

Journal for the History of Analytical Philosophy

Volume 7, Number 2

Editor in Chief

Marcus Rossberg, University of Connecticut

Editorial Board

Annalisa Coliva, UC Irvine

Henry Jackman, York University

Kevin C. Klement, University of Massachusetts

Consuelo Preti, The College of New Jersey

Anthony Skelton, Western University

Mark Textor, King's College London

Audrey Yap, University of Victoria

Richard Zach, University of Calgary

Editor for Special Issues

Sandra Lapointe, McMaster University

Review Editors

Sean Morris, Metropolitan State University of Denver

Sanford Shieh, Wesleyan University

Design and Layout

Daniel Harris, Hunter College

Kevin C. Klement, University of Massachusetts

ISSN: 2159-0303

jhaponline.org

© 2019 Peter Pagin

Compositionality in Davidson's Early Work

Peter Pagin

Davidson's 1965 paper, "Theories of Meaning and Learnable Languages", has (at least almost) invariably been interpreted, by others and by myself, as arguing that natural languages must have a compositional semantics, or at least a systematic semantics, that can be finitely specified. However, in his reply to me in the Żegleń volume, Davidson denies that compositionality is in any need of an argument. How does this add up?

In this paper I consider Davidson's first three meaning theoretic papers from this perspective. I conclude that Davidson was right in his reply to me that he never took compositionality, or systematic semantics, to be in need of justification. What Davidson had been concerned with, clearly in the 1965 paper and in "Truth and Meaning" from 1967, and to some extent in his Carnap critique from 1963, is (i) that we need a *general theory* of natural language meaning, (ii) that such a theory should not be in conflict with the learnability of a language, and (iii) that such a theory should *bring out* how knowledge of a finite number of features of a language suffices for the understanding of all the sentences of that language.

Special Issue:

Donald Davidson: Looking Back, Looking Forward

Edited by Claudine Verheggen

Compositionality in Davidson's Early Work

Peter Pagin

1. Introduction

In the Summer of 1995, Urszula M. Żegleń, aided by Ernie Lepore, organized a conference outside Kazimierz Dolny, near Lublin, in Poland.¹ The topic was the philosophy of language and mind of Donald Davidson. Professor Davidson himself was there, and he responded generously to the talks.²

My talk at the conference, like the paper based on it (Pagin 1999), concerned the possibility of justifying *compositionality* on the basis of the principles governing *radical interpretation*, in particular *the principle of charity*. In the talk, I argued that such justification would not work. The principles of radical interpretation fall short of justifying the principle.

The function version of the principle of compositionality can be informally stated as

(PCF) The meaning of a complex expression is a function of the meanings of its parts and its mode of composition.

The principle has the form a proposition, but leaves unspecified what language is concerned, what the syntactic theory, and thereby the constituent structure, for the language is, and which semantic values are intended by "meaning". All these parameters must be fixed before we have a proposition with a truth value.

The *substitution* version can be informally stated as

¹It was also thanks to Ernie that I was able to take part.

²Żegleń (1999) is a collection of papers from the conference, together with replies by Professor Davidson, edited by Professor Żegleń.

(PCS) If in a complex expression A a part e is replaced, in one or more occurrences, by a part e' that has the same meaning as e , the resulting expression A' has the same meaning as A .

For the purposes of the current paper, these two principles can be taken as equivalent.³

Justification here amounts to providing reasons for the more general claim that natural languages do have a compositional semantics. That is, there may be overarching features of language and human communication that can provide reasons for believing that natural languages *in general* have compositional semantics.

The background for raising the question was, firstly, that successful linguistic communication, by means of using new sentences, does provide such a justification, and, secondly, that within the framework of radical interpretation, this justification is not available.

As for the first part, the idea is that the assumption that natural language meaning is compositional best explains how a hearer manages to understand what a speaker means, even when the sentence used is new to the hearer (and perhaps also to the speaker).⁴ The idea is closely related to the argument by Frege in the famous opening paragraph of "Compound Thoughts":

³For more precise statements and discussion of the relation, see Pagin and Westerståhl (2010a).

⁴In fact, what is required for the possibility of *working out* the meaning of a new sentence is that the semantics is *recursive*: a semantic function μ is *recursive* (in a sense analogous to that of recursive functions on natural numbers) iff the μ value of a complex expression is a *recursive* function of the μ values of the parts, the parts themselves, and the mode of composition. See Pagin and Westerståhl (2010a,b). Note that since *the parts themselves* occur as arguments, a recursive semantics need not be compositional.

The further reason for desiring that the semantics be not only recursive but also compositional is that compositionality is a necessary condition for *low computational complexity*. In this sense, a compositional semantics offers the *best* explanation. See Pagin (2012a,b, 2016, 2017).

It is astonishing what language can do. With a few syllables it can express an incalculable number of thoughts, so that even a thought grasped by a terrestrial being for the very first time can be put into a form of words which will be understood by somebody to whom the thought is entirely new. This would be impossible, were we not able to distinguish parts in the thought corresponding to the parts of a sentence, so that the structure of the sentence serves as an image of the structure of the thought. (Frege 1923, 1)

As for the second part, it was early on an element in Davidson's discussion of interpretation that whether two speakers communicate successfully, or speak the same language, is to be determined by applying radical interpretation itself:

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

(Davidson 1973, 125)

If it has to be determined by means of radical interpretation (with the principle of charity) whether linguistic communication is successful, and the general format of a semantic theory for a natural language is part of what goes into the radical interpretation project, then it has to be determined whether the semantics be compositional or not *before* radical interpretation is even applied, and hence *before* it is determined that communication is successful.

The Frege style argument from communication is therefore not available from Davidson's point of view. Rather, that semantics must be compositional has to be a feature of radical interpretation already in place. I argued in the talk that features of radical interpretation do not suffice for such a justification. In a sense, the contrary situation obtains: it is easier to satisfy the requirement of the principle of charity if the semantics is *not* constrained by having to satisfy compositionality.

Professor Davidson's response at the conference, as far as I remember it, had three parts. The first part gave a background to his interest in having a general, compositional, semantic framework. He said he had with interest been following the debate in *Analysis* over belief sentences, starting with Church (1950), and he had been wondering what the general *rules* were for getting the semantics for belief sentences *right*. A few years later he had come across Tarski's work on truth (Tarski 1935; 1983, 152–278) and realized that he had found what he was looking for: any semantic analysis should fit into the format Tarski had provided. This was the test. The second part was that it had not occurred to him that this choice needed any further *justification*. He had simply been delighted at finding a framework that did what he had asked for. The third part was concerned with conventions and the distinction between prior and passing theories (in Davidson 1986; this question need not concern us here). At least, this is my recollection of the response.

The written reply to the paper (Davidson 1999) still bears a fair bit of resemblance to the immediate response (as I remember it). But in the written reply, Davidson has taken the further step of explicitly *rejecting* the demand for justification of compositionality. The picture I had given of starting out with the idea of radical interpretation, and then going on to justify compositionality on that basis, is simply at odds with how he, Davidson, thought of radical interpretation. The reply ends as follows:

Clearly, on my understanding of radical interpretation, compositionality is built in at every stage. Does Pagin perhaps think, as some of my other commentators seem to think, that for me charity just enjoins us to make as good sense of each utterance, taken in isolation, as we can? Or that charity simply means 'maximise truth'? I urge such readers to reread the introduction to *Inquiries into Truth and Interpretation* (1984a). The reason why compositionality and 'radical interpretation' cannot ever conflict, given my understanding of radical interpretation, is that I view it as given that any theory of meaning is compositional, and then, and only, then, ask how we can tell that a speaker is speaking in accord with a specific compo-

sitional theory. It is only at this point that radical interpretation has a role to play. This picture may not always have been as plain in my writings as it should have been, but it has always been there, and it has certainly been the position I have been at pains to emphasise in all my writings in recent years. (Davidson 1999, 70; all emphasis, references, and quotation, in the original.)

This reply is remarkable in at least two respects. Firstly, why shouldn't the question of the justification of compositionality arise in the first place? That Davidson took it as an integral part of the idea of radical interpretation does not entail that he was *justified* in doing so. Is there any further reason in Davidson's framework for thinking that the question simply should not arise?⁵

Secondly, Davidson is the author of one of the standard arguments for compositionality, *the learnability argument*, presented in Davidson (1965). Davidson has been standardly interpreted, by myself and all or virtually all other interpreters, as providing an argument for compositionality. If he did, why did he reject the question in his reply? And if he didn't provide an argument for compositionality, what was he doing instead? We shall return to this question.

In the following, I shall be concerned with three papers by Davidson from the 1960s (Davidson 1963, 1965, 1967), where

⁵It seems that the answer to this question is negative. Rather, that natural language has (loosely) compositional semantics seemed to be taken for granted in Davidson's philosophical context, and thus did not appear to require an argument.

An anonymous reviewer draws attention to another passage of the Reply:

I assume (on the basis of learnability, creativity and common sense) that any successful interpretation of an utterance must fit that utterance into a compositional scheme. (Davidson 1999, 69)

This passage is, perhaps contrary to first appearance, perfectly consonant with the interpretation put forward in the paper. In this passage Davidson talks about *the format of the output* of radical interpretation, i.e., a meaning theory. Such a theory must instantiate a "compositional scheme". This is precisely what I argue for in the present paper.

the role of compositionality and the format of Tarskian truth theories emerge. These papers are discussed in a section each. In the concluding section, I summarize the findings.⁶

2. The Method of Extension and Intension (1963)

Most of Davidson (1963), Davidson's long contribution to the Carnap Schilpp volume, is an extensive criticism of the peculiarities of Carnap's semantic system in *Meaning and Necessity* (1947/1956). We need not concern ourselves with the details of designation, L-designation, etc. However, the fifth and sixth sections are of considerable interest in the present context. They also show that the reference to the *Analysis* debate over belief sentences was not only of autobiographical interest.

In Section V of the paper, Davidson discusses Carnap's analysis of belief sentences and the criticism by Alonzo Church (Church 1950). Carnap's account (1947/1956, 61–62) is exemplified below. Here (b) is the analysis of (a):

- (Carnap)
- a. John believes that D
 - b. There is a sentence \mathfrak{S}_i in a semantical system S' such that (a) \mathfrak{S}_i is intensionally isomorphic to "D" as a sentence of English and (b) John is disposed to an affirmative response to \mathfrak{S}_i as a sentence of S' .

⁶The term "compositionality" has very few occurrences in the Davidson corpus, and I think most of them actually are in the Reply. The main reason is that the term was not established at the time. It was introduced into the literature in Katz and Fodor (1963) (concerning understanding), and did not gain wider currency until the mid-seventies, partly thanks to merging the term with Montague-style semantics. The term was finally canonized in Partee (1984), which appeared after all the papers in Davidson (1984).

Davidson's *idea* of compositionality, as it appears in the papers considered here, is closely related to the canonical format (PCF), but is not the same. In particular, the distinction between compositional and recursive semantics is not reflected in Davidson's formulations. These aspects will not, however, be at issue here.

The relation of intensional isomorphism between linguistic expressions is roughly as follows: two atomic expressions are intensionally isomorphic iff they are co-intensional, and two complex expressions are intensionally isomorphic iff they have the same structure and corresponding parts are intensionally isomorphic. The formal definition is given at (Carnap 1947/1956, 59), where Carnap uses his technical notion of L-equivalence rather than co-intensionality. L-equivalence between two sentences holds just in case they are true in exactly the same state descriptions (Carnap 1947, 10–11).

Carnap's account has the crucial property that the analysis does not *say* what the belief content is, but only *mentions* the embedded sentence. The loss of information is stressed by Alonzo Church (1950), who brings out the point by appealing to the Langford translation test. Basically, the objection is the following: "D" is an English sentence in (a) and (b). In a translation to German, it should be translated in (a), since it is *used* there, but not in (b), since it is only *mentioned* there by being quoted, in order to preserve truth value. But that means that a German speaker who does not understand English will not, despite knowing that the German translation of (b) is true, know what John believes, in contrast to knowing the truth of a translation of (a). Church finishes up with an extra twist: if we relax the translation relation by way of translating quoted expressions as well as used expressions, this problem is eliminated, but then a German translation will not be intensionally isomorphic to the English original, or to a translation into French, since the translations in the different languages contain reference to *different* sentences, the respective translations of "D".⁷

Davidson objects to Church's criticism as follows. We start out with a belief sentence in English:

⁷Compare "The word 'dog' has three letters", translated into German as "Das Wort 'Hund' hat drei Buchstaben", where a truth is translated into a falsehood.

- (1) Plato believed that the realm of ideas is real.

We construct Carnap's analysis of (1):

- (2) There is a sentence S_i in a semantical system S' such that (a) S_i is intensionally isomorphic to "The realm of ideas is real" as an English sentence and (b) Plato was disposed to an affirmative response to S_i as a sentence of S' .

We then translate (1) into German:

- (3) Platon glaubte dass das Ideenreich wirklich ist.

Likewise, we construct in German the analysis of (3):

- (4) Es gibt einen Satz S_i in einem semantischen System S' für den gilt dass (a) S_i mit "Das Ideenreich ist wirklich." als einem deutschen Satz intensional isomorphisch ist und (b) Platon zu einer affirmativen Reaktion auf S_i , als einem Satz in S' , geneigt war.

Davidson then proceeds as follows (I have replaced his sentence naming by my own):

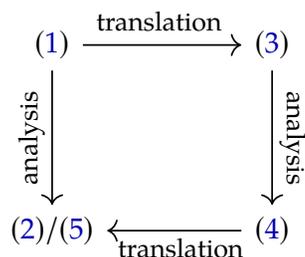
None of the force of Church's first objection to Carnap's analysis will be lost if we proceed to reason as follows: the meaning of [(1)] will be preserved if it is translated into another language, say German (let us call the resulting sentence [(3)]); next we may apply Carnap's analysis to [(3)] to produce an interpretation (in German) of [(3)] (call this [(4)]); finally, we translate [(4)] back into English. The result will be:

- [(5)] There is a sentence S_i in a semantical system S' such that (a) S_i is intensionally isomorphic to 'Das Ideenreich ist wirklich' as a German sentence and (b) Plato was disposed to an affirmative response to S_i as a sentence of S' .

Church contends that if Carnap's method of analysis were correct, [(5)] would 'convey the same information' as [(2)]. In particular he remarks that [(2)] and [(5)] are not intensionally isomorphic.

The demand that [(5)] and [(2)] be intensionally isomorphic is unreasonable since [(1)] and [(2)] obviously are not intensionally isomorphic. According to Carnap's view of analysis, an analysis is true and interesting if and only if analysans and analysandum are L-equivalent but not intensionally isomorphic. Since [(2)] is intended as an analysis of [(1)], it is presumably not intended to be intensionally isomorphic to it. (Davidson 1963, 343)

We can picture the situation with the following diagram:



From Church's point of view, the problem is that (5) and (2) are not intensionally isomorphic, because containing reference to different sentences. Davidson objects to Church that the requirement of isomorphism here is "unreasonable". Davidson's idea seems to be that since the relation between analysans and analysandum must be one of L-equivalence and must *not* be one of intensional isomorphism, there is no basis for requiring intensional isomorphism between two distinct acceptable analyses. Rather, we should settle for the relation of L-equivalence between (2) and (5), a condition that is met.⁸

As far as I understand it, this means that (5) should be seen as an equally good analysis of (1) as (2) is. This could be seen as an unwelcome result, since someone who understands English but does not understand German will understand (2) but not (5). But Davidson rejects this objection as well, saying: "The fact that someone who understood [(2)] might not understand [(5)] is no more against Carnap's analysis than the fact that someone who

⁸Putnam (1954, 115–16) similarly objects to Church that two analyses can be regarded as correct even if they are not intensionally isomorphic.

understood [(1)] might not understand [(2)]" (Davidson 1963, 345).

After a brief discussion of a related objection by Church, concerning iterated belief sentences, Davidson sums up the situation as follows:

In the light of the foregoing considerations it seems that Church has presented no reasons for rejecting Carnap's analysis of belief sentences, provided Carnap's criteria of a successful and correct analysis are accepted. (Davidson 1963, 346)

What exactly Carnap's general criteria are of an adequate analysis is not completely clear. Davidson is certainly right in claiming that the analysans must be L-equivalent with the analysandum, but perhaps not in adding that *in addition failing to be intensionally isomorphic* is sufficient for adequacy. Judging from Carnap's general aims and his remarks about the problem of belief sentence analysis (Carnap 1947, 53), we can distill two conditions on belief sentence analysis:

- (C-con) i) The analysans must be L-equivalent with the analysandum.
- ii) The analysans must make explicit the conditions under which substitutions in the belief context of a belief sentence preserve truth.

Let's assume that condition (i) is met (neither Church nor Davidson objects to it). We may further assume regarding (ii) that it is a sufficient condition for the preservation of truth value through a substitution in the belief context that the two embedded sentences are intensionally isomorphic. That is

- (Sub) If $\ulcorner p \urcorner$ and $\ulcorner q \urcorner$ are intensionally isomorphic, then the sentences $\ulcorner X \text{ believes that } p \urcorner$ and $\ulcorner X \text{ believes that } q \urcorner$ agree in truth value.

We can further assume that the condition is necessary in the sense for any true sentence $\ulcorner X \text{ believes that } p \urcorner$ we can find some

sentence $\ulcorner q \urcorner$ which is L-equivalent but not intensionally isomorphic with $\ulcorner p \urcorner$ such that $\ulcorner X$ believes that $q \urcorner$ is false.

On those assumptions, Carnap's analysis does satisfy (C-con). Davidson's objection to Church seems to be that Church imposes a misplaced *further* condition of informativeness of the analysis. In a very weak form, it can be stated as:

(I-con) If S' is an analysis of S in a language L , and X is a speaker of L , then if X knows that S is true, then X knows that S' is true.

This condition is not met by Carnap's analysis, since (1) is an English sentence, (5) is an English sentence that is an analysis of (1), but a speaker who knows English but not German may know that (1) is true and not know that (5) is true.

It is not completely clear to me whether Davidson takes the (I-con) condition to be flawed *in general* or whether Church is at fault for imposing a condition that is *irrelevant to Carnap's aims*. Either way, Davidson does raise the question what the rules for getting the analysis right are, and claims implicitly that, one way or another, Church has gotten them wrong.

Nevertheless, Davidson agrees with Church that Carnap's analysis has a badly motivated feature. He continues:

What Church has demonstrated is that Carnap has not given a semantical analysis of belief sentences in the sense in which he has given a semantical analysis of intensional contexts. For intensional contexts, Carnap lays down rules for the interchange of expressions based on the semantic relations between those expressions and entities; to paraphrase Carnap, he applies his method of meaning analysis to the designators within intensional sentences in order to show how out of the meanings of other expressions the meanings of sentences are constituted. Although Carnap's analysis of belief sentences makes use of the semantical notion of intensional isomorphism, it does not provide a semantical analysis of sentences like (42), [(1)] and (46) in the sense of showing how out of the meanings of the expressions of less than sentential scope the meanings of the sentences are constituted. Rather the analysis translates such

sentences as wholes into other sentences to which, *then*, Carnap's full semantical analysis (in terms of the method of extension and intension) may be applied. (Davidson 1963, 346)

On one way of understanding it, Davidson's objection here is that Carnap's analysis (2) of (1) does *not* provide a *complete* semantic analysis of (1). In order to arrive at the extension or the intension of (1), we would need to proceed to apply Carnap's semantics *separately* to the embedded sentence "The realm of ideas is real" as an English sentence (or to "Das Ideenreich ist wirklich" as a German sentence, or ...). The truth-conditions of (2) do not depend on what meaning Carnap's semantics gives the sentence.

We can bring out the deficiency as follows. In possible situation A , Plato believed (in the intuitive sense of "believe") that the realm of ideas is real, and was disposed to an informative response to a Greek sentence S that means that the realm of ideas is real. In situation A , furthermore, the English sentence "The realm of ideas is real" has a meaning different from its actual meaning and is *not* intensionally isomorphic to S , but *is* intensionally isomorphic to a Greek sentence S' that means *that the god of birds is fast*. In situation A , Plato was not disposed to an affirmative response to any Greek sentence that means that the god of birds is fast.

In possible situation B , Plato did *not* believe that the realm of ideas is real, but he did believe *that the god of birds is fast*, and was disposed to an affirmative response to a Greek sentence S' that means that the god of birds is fast. In situation B , again, the sentence "The realm of ideas is real" is intensionally isomorphic to S' .

Now, intuitively, (1) is true with respect to A and false with respect to B . The analysis (2), by contrast, is true with respect to B and false with respect to A . The result is clearly unwelcome. This brings out the fact that in Carnap's analysis the interpretation of the embedded sentence is *decoupled* from the truth-conditions of the analysis of the belief sentence. If we let the meaning of the

embedded sentence vary, the truth value of the belief sentence analysis (2) will follow the *sentence* “The realm of ideas is real”, not its *current meaning*.

In the quote from (1963, 346), as I understand it, Davidson complains that Carnap’s analysis does not show how the meaning of a belief sentence depends on the meanings of *parts* of the embedded sentence. This complaint agrees with the interpretation of the passage just given. The objection could be understood as a complaint that Carnap’s analysis is not compositional. It is clear, however, that in one sense, Carnap’s analysis *is* compositional. Carnap proposes to use “synonymous” as synonymous with “intensionally isomorphic” (Carnap 1947, 56). Hence, there is a corresponding notion of *meaning*, shared between any two expressions that are intensionally isomorphic. As is stated in (Sub) above, Carnap’s analysis satisfies (PCS), the substitution version of compositionality. Nevertheless, Carnap’s analysis is incomplete, since it falls short of telling us what the meaning of the belief sentence is in terms of the *actual* meaning of the embedded sentence. All that is required is that there is *some* meaning that makes the embedded sentence intensionally isomorphic to a sentence Plato was disposed to affirm.

However, the way Davidson proceeds after the quoted passage suggests that he is tracking a different idea. For Davidson goes on to emphasize the contrast between *direct* interpretation of a natural language sentence and the indirect interpretation of the sentence via a direct interpretation of a *transformation* or *regimentation* of that sentence. His complaint against Carnap, understood in this light, is then that Carnap does not provide any good reason for treating modal sentences by the direct method and belief sentences by the indirect method. In particular, Davidson points to Carnap’s criticism of Russell’s theory of descriptions, according to which definite descriptions do not have any independent meaning; only transformations of sentences containing them do (Carnap 1947, 140–41). Davidson complains that Carnap does not provide any good reason for turning to the indirect method himself in the case of belief sentences.

Either way, it is clear that Davidson, at the end of the section and in the final section of Davidson (1963), is calling for a *general theory of language* for choosing analyses of individual constructions, such as the “believe-that” construction. The direct method, in which the truth-conditions of an entire sentence are derived directly from the meanings of the parts, is in general preferable, but this does not hold for artificial constructions that we find neither in natural language nor in science and the accommodation of which may have negative effects more generally (quantification into modal contexts; Davidson 1963, 348–49). Davidson writes in the final paragraph:

But in the end it would seem that a general question of policy must be raised. In attempting to achieve generality, any theory of language will have the task of deciding which problems, which sorts of sentences and contexts, are to be dealt with by head-on methods, and which are better handled by preliminary transformations, translations and analyses. Even the decision to transform ‘John laughed and cried’ into ‘John laughed and John cried’ before confiding it to the formal system represents, on a low level, such a decision. More serious are the decisions to accept whole classes of sentences (perhaps the simple modalities, or belief sentences) as appropriate for treatment *without* gross transformation. In systems in which intensional entities are enlisted as meanings, or non-garden varieties of truth are introduced, decisions directly to accommodate one or another area in the total linguistic territory become especially binding since they influence the interpretation not only of the contexts they were devised to explicate but also the interpretation of other contexts. (Davidson 1963, 349)

The call for a general theory as a method for deciding individual cases seems to be the upshot of the paper. Summing up the relevant ingredients of Davidson’s paper, we can say that

- (1963) Davidson
- i) requires that analyses of natural language constructions should be made on the basis of a general theory of language;

- ii) requires that analyses of natural language constructions by default should show how the meaning of a sentence depends on the meanings of its parts;
- iii) allows an indirect method of interpretation only when it is motivated by general concerns.

That the general method should be compositional, in the sense of “showing how out of the meanings of the expressions of less than sentential scope the meanings of the sentences are constituted”, appears to be taken for granted. It is neither emphasized, nor argued for.

3. Theories of Meaning and Learnable Languages (1965)

The second part of Davidson (1965) is concerned with showing that certain semantic theories have the effect of rendering a language they would be true of unlearnable. Davidson’s examples are theories of quotation, Scheffler’s and Quine’s theories of belief sentences, and Church’s infinite hierarchy of sense and denotation in Church (1951). The languages would be unlearnable because they would have an infinite number of semantic primitives.

The connection between learnability and finiteness is spelled out early in the paper in the following well-known passage:

When we can regard the meaning of each sentence as a function of a finite number of features of the sentence, we have an insight not only into what there is to be learned; we also understand how an infinite aptitude can be encompassed by finite accomplishments. For suppose that a language lacks this feature; then no matter how many sentences a would-be speaker learns to produce and understand, there will remain others whose meanings are not given by the rules already mastered. It is natural to say that such a language is unlearnable. This argument depends, of course, on a number of empirical assumptions: for example, that we do not at some point suddenly acquire an ability to intuit the meanings of sentences on

no rule at all; that each new item of vocabulary, or new grammatical rule, takes some finite time to be learned; that man is mortal. (Davidson 1965, 8–9)

This passage has been interpreted, also by myself, as an argument that natural language semantics must be compositional. Unless it is compositional, since the language does contain infinitely many meaningful sentences, the language would be unlearnable. And since we do learn natural languages, they are learnable.

As Davidson himself says, this argument does build on some assumptions. One is that the language does have an infinite number of sentences, and another (unless this is included in the notion of a sentence) that infinitely many of those sentences do have a definite meaning.⁹ Let’s call such a language *infinitely rich*. If we are allowed to assume that natural languages are infinitely rich, then we do have a good argument for the claim that the meaning of complex expressions must be at least somehow computable. At least no person, *whether or not mortal*, can learn all the meaningful sentences, one at a time, in any finite time (provided that the time needed for learning a new sentence does not rapidly approach zero).

The problem is that we should not just help ourselves to the assumption. The claim that natural languages are infinitely rich is a strong claim about natural languages. Simply assuming it is question-begging. And it is not something we get directly from observations of natural language speakers. It needs a more

⁹Actually, if infinity is required, we would have to add that infinitely many different meanings are expressed by the infinitely many meaningful sentences. It would not be difficult to learn a language with infinitely many meaningful sentences if they all mean the same, or if e.g., sentences with an even number of letters all mean that *p* and those with an odd number all mean that *q*. So we would have to require an infinite number of equivalence classes of synonymous sentences.

But perhaps infinity is not needed. A huge finite number of sentences to be learned might serve as well as a basis for the argument. This has been pointed out e.g., by Richard Grandy (1990).

theoretical justification. Hence, the argument, as it stands, fails because of an unjustified premise.

But worse than that, if we *could* justify the assumption that speakers master an infinitely rich language *without* compositional structure, then we would know in advance that the learnability argument is flawed, because then, if speakers learn such a language from each other we know that it cannot be correctly explained by means of compositionality, simply because we know that the language *isn't* compositional. In fact, learnability requires that the semantics be computable, and computability does not entail compositionality (see [Pagin and Westerståhl 2010b](#), [Pagin 2016](#)).

As an argument for compositionality, the learnability argument therefore fails, both because of relying on the unjustified and even question-begging assumption of infinite richness and because of assuming that learnability entails compositionality.

There is, however, a different interpretation of the paper, according to which Davidson does not make these mistakes. He does not make the second mistake, since he does not make the distinction between compositional and computable semantics. And he does not make the first mistake, since he does *not* argue for compositionality (or that natural language meaning must be computable). Rather, on this interpretation, Davidson *takes for granted* that in natural language, the meaning of a complex expression depends on the meanings of its parts, and that it is therefore compositional in a loose sense. He also takes for granted that natural languages are infinitely rich, but this is a minor addition to the assumption that they have something like a compositional semantics.

On this interpretation, what Davidson argues is that our general *meaning theory* should *bring this out*. It should be a feature of semantic theory that it shows *how* the meaning of a sentence depends on the meanings of the parts, and thus *shows that* the meaning of a sentence is “a function of a finite number of features of the sentence”. Davidson indeed emphasizes that Tarski’s the-

ories of truth satisfy this requirement. He says earlier on the same page:

Though no doubt relativized to times, places, and circumstances, the kind of structure required seems either identical with or closely related to the kind given by a definition of truth along the lines first expounded by Tarski, for such a definition provides an effective method for determining what every sentence means (i.e. gives the conditions under which it is true). ([Davidson 1965](#), 8)

It is doubly charitable to Davidson to interpret him in the proposed alternative way. Firstly, it makes the argument in the paper consistent with his later claim that it did not occur to him that compositionality required justification. And secondly, it avoids attributing to him a question-begging argument.

We can therefore sum up:

- (1965) Davidson
- i) requires that a natural language meaning theory show how the meaning of a sentence depends on the meanings of its parts
 - ii) requires that a natural language meaning theory show how natural languages are learnable
 - iii) holds that the first is a consequence of the second.

This interpretation also agrees well with the opening paragraph of [Davidson \(1967\)](#), to which we now turn.¹⁰

¹⁰An anonymous reviewer urged that I spell out more precisely what Davidson’s argument is in the quoted passage from (1965, 8–9). The reviewer offers several alternatives, and has an objection to each, based on my own objections against the interpretation of the passage as an argument for compositionality. The reviewer’s argument C reads:

- (C)
- i) A semantic theory for L should show how L is learnable.
 - ii) Natural languages are infinitely rich.
 - iii) *Therefore*: A semantic theory for a language L should show how the meanings of sentences of L depend on the meanings of their parts.

4. Truth and Meaning (1967)

It is much clearer in [Davidson \(1967\)](#) that Davidson takes compositionality (in the loose sense) as a point of departure. The paper opens:

It is conceded by most philosophers of language, and recently by some linguists, that a satisfactory theory of meaning must give an account of how the meanings of sentences depend upon the meanings of words. Unless such an account could be supplied for a particular language, it is argued, there would be no explaining the fact that we can learn the language: no explaining the fact that, on mastering a finite vocabulary and a finite set of rules, we are prepared to produce and understand any of a potential infinitude of sentences. I do not dispute these vague claims, in which I sense more than a kernel of truth. Instead I want to ask what it is for a theory to give an account of the kind adumbrated. ([Davidson 1967](#), 304)¹¹

Moreover, a later passage adds that Davidson also assumes that the alleged abilities of natural languages speakers can be explained by compositionality (in the loose sense):

In this paper I have assumed that the speakers of a language can effectively determine the meaning or meanings of an arbitrary ex-

The reviewer's objection is that premise (Cii) is still unargued for, and hence on this interpretation, the argument is flawed by my own standards. This is not correct, however. As is suggested in the main text, we can argue for premise (Cii) from the unstated extra premise that L has a compositional semantics. Of course, this conclusion does not strictly follow, since a compositional semantics need not give infinitely many *different* meanings to sentences, even if it does give meaning to infinitely many sentences. But once we have assumed a compositional semantics for an infinite syntax, the possibility of infinite richness is explained. It is just one option for the semantics.

Except that Davidson did not explicitly make the distinction between an infinite syntax and an infinitely rich language, I think that Argument C corresponds well to how Davidson was thinking. I am not claiming, however, that Davidson actually was prepared to back up the infinite richness claim by appeal to compositionality. Rather, I think both were taken as data. It is just that we *can* justify the first from the second.

¹¹In the passage, there is a footnote reference to [Davidson \(1965\)](#).

pression (if it has a meaning), and that it is the central task of a theory of meaning to show how this is possible. I have argued that a characterization of a truth predicate describes the required kind of structure, and provides a clear and testable criterion of an adequate semantics for a natural language. ([Davidson 1967](#), 320)

The second part of this later passage also rather well characterizes what Davidson is up to in the 1967 paper. Oversimplifying somewhat, we can say that Davidson (i) gives a step-by-step motivation for using Tarskian T-theories as theories of meaning, and (ii) considers a range of problems in so doing. I shall here focus on (i).

As for (i), it is taken as given both that we need a (loosely) compositional semantics and that a Tarskian T-theory meets that condition. Beyond that, the argument for the Tarskian format largely proceeds by discrediting alternatives:

1. Davidson rejects the idea (pretty much the idea presented in [Katz and Fodor 1963](#)) that lexical meanings plus syntax can deliver a compositional semantic theory ([Davidson 1967](#), 307–08). Davidson's point is simply that we need to understand the semantic significance of the syntactic construction in order to get from the meanings of parts to the meanings of the whole. To drive the point home, he refers to belief sentences as an example: in this case we know both the lexical meanings of the simple parts and the syntax, but we are still unclear about how to get the meanings of the entire sentences.

2. Davidson further argues that the condition on a theory of meaning, that it give the meaning of each sentence in the language, does require that the semantic relation, . . . *means* . . . , is characterized by means of recursion over syntax (thereby giving the semantic significance of syntactic constructions). It does *not* require, however, that meanings are treated as *entities* ([Davidson 1967](#), 306–07). Moreover, Davidson claims that no account has been given (in particular not by Frege or Church) where the condition is met and meaning entities play an important role.

In particular, we can state the *substitution* version of compositionality by appeal to synonymy, without appeal to meanings as entities:

Meanings as entities, or the related concept of synonymy, allow us to formulate the following rule relating sentences and their parts: sentences are synonymous whose corresponding parts are synonymous ('corresponding' here needs spelling out of course). (Davidson 1967, 307)

The rule is equivalent to the substitution version, since synonymous substitutions are precisely what provide a new sentence whose parts are synonymous to the corresponding parts of the original sentence. No doubt, Carnap's notion of intensional isomorphism is an historical antecedent to this passage.

3. Giving up on meanings as entities, we can change the format of specifying the meaning of sentences. We move from the format 's means m', with "m" a singular term, to 's means that p', where "p" holds the place of a sentence in the meta-language (Davidson 1967, 309).

4. A remaining problem is that "means that..." is a non-extensional context, which will create difficulties for the derivations in the theory. The way to solve this problem is to replace "means that" by an extensional predicate (Davidson 1967, 309). What matters is that the theory pairs up the right s and p, not which predicate is used for the pairing.

5. Attempting to avoid the problems of non-extensional context suggests the extensional format

(T) s is T iff p.

where "s" is to be replaced by a name of an object language sentence and "p" by a meta-language sentence, appropriately related to what is referred to by what replaces "s". In case context dependence is not involved, the former is just what the latter refers to, or a translation of it. Davidson notes (1967, 309–10) that any two predicates "T" satisfying the schema would have the same extension.

6. Under the stipulated relation between "s" and "p", the truth predicate is one such predicate. Because of that, (T) is in fact equivalent to Tarski's Convention T (Tarski 1983, 187–88). Hence, since Tarski's format for a formal truth definition can be seen as specifying truth-conditions by means of recursion over syntax, this format can be used as a compositional theory of meaning.

Summing up, we can say that

(1967) Davidson

- i) requires that a theory of meaning give the meaning of each sentence in the language in a (loosely) compositional manner;
- ii) argues that this task is best achieved with an extensional meaning predicate;
- iii) argues that such an extensional predicate is a truth predicate, and that therefore Tarski's model of formal truth definitions satisfies the requirements on a theory of meaning.

5. Conclusion

It is fairly clear from these considerations that Davidson took it as a matter of course that meaning in natural language is (loosely) compositional, at least in the sense that the meaning of a sentence depends on the meanings of its parts. He never argued for this claim, only for the further claim that semantic theories must be compositional, in the sense of *showing* how the meanings of sentences depend on the meanings of their parts.

Simplifying somewhat, we can say that the main claims in the three papers considered are the following:

- (Davidson 1963) The choice of analysis of particular natural language constructions must be guided by a general theory of meaning.

- (Davidson 1965) A general theory of meaning must show how languages are learnable, and must therefore show how the meaning of a sentence depends on the meanings of its parts.
- (Davidson 1967) The best way of constructing a general theory of meaning uses an *extensional* meaning predicate, and therefore the format of Tarski's truth definitions offers the best model for a general meaning theory.

Acknowledgements

I am grateful to an anonymous reviewer for bringing up points that have led to an improved paper. I am especially grateful to Kathrin Glüer for helping me to a better understanding of Davidson's philosophy through many discussions over the years.

Peter Pagin
Stockholm University
peter.pagin@philosophy.su.se

References

- Carnap, Rudolf, 1947. *Meaning and Necessity*, 1st ed. Chicago: University of Chicago Press.
- , 1956. *Meaning and Necessity*, 2nd ed. Chicago: University of Chicago Press.
- Church, Alonzo, 1950. "On Carnap's Analysis of Statements of Assertion and Belief." *Analysis* 10: 97–99.
- , 1951. "A Formulation of the Logic of Sense and Denotation." In *Structure, Method, and Meaning: Essays in Honor of Henry M. Sheffer*, edited by P. Henle, H. Kallen and S. Langer, pp. 3–24. New York: Liberal Arts Press.
- Davidson, Donald, 1963. "The Method of Extension and Intension." In *The Philosophy of Rudolf Carnap (Library of Living Philosophers)*, edited by P. Schilpp, pp. 311–49. La Salle, IL: Open Court.
- , 1965. "Theories of Meaning and Learnable Languages." In *Logic, Methodology and Philosophy of Science II*, edited by Y. Bar-Hillel, pp. 3–17. Amsterdam: North-Holland. Reprinted in Davidson (1984), pp. 1–15.
- , 1967. "Truth and Meaning." *Synthese* 17: 304–23. Reprinted in Davidson (1984), pp. 17–36.
- , 1973. "Radical Interpretation." *Dialectica* 27: 313–28. Reprinted in Davidson (1984), pp. 125–39.
- , 1984. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.
- , 1986. "A Nice Derangement of Epitaphs." In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, edited by Ernest Lepore, pp. 433–46. Cambridge: Basil Blackwell.

- , 1999. "Reply to Pagin." In *Żegleń (1999)*, pp. 68–70.
- Frege, Gottlob, 1923. "Gedankengefüge." *Beiträge zur Philosophie des Deutschen Idealismus* 3: 36–51. Translated as "Compound Thoughts," by R. Stoothoff. *Mind* 72 (1963): 1–17.
- Grandy, Richard, 1990. "Understanding and the Principle of Compositionality." *Philosophical Perspectives* 4: 557–72.
- Katz, Jerrold and Jerry Fodor, 1963. "The Structure of a Semantic Theory." *Language* 39: 170–210.
- Pagin, Peter, 1999. "Radical Interpretation and Compositional Structure." In *Żegleń (1999)*, pp. 55–68.
- , 2012a. "Communication and the Complexity of Semantics." In *The Oxford Handbook of Compositionality*, edited by W. Hinzen, E. Machery and M. Werning, pp. 510–29. Oxford: Oxford University Press.
- , 2012b. "Truth-theories, Competence, and Semantic Computation." In *Davidson's Philosophy: A Reappraisal*, edited by G. Preyer, pp. 49–75. Oxford: Oxford University Press.
- , 2016. "Compositionality, Computability, and Complexity." Unpublished draft.
- , 2017. "How [Not] To Justify Compositionality." Unpublished draft.
- Pagin, Peter and Dag Westerståhl, 2010a. "Compositionality I: Definitions and Variants." *Philosophy Compass* 5: 250–64.
- , 2010b. "Compositionality II: Arguments and Problems." *Philosophy Compass* 5: 265–82.
- Partee, Barbara, 1984. "Compositionality." In *Varieties of Formal Semantics*, edited by F. Landman and F. Veltman, pp. 281–311. Dordrecht: Foris.
- , 2004. *Compositionality in Formal Semantics: Selected Papers by Barbara H. Partee*. Oxford: Blackwell.
- Putnam, Hilary, 1954. "Synonymity, and the Analysis of Belief Sentences." *Analysis* 14: 114–22.
- Tarski, Alfred, 1935. *Der Wahrheitsbegriff in den formalisierten Sprachen*. Leopoli: Studia Philosophica.
- , 1983. *Logic, Semantics, Meta-Mathematics*, 2nd ed., edited by J. Corcoran. Indianapolis: Hackett Publishing Company.
- Żegleń, Urszula, ed., 1999. *Donald Davidson. Truth, Meaning and Knowledge*. London: Routledge.