perform a linguistic act without at the same time performing a communicative act. But is this really what Mrs Malaprop does? I have severe doubts, for if she did not intend her interpreter to

interpret her as first-meaning that such-and-such a piece of text is a nice arrangement of epithets, it would be impossible for her to satisfy her malicious desire to mislead him. It is, thus, more straightforward to regard Mrs Malaprop's linguistic act as a lie and, thus, as a *communicative act*, than to follow Talmage's opinion that it is a linguistic but not a communicative act. The 'misanthropist Mrs Malaprop case' therefore fails to back Talmage's claim.

But what about (b)? Are the notions of literal and first meaning as distinct as Talmage believes them to be? The decisive question is whether Talmage is right in claiming the notion of literal meaning to be free of communicative intentions and the notion of first meaning to be free of the ascription of truth-conditions. A short reflection reveals, however, that both these claims are wrong: Davidson has *always* considered his notion of *literal meaning* to be connected to the speaker's communicative intentions. The reason is that without such a connection, the truth-conditional approach would never have had the slightest chance to be successfully applied to natural languages. This is particularly obvious with respect to the *paradigmatic situation* of radical interpretation: what the interpreter aims at with his testing of T-Sentence-hypotheses is to discern what the alien speaker intends him to *understand*. The radical interpreter – even the one from the pre-NDE theory of Radical Interpretation – thus tries to discern the alien speaker's *communicative intentions* by means of determining the literal meanings of his utterances. Consequently, it is foundational to the theory of Radical Interpretation that it is possible to grasp the speaker's communicative intention(s) by means of grasping what the speaker is holding to be true. I should probably add that saying this does not entail the assumption that Davidson ever assumed the speaker's communicative intentions to be *restricted* to what the truth-conditions of his uttered sentences reveal. What Davidson claims throughout is only that there *is* a connection between *what is held to be true* and *what is intended to be understood*, and that this connection suffices, in principle, to describe the conceptual preconditions of language-understanding.

It is the very same reason why Talmage wrongly assumes the notion of *first meaning* to be exclusively about communicative intentions: an uttered sentence's truth-conditions remain what determines the shape of the post-NDE-interpreter's interpretation theories, for his passing and prior theories are, as I emphasized in subsection 6.1.1 (p. 80ff), exactly the same Tarski-style truth-theories as the pre-NDE-interpreter's interpretation theories. This is why I said earlier that the notion of first meaning, although diverging in many respects from literal meaning, is still *essentially* the same notion of meaning (cf. subsection 6.1.2, p. 83). Talmage's attempt to regard the notions of literal and first meaning as fundamentally distinct from each other is, thus, ill-founded. But if it is impossible to distinguish the two notions along the ways Talmage proposes, it is impossible to claim that they could coexist side by side. Accordingly, it is impossible to claim that it is not the notion of first meaning, but rather the notion of literal meaning that has to be regarded as foundational to the NDE-modified theory of Radical Interpretation. Talmage's argument against Davidson's NDE-modified theory of Radical Interpretation is thus invalid.

Having considered Davidson's NDE-modified theory of Radical Interpretation from all sides, it is now time to come back to the central question of this study: is it possible, as Davidson claims, to arrive at an adequate conceptual description of the possibility of language-understanding if one relies on a notion of linguistic meaning that is restricted to the truth-conditions of sentences? Before turning to this question in the subsequent (and concluding) chapter, allow me to close the present one by briefly summarizing the findings regarding the characteristics of Davidson's notion of first meaning.

6.4 The characteristics of the notion of first meaning

According to my judgment, the following five characteristics of the notion of first meaning deserve special emphasis:

(1) The semantically foundational notion of meaning in Davidsonian semantics (which is now called "first meaning" instead of "literal meaning") does not apply to rule or convention-governed shared languages, but to *idiolects relativized to periods of time*. This is the content of the (in)famous *No-Language Postulate*. Although it implies that shared languages are not prerequisite to language-understanding, it does *not* imply that shared languages are not possible within the Davidsonian framework: as I have tried to show, the notion of "shared language" that is questioned by Davidson is far more restricted than is often assumed by his opponents and interpreters (for instance: by Dummett, Hacking, Bar-On/Risjord and Ramberg).

(2) The semantically foundational notion of meaning in Davidsonian semantics has to be understood as something highly *volatile* and as being *subject to constant change*. It thus differs strictly from conventional and rule-governed notions of linguistic meaning (as well as from the pre-NDE notion of literal meaning).

(3) First meaning is as restricted to the assignment of truth-conditions of uttered sentences as was its predecessor (i.e. the notion of literal meaning). First meaning is thus the *essentially* same notion of meaning as the notion of literal meaning. Consequently, the introduction of first meaning should not be taken as a sign of Davidson's assigning his theory of Radical Interpretation a new function: its primary function remains to show his truth-conditional approach to semantics to be applicable to natural languages.

(4) The claim that first meaning is independent of linguistic rules and linguistic conventions implies the idea that there is *no relevant boundary between the knowing (interpreting, understanding, speaking, learning) of a language and the knowing of one's way around in the world in general*. In Davidson's NDE-modified theory of Radical Interpretation, the knowing of a language thus loses its exceptional position within the whole range of rational activities.

(5) The knowing (interpreting, understanding, speaking, learning) of a language is supposed to be an *essentially social activity*. But in contrast to most of his opponents, Davidson takes this social aspect to be anchored in the language-users intention to understand and to make oneself understood, rather than in the sharing of linguistic conventions and/or rules.

7 Concluding remarks

It is now time to return to the question with which this study began: is it possible, as Davidson claims, to arrive at an adequate conceptual description of the possibility of language-understanding if one relies on a notion of linguistic meaning that is restricted to the truth-conditions of sentences?

One way to answer this question is to scrutinize the evolution of this notion of meaning over the course of the development of the theory of Radical Interpretation. Davidson took the first steps towards this theory already in his essay *Truth and Meaning* (Davidson 1967), where the foundational notion of meaning was referred to by the name "literal meaning". Since Davidson presupposed at that time that speakers use their languages in very stable ways, he assumed the literal meanings of their uttered sentences to be determinable by *one single* Tarski-style truth-theory. The consequence of this assumption was that a sentence's (but likewise a phrase's or word's) literal meaning had to be taken to be very rigid and constant over time. It is not clear from Davidson's essays whether he really (and naively) believed natural languages to be used in such stable ways or whether he merely allowed himself to temporarily simplify the linguistic reality in order to nail down the key features of his theory of Radical Interpretation. But whatever his motive was: he removed this unjustified assumption in *A Nice Derangement of Epitaphs* (Davidson 1986), where he proposed a thorough modification of the theory of Radical Interpretation. Instead of assuming the radical interpreter to construct a single Tarski-style truth-theory to mirror the entire linguistic competence of such-and-such a speaker, Davidson now had the interpreter use *a constantly changing pair* of such theories.

The NDE-modification had its price: claiming that the radical interpreter works with pairs of constantly changing Tarski-style truth-theories implies that none of these theories, and at no point in time, ever mirrors the entire linguistic competence of a speaker. The NDE-modified theory of Radical Interpretation thus envisages representations of a speaker's linguistic competence that remain valid only over relatively short periods of time. This drawback is matched, however, by a far more valuable merit of the modification: it allows the radical interpreter to reorganize his interpretation theories in whatever directions and as often as he judges this to be necessary. The theory of Radical Interpretation thus loses its original static structure and now provides the flexibility which enables the radical interpreter to deal with whatever idiosyncrasies and inconstancies he may encounter in a speaker's linguistic behaviour.

The development from the static original theory to the flexible NDE-modified version does not remain without an effect on the theory's foundational notion of linguistic meaning: it leads, as we saw in chapter 6, to the replacement of the notion of literal meaning by the notion of first meaning (cf. p. 82). Both these notions are, to be sure, *truth-conditional* notions of meaning. They differ, however, in many other respects: the most obvious one is that a sentence *always* keeps the same literal meaning, while it changes its first meaning according to the particular way it is used. A sentence's literal

meaning is, thus, *constant*, while its first meaning is *volatile* (cf. sections 5.4, 6.1.1 and 6.4). The contrast between the constancy of a sentence's literal meaning and the volatility of its first meaning emphasizes how dramatic the development of the theory of Radical Interpretation actually is.

But this is not all: Davidson's replacement of the foundational notion of linguistic meaning also allows him to clarify what remained vague earlier or what his pre-NDE theory of Radical Interpretation did not deal with at all. A particularly obvious clarification concerns the question as to what notion of language his respective foundational notions of linguistic meaning are supposed to apply. Regarding the pre-NDE notion of *literal meaning*, Davidson tells us only the following:

(a) Literal meaning is no convention- or rule-governed notion of meaning (cf. chapter 5). Accordingly, literal meaning does not apply to any notion of "language" which is constituted thanks to the speaker-interpreter's obedience to a particular set of linguistic conventions/rules.

(b) Idiolects play an important role in the pre-NDE theory of Radical Interpretation; they do so because the radical interpreter is always supposed to interpret the linguistic behaviour of *individuals*.

These two facts are obviously insufficient to spell out the notion of "language" to which literal meaning should be taken to apply. It clearly does not apply to shared language in the sense that *"many philosophers and linguists have supposed"* (Davidson 1986, p. 107). But is literal meaning already supposed to apply to idiolects? The importance of idiolects for the radical interpreter might support this assumption, but it seems to me to rest on rather unstable grounds, since Davidson never showed signs that he doubted the importance of shared language prior to his *A Nice Derangement of Epitaphs*: there is not even a hint at such a view in his *Communication and Convention* (Davidson, 1984), where he presents his arguments against conventionalism and which was published just two years before *A Nice Derangement of Epitaphs*. The notion of "language" to which literal meaning applies thus remains pretty unclear. However, this odd situation is removed by the NDE-modification, since it is completely clear to what notion of language the newly introduced notion of first meaning applies to: the speaker's idiolects, relativized to periods of time.

Another aspect where a clarification during the development of the theory of Radical Interpretation takes place concerns the *social aspect of language* – or, as I had better say, the social aspect of *linguistic meaning*, for linguistic meaning is, after all, what the speaker and the interpreter are still supposed to share after the introduction of the *No-Language Postulate*. We cannot say much about the social aspect of the notion of literal meaning, since Davidson considers language-understanding in his pre-NDE theory of Radical Interpretation exclusively from the interpreter's perspective. This is why we know quite a lot about the interpreter's procedure (cf. section 2.2), but hardly anything about the speaker's role. The only thing we are justified to claim about the speaker is

that he attempts *not* to interfere with the interpreter's efforts to allocate the right truth-conditions to his (= the speaker's) uttered sentences. This would appear as a pretty *passive* and *vague* contribution to these sentence's respective literal meanings. The situation looks entirely different, however, as soon as the speaker becomes the interpreter's equally active partner in the NDE-modified theory of Radical Interpretation (cf. section 6.1.3, p. 85f): successful interpretation and communication is now no longer understood as the result of an immense effort on the part of the interpreter and a rather passive and vague contribution on the part of the speaker, but as the result of a mutual effort which requires active contributions from both sides. *"What matters to successful linguistic communication is"*, Davidson says,

> *the intention of the speaker to be interpreted in a certain way (...) and the actual interpretation of the speaker's words along the intended lines through the interpreter's recognition of the speaker's intentions (...)* (Davidson 2005b, p. 51).

The speaker is now supposed to actively interfere with the interpreter's proceedings. This, however, means that he is supposed to influence the ways the interpreter allocates truth-conditions to his (= the speaker's) uttered sentences – and this is obviously the same as to suppose that he (= the speaker) actively influences the constitution of the respective first meanings of his sentences. The speaker's contribution to the social aspect of meaning is thus no longer passive and vague; it is as active as the interpreter's contribution. Indeed, the speaker and interpreter are supposed to be *equal* partners – and this equality leaves its traces in the foundational notion of meaning in Davidson's theory of Radical Interpretation: speaker and interpreter both contribute to the constitution of a sentence's first meaning, and this is what embodies the social aspect of linguistic meaning in Davidson's NDE-modified theory of Radical Interpretation.[46]

A further interesting clarification concerns Davidson's removal of the boundary between knowing/learning a language and finding one's way around in the world in general. Davidson presents the dissolution of this boundary as a necessary consequence of the NDE-modification. This is certainly correct. Reflection reveals, however, that there may *never* have been such a boundary within Davidson's approach. Consider the field linguist's position in the paradigmatic situation of Radical Interpretation: this is, as we saw, a position of complete ignorance with respect to the alien speaker's linguistic behaviour. However, if the field linguist lacks specific knowledge about the alien speaker's linguistic behaviour, he is very likely to be just as ignorant with respect to the other cultural achievements of the alien community. But on this assumption, there is obviously no categorical difference between the field linguist's attempt to learn/understand the alien language and the field

---

[46] Saying this is *not* to say that the NDE-modified theory of Radical Interpretation abandons the interpreter-centred perspective: the speaker influences the radical interpreter's interpretation theories, but it is still exclusively the interpreter who constructs these theories. Accordingly, it is eventually the interpreter who, for better or worse, decides whether such-and-such an utterance first-means so-and-so.

linguist's attempt to learn/understand any of the alien community's non-linguistic activities (say, their games, rituals, and religious ceremonies). In order to learn/understand either, the interpreter has to begin from scratch. This is an interesting fact, I think, for it clarifies an aspect of the scenario of Radical Interpretation which Davidson never made explicit prior to *A Nice Derangement of Epitaphs*. It clarifies, namely, the nature of the setting from which the foundational notion(s) of meaning in Davidson's theory of Radical Interpretation emerge.

The NDE-modification of the theory of Radical Interpretation thus serves various purposes: it emphasizes what Davidson does not discuss in the early years, it clarifies what remains vague or unclear before, and – most importantly – it allows the radical interpreter to construct interpretation theories which are enormously flexible, and which allow him to do justice to whatever unexpected and unconventional language uses he might come across. I think that it is thus fair to conclude that the NDE-modified theory of Radical Interpretation allows for an adequate conceptual description of the possibility of language-understanding. Consequently, the central question of this study, in my opinion, has to be answered affirmatively: it *is* possible to arrive at an adequate conceptual description of language-understanding in terms of a truth-conditional notion of linguistic meaning; it is possible, more precisely, if the truth-conditional notion of linguistic meaning is Davidson's notion of *first meaning*.

There is, however, a proviso to this conclusion. This proviso is that there are *no* sentences that cannot be analyzed as having truth-conditions. Davidson's paratactic analysis shows that it is possible to ascribe truth-conditions to many more kinds of sentences than one might expect at first sight. This goes, for instance, for non-indicative sentences (cf. section 5.3), quotations (Davidson, 1979a), and for sentences in indirect speech (Davidson 1968). This clearly gives reason for hope; but since we cannot know whether there aren't any further kinds of recalcitrant sentences, there is no guarantee that the paratactic treatment will *always* be successful. There remains, thus, the danger that future discoveries could demand a re-evaluation of Davidson's attempt to apply his truth-conditional approach to the semantics of natural languages.

So far, I have only considered whether Davidson successfully arrives at his goal from within his own framework. That he does so is not only my point of view; it is conceded even by some of his most ardent adversaries (for instance by Strawson (1984), Dummett (1993) and Hacking (1984)). But what if one considers his work from a more distant, i.e. from an *external* perspective? Here, the agreement immediately ceases: considering Davidson's work from such a perspective implies considering whether his objective is actually an objective that is worth pursuing, and whether Davidson does so from justified preconditions. It is especially the latter point which is heavily disputed. Critical voices have been raised, for instance, with respect to the status Davidson assigns to the concept of truth (cf. chapters 3 and 4). Other frequent criticisms concern the Quinean heritage in Davidson's philosophy, particularly his holistic convictions and the idea that an approach to semantics should not make use of intensional concepts (such as communicative intentions). Some of these

criticisms – but not all – have been an issue on the last hundred or so pages. Although I hope to have showed them not to undermine Davidson's approach, it is certainly true that there remain open questions. One such open question concerns Davidson's conception of the concept of truth: it appears to me to be a very plausible conception – but it is, as I regretfully mentioned in section 4.4, rather just the sketch of a general idea than a subtly elaborated (let alone proved) thesis. A similar remark might be made about Davidson's holism: it clearly has its plausibility, but the fact that its correctness is still heavily disputed (cf. Fodor/Lepore 1992, 1993) shows that the last word has not yet been spoken.

However: to call into question what others take for granted is the philosophical activity *par excellence*. A philosophical debate about Davidson's work thus quite naturally involves the calling into question of his aims and presuppositions. Accordingly, one should not be impressed too much by the situation, as long as there is no *proof* that Davidson sets out with his project from untenable presuppositions, or that his theory involves inconsistencies or contradictions, or that the theory implies obviously wrong or absurd claims. That there actually is *no* such inconsistency, contradiction, falsity or absurdity to be found in Davidson's work is what I hope to have shown in the present study.

8 Bibliography

Aristotle 1993: Metaphysics, Book Γ, translated by Christopher Kirwan, Oxford

Avramides, Anita 1997: Intention and Convention, in: Hale/Wright 1997, pp. 60–86

Bar-On, Dorit; **Risjord**, Mark 1992: Is There Such a Thing as a Language?, in: Canadian Journal of Philosophy, No. 22, pp. 163–190

Bennett, Jonathan 1985: Critical Notices to Donald Davidson's "Inquiries into Truth and Interpretation", in: Mind, No. 94, pp. 601–626

Bennett, Jonathan 1990: Linguistic Behaviour, Indianapolis, Cambridge

Blackburn, Simon 1984: Spreading the Word, Oxford

Burge, Tyler 1979: Reasoning about Reasoning, in: Philosophia, No.70, pp. 651–656

Davidson, Donald 1965: Theories of Meaning and Learnable Languages, in: Davidson 2001, pp. 3–15

Davidson, Donald 1967: Truth and Meaning, in: Davidson 2001, pp. 17–36

Davidson, Donald 1968: On Saying That, in: Davidson 2001, pp. 93–108

Davidson, Donald 1970: Mental Events, in: Davidson 2001b, pp. 207–227

Davidson, Donald 1970a: Semantics for Natural Languages, in: Davidson 2001, pp. 55–64

Davidson, Donald 1973: Radical Interpretation, in: Davidson 2001, pp. 125–139

Davidson, Donald 1974: Belief and the Basis of Meaning, in: Davidson 2001, pp. 141–154

Davidson, Donald 1974a: On the Very Idea of a Conceptual Scheme, in: Davidson 2001, pp. 183–198

Davidson, Donald 1975: Thought and Talk, in: Davidson 2001, pp. 155–170

Davidson, Donald 1977: Reality Without Reference, in: Davidson 2001, pp. 215–225

Davidson, Donald 1978: What Metaphors Mean, in: Davidson 2001, pp. 245–264

Davidson, Donald 1979: Moods and Performance, in: Davidson 2001, pp. 109–121

Davidson, Donald 1979a: Quotation, in: Davidson 2001, pp. 79–92

Davidson, Donald 1979b: The Inscrutability of Reference, in: Davidson 2001, pp. 227–241

Davidson, Donald 1982: Rational Animals, in: Davidson 2001a, pp. 95–105

Davidson, Donald 1983: A Coherence Theory of Truth and Knowledge, in: Davidson 2001a, pp. 137–153

Davidson, Donald 1984: Communication and Convention, in: Davidson 2001, pp. 265–280

Davidson, Donald 1985: Incoherence and Irrationality, in: Davidson 2004, pp. 189–198

Davidson, Donald 1986: A Nice Derangement of Epitaphs, in: Davidson 2005, pp. 89–107

Davidson, Donald 1990: The Structure and Content of Truth, in: The Journal of Philosophy, Vol. 87, No. 6, pp. 279–238

Davidson, Donald 1992: The Second Person, in: Davidson 2001a, pp. 107–121

Davidson, Donald 1994: The Social Aspect of Language, in: Davidson 2005, pp. 109–125

Davidson, Donald 1994a: Radical Interpretation Interpreted, in: Philosophical Perspectives – Logic and Language, Vol. 8, pp. 121–128

Davidson, Donald 1997: The Emergence of Thought, in: Davidson 2001a, pp. 123–134

Davidson, Donald 1997a: Seeing Through Language, in: Davidson 2005, pp. 127–141

Davidson, Donald 1999: Reply to W. V. Quine, in: Hahn 1999, pp. 80–86

Davidson, Donald 2001: Inquiries into Truth and Interpretation, Oxford

Davidson, Donald 2001a: Subjective, Intersubjective, Objective, Oxford

Davidson, Donald 2001b: Essays on Actions and Events, Oxford

Davidson, Donald 2004: Problems of Rationality, Oxford

Davidson, Donald 2005: Truth, Language, and History, Oxford

Davidson, Donald 2005a: Truth and Predication, Cambridge (MA), London

Davidson, Donald 2005b: The Content of the Concept of Truth, in: Davidson 2005a, pp. 49–75

Davidson, Donald 2005c: What More Is There to Truth?, in: Davidson 2005a, pp. 29–48

Davidson, Donald; **Harman**, Gilbert (eds.) 1973: Semantics of Natural Language, Dordrecht

Donnellan, Keith 1966: Reference and Definite Descriptions, in: The Philosophical Review, Vol. 75, No. 3, pp. 281–304

Donnellan, Keith 1968: Putting Humpty Dumpty Together Again, in: The Philosophical Review, Vol. 77, No. 2, pp. 203–215

Dummett, Michael 1974: Frege. Philosophy of Language, London

Dummett, Michael 1975: What is a Theory of Meaning? (I), in: Dummett 1993, pp. 1–33

Dummett, Michael 1976: What is a Theory of Meaning? (II), in: Dummett 1993, pp. 34–93

Dummett, Michael 1978: Truth and Other Enigmas, Cambridge (MA)

Dummett, Michael 1978a: What do I Know when I Know a Language? in: Dummett 1993, pp. 94–116

Dummett, Michael 1989: Language and Communication, in: Dummett 1993, pp. 166–187

Dummett, Michael 1989a: "A Nice Derangement of Epitaphs": Some Comments on Davidson and Hacking, in: Lepore 1989, pp. 459–476

Dummett, Michael 1993: The Seas of Language, Oxford

Dummett, Michael 1993a: Mood, Force, and Convention, in: Dummett 1993, pp. 202–223

Dummett, Michael 1993b: Origins of Analytical Philosophy, Cambridge (MA)

Dummett, Michael 1994: The Logical Basis of Metaphysics, Cambridge (MA)

Dummett, Michael 1994a: Meaning, Knowledge, and Understanding, in: Dummett 1994, pp. 83–116

Dummett, Michael 1994b: Reply to Davidson, in: McGuinness/Oliveri 1994, pp. 257–267

Etchemendy, John 1988: Tarski on Truth and Logical Consequence, in: The Journal of Symbolic Logic, pp. 51–79

Evnine, Simon 1991: Donald Davidson, Stanford

Fodor, Jerry; **Lepore**, Ernest 1992: Holism. A Shopper's Guide, Oxford

Fodor, Jerry; **Lepore**, Ernest (eds.) 1993: Holism. A Consumer's Update, Amsterdam

Fodor, Jerry; **Lepore**, Ernest 1994: Is Radical Interpretation Possible?, in: Philosophical Perspectives – Logic and Language, Vol. 8, pp. 101–119

Fosnot, Catherine 2005: Constructivism: Theory, Perspectives, and Practics, New York

Frege, Gottlob 1891: Funktion und Begriff, in: Frege 1994, pp. 18–39

Frege, Gottlob 1892: Über Sinn und Bedeutung, in: Frege 1994, pp. 40–65

Frege, Gottlob 1918: Der Gedanke: Eine logische Untersuchung, in: Frege 1993, pp. 30–53

Frege, Gottlob 1993 (ed. by Günther Patzig): Logische Untersuchungen, Göttingen

Frege, Gottlob 1994 (ed. by Günther Patzig): Funktion, Begriff, Bedeutung, Göttingen

Glüer, Kathrin 1995: Relations and Transitions – An Interview with Donald Davidson, in: Dialectica, Vol. 49, No. 1, pp. 75–86

Green, Karen 2001: Davidson's Derangement: On the Conceptual Priority of Language, in: Dialectica, Vol. 55, No. 3, pp. 239–258

Grice, Paul 1948/1957: Meaning, in: Grice 1989, pp. 213–223

Grice, Paul 1967/1987: Utterer's Meaning and Intentions, in: Grice 1989, pp. 86–116

Grice, Paul 1967/1987a: Utterer's Meaning, Sentence-Meaning, and Word-Meaning, in: Grice 1989, pp. 117–137

Grice, Paul 1987: Retrospective Epilogue, in: Grice 1989, pp. 339–385

Grice, Paul 1989: Studies in the Way of Words, Cambridge (MA), London

Hacking, Ian 1984: On the Frontier, in: The New York Review of Books, 20th December 1984

Hacking, Ian 1989: The Parody of Conversation, in: Lepore 1989, pp. 447–458

Hahn, Lewis E. (ed.) 1999: The Philosophy of Donald Davidson. The Library of Living Philosophers, Vol. XXVII, Chicago

Hale, Bob; **Wright**, Crispin (eds.) 1997: A Companion to Philosophy of Language, Oxford

Hale, Jane 1997: Radical Interpretation, in: Hale/Wright: 1997, p. 175–196

Hanfling, Oswald 1980: Does Language Need Rules?, in: The Philosophical Quarterly, Vol. 30, No. 120, pp. 193–205

Harnish, Robert M. 1994: Mood, meaning and speech act, in: Tsohatzidis 1994, pp. 407–459

Joseph, Marc 2004: Donald Davidson, Chesham

Kirkham, Richard L. 1997: Theories of Truth. A Critical Introduction, Cambridge (MA)

Laakso, Aarre 2004: In Defense of Convention. Why Davidson's Objection Won't Work, published online: www.laaksmi.com/aare/cv/defcon.pdf; University of California at San Diego

Lepore, Ernest 1989: Truth and Interpretation. Perspectives on the Philosophy of Donald Davidson, Oxford

Lepore, Ernest 2004: An Interview with Donald Davidson, in: Davidson 2004, pp. 231–265

Lepore, Ernest; **Ludwig**, Kirk 2005: Donald Davidson. Meaning, Truth, Language, and Reality, Oxford

Lepore, Ernest; **Ludwig**, Kirk 2006: The Reality of Language: On the Davidson/Dummett Exchange, published online: www.ruccs.rutgers.edu/faculty/LeporeSelPub.html, Rutgers University

Lepore, Ernest; **Ludwig**, Kirk 2007: Donald Davidson's Truth-Theoretic Semantics, Oxford

Lewis, David 1968: Languages and Language, in: Lewis 1984, pp. 163–188

Lewis, David 1969: Convention. A Philosophical Study, Cambridge (MA)

Lewis, David 1973: General Semantics, in: Davidson/Harman 1973, pp. 169–217

Lewis, David 1984: Philosophical Papers, Vol I, Oxford

Linsky, Leonard (ed.) 1952: Semantics And The Philosophy of Language, Urbana

Ludwig, Kirk (ed.) 2003: Donald Davidson, Cambridge

Lycan, William G. 2000: Philosophy of Language. A contemporary introduction, London

MacKay, Alfred F. 1968: Donnellan and Humpty Dumpty on Referring, in: The Philosophical Review, Vol. 77, No. 2, pp. 197–202

Margalit, Avishai (ed.) 1979: Meaning and Use, Dordrecht

McDowell, John 1980: Meaning, Communication, and Knowledge, in: McDowell 2001, pp. 29–50

McDowell, John 2001: Meaning, Knowledge, and Reality, Cambridge (MA)

McGuinness, Brian; **Oliveri**, Gianluigi (eds.) 1994: The Philosophy of Michael Dummett, Dordrecht

Phillips, Denis C. (ed.) 2000: Constructivism in Education. Opinions and Second Opinions on Controversial Issues, Urbana

Putnam, Hilary 1985/86: A Comparison of Something with Something Else, in: New Literary History, No. 17, pp. 61–79

Quine, Willard V. O. 1948: On What There Is, in: Quine 1999, pp. 1–19

Quine, Willard V. O. 1951: Two Dogmas Of Empiricism, in: Quine 1999, pp. 20–46

Quine, Willard V. O. 1960: Word and Object, Cambridge (MA)

Quine, Willard V. O. 1979: Comments, in: Margalit 1979, pp. 21–22

Quine, Willard V. O. 1980: Things and Their Place in Theories, in: Quine 1981, pp. 1–23

Quine, Willard V. O. 1981: Theories and Things, Cambridge (MA), London

Quine, Willard V. O. 1999: From A Logical Point Of View. Nine Logico-Philosophical Essays, Cambridge (MA), London

Ramberg, Bjørn 1989: Donald Davidson's Philosophy of Language. An Introduction, Oxford

Ramsey, Frank 1927: Facts and Propositions, in: Proceedings of the Aristotelian Society, No. 7, Vol. 1, pp. 153–170

Rorty, Richard 1986: Davidson and Truth, in: Lepore 1989, pp. 333–353

Rumfitt, Ian 1995: Truth Conditions and Communication, in: Mind, New Series, Vol. 104, No. 416, pp. 827–862

Schiffer, Stephen 1972: Meaning, Oxford

Schiffer, Stephen 1987: Remnants of Meaning, Cambridge (MA)

Soames, Scott 1984: What is a Theory of Truth?, in: The Journal of Philosophy, No. 81, pp. 411–429

Strawson, Peter F. 1964: Intention and Convention in Speech Acts, in: Strawson 1971, pp. 149–169

Strawson, Peter F. 1969: Meaning and Truth, in: Strawson 1971, pp. 170–189

Strawson, Peter F. 1971: Logico-Linguistic Papers, London, New York

Strawson, Peter F. 1984: Hot Pursuit, in: The London Review of Books, Vol. 6, No. 13, 19th july 1984, p. 17

Talmage, Catherine J. L. 1994: Literal Meaning, Conventional Meaning, and First Meaning, in: Erkenntnis No. 40, pp. 213–225

Talmage, Catherine J. L. 1996: Davidson and Humpty Dumpty, in: Noûs, Vol. 30, No. 40, pp. 537–544

Tarski, Alfred 1944: The Semantic Conception Of Truth And The Foundations Of Semantics, in: Linsky 1952, pp. 13–47

Tsohatzidis, Savas L. (ed.) 1994: Foundations of speech act theory. Philosophical and linguistic perspectives, London, New York

Updike, John 1997: Toward the End of Time, London

Weiss, Bernhard 2002: Michael Dummett, Princeton, Oxford

Wheeler, Darell 2003: On Davidson, New York

Wilson, Neil L. 1959: Substances Without Substrata, in: Review of Metaphysics, No. 12, pp. 521–539

Wittgenstein, Ludwig 1952/2009: Philosophical Investigations, translated by G. E. M. Anscombe, P. M. S. Hacker and J. Schulte, Oxford

Wood, Allen 2002: Relativism, published online: www.stanford.edu/~allenw/recentpapers.htm; Stanford University