

Is Compositionality a Trivial Principle?

Richard G. Heck, Jr.

Department of Philosophy, Brown University

I

“The meaning of a sentence is determined by the meanings of its parts and how those parts are put together.” Such is the observation that supports the demand that semantic theories, ‘theories of meaning’, should be *compositional*. Very roughly, the demand is that such theories should characterize the meanings of sentences (and other sorts of complex expressions) ‘recursively’, by ascribing meaning to sub-sentential expressions, and to various forms of syntactic combination, and then deriving the meanings of sentences therefrom. Only a theory that met this constraint, it is thought, could properly be described as answering the question how the meaning of a sentence is determined by the meanings of its parts and the way the parts are combined.¹

There are a number of problems that arise in connection with the question how the principle of compositionality should properly be formulated. I shall not be discussing these issues here. What I want to discuss, rather, is the question whether compositionality is a substantial constraint at all. Paul Horwich (1997), among others, has argued that it is not, in part because it is an utterly trivial matter to formulate semantic theories that are compositional.² I shall explain what sorts of ‘semantic theories’ Horwich has in mind shortly.

I shall argue, as against Horwich, that the trivial semantic theories he formulates not only do not explain the phenomenon of compositional-

¹ The *locus classicus* for this conception is of course Davidson’s paper “Truth and Meaning” (Davidson, 1984).

² This paper is essentially reprinted in Horwich’s book *Meaning* (Horwich, 1998), with some minor changes. Page references in the text are given in the form “000/000”, with the first pages number(s) referring to the paper; the second, to the book. Terrence Parsons (1996) and Mark Johnston (1998) have expressed relevantly similar views, and they have since become common among self-described deflationists.

ity but are explanatorily wholly vacuous. I firmly believe that the arguments I shall offer are conclusive against Horwich's particular attempt to evade the constraints that the structured, articulated character of language is usually thought to impose on semantic theory. Since I do not myself see any other way of evading these constraints, I believe that these arguments also establish that compositionality is a substantive constraint, but I do not claim as much for the arguments on this score. Nor, just to be clear, do I claim that they establish any particular version of the thesis of compositionality, whose detailed formulation, as I have already said, is a matter of some controversy.

The view I want to refute is thus that compositionality imposes no substantial constraints upon semantic theories, because the alleged constraint is trivially satisfiable. How so? Very roughly, the idea is this. Consider a simple sentence, such as "dogs bark". This sentence means that dogs bark.³ Introducing some notation, it means: dogs bark.⁴ The problem is to see how this fact, that "dogs bark" means dogs bark, might be seen as determined by facts about the meanings of the words "dogs" and "bark" and how they are combined in the sentence "dogs bark". Formally, the problem is to formulate a theory in which the fact that "dogs bark" means dogs bark is derivable from assumed facts about the meanings of these words and the semantic significance of the relevant mode of combination.

Horwich claims that it is easy to formulate a theory meeting the mentioned requirements, roughly as follows (506/158):

- "dogs" means dogs;
- "bark" means bark;
- a sentence of the form N-V means:
 $\text{Comp}_{N-V}(\text{the meaning of N, the meaning of V})$;
- "dogs bark" is a sentence of the form N-V.

Here, " Comp_{N-V} " denotes a function from meanings to meanings satisfying the following condition, which I am quoting from Horwich and which I shall call COMP (529/180):

³ I shall here ignore the complexities introduced by context-dependence. Accounting for them can only make matters worse.

⁴ Horwich uses capital letters, thus: DOGS BARK. Parsons uses the more traditional circumflex, thus: $\hat{(\text{dogs bark})}$. I use underlining for neutrality, and because it is easier to read.

The meaning of the result of applying combinatorial procedure P to a sequence of primitives = The result of applying P to the sequence of the meanings of those primitives.

The first four premises imply that “dogs bark” means $\text{Comp}_{N-V}(\text{dogs}, \text{bark})$, and it is then supposed to follow from COMP that dogs bark = $\text{Comp}_{N-V}(\text{dogs}, \text{bark})$. (Thus, N-V is a “combinatorial procedure”.) So, says Horwich, from the stated assumptions, we can derive that “dogs bark” means dogs bark, which is “just what we wanted to explain” (506/157). Thus, the theory shows us how the meaning of this sentence is determined by the meanings of its parts and how they are combined. It seems plausible that the theory can easily be extended to other sorts of sentences, so we seem to have been given a recipe for a compositional theory of meaning for the whole of English.

Horwich remarks that his main claim is “that compositionality imposes no constraint on how the meaning properties of the *primitives* are constituted” (509/159-60). Theories of meaning of this kind, which we may call *direct meaning-theories*, are neutral regarding such matters as whether the meaning of a (primitive) predicate is a set, or a property, or whatever, or whether the meaning of a proper name is its referent: More generally, direct meaning-theories are supposed to make no non-trivial assumptions about what sorts of things the meanings of primitive expressions are. On the other hand, Horwich does commit himself to a specific view about the meanings of sentences.⁵ On this “deflationary proposal”, a sentence’s meaning what it does just consists in its being constructed in the way it is from parts that mean whatever they do: For “dogs bark” to mean dogs bark just is for it to be constructed, *via* the combinatorial procedure N-V, from words meaning what “dogs” and “bark” do, viz., dogs and bark (504-5/155-6). Horwich’s thesis is that, if we accept his deflationary conception of what constitutes a sentence’s meaning what it does—his “constitutive thesis”, as I shall call it—it is a trivial matter to explain how the meaning of a complex expression is determined by the meanings of its parts and how they are combined.

Not everyone who has been attracted by the simplicity of meaning-theories of this kind has endorsed Horwich’s constitutive thesis. I shall consider its significance later. I will first argue that direct meaning-theories are, by themselves, explanatorily vacuous: Despite appear-

⁵ And other complex expressions. Henceforth, I shall speak just of sentences, leaving it understood that remarks applying to them are also supposed to apply to other sorts of complex expressions.

ances, a direct meaning-theory does *not* allow us to explain how the meaning of a sentence is determined by the meanings of its parts.

II

In the context of a direct meaning-theory, claims about how a sentence is constructed from its parts (and, indeed, what its parts are) play no essential role: Such theories are, in a sense I shall explain, insensitive to facts about the structures of sentences. Let me illustrate this phenomenon. The story I told above about how the meaning of the sentence “dogs bark” is determined by the meanings of its parts differs from the story Horwich tells. His story appeals to the following principles:

- “dog” means dog;
- “bark” means bark;
- a sentence of the form Ns-V means:
 $\text{Comp}_{Ns-V}(\text{the meaning of N, the meaning of V})$;
- “dogs bark” is a sentence of the form Ns-V.

The difference here is that the two theories ascribe different structures to the sentence “dogs bark” and so explain its meaning dogs bark in different ways. My story speaks of the meaning of “dogs” and treats this sentence as being of the form N-V. Horwich’s story, on the other hand, speaks of the meaning of “dog” and treats the sentence as being of the form Ns-V. One might prefer Horwich’s treatment on a number of grounds. One might, for example, think it a mistake to suppose that “dogs” is a primitive whose meaning has nothing to do with that of “dog”: On this view, the word “dogs” is, despite superficial appearances, semantically complex, consisting of the common noun “dog” and the plural affix “-s”.⁶ What is puzzling, however, is that, so far as these theories’

⁶ Similarly, it is widely thought that the affixes by means of which tense is expressed in English are the morphological realization of underlying syntactic features of (tensed) sentences. In that sense, we might say that “-ed” is a word—an expression with independent semantic significance—even though it is not realized as what we call a “word” in ordinary English. And if one adopts such a view, then one should also say that “ran” is the morphological realization of what is, at a deeper level, just what one would expect it to be, namely, “runned” (or something like that).

Anyone familiar with actual work in semantics will recognize that there are deeper issues here, since the most usual reading of “dogs bark” treats it as a generic. (It has other readings as well: “What happens at night in the country?” “Dogs bark.”)

ability to deliver the wanted theorems is concerned, it just doesn't seem to matter whether my theory is correct, or Horwich's theory is correct, or, indeed, whether neither of them is.

Here's another example. Consider the sentence "John believes that pigs fly". In older work, it is often treated as consisting of a noun-phrase, "John", an operator, "believes that", and the sentence on which it operates, "pigs fly". Contemporary syntacticians, on the other hand, regard the sentence as consisting (very roughly) of the noun-phrase, "John", a transitive verb, "believes", taking a clausal complement, and the complement itself, "that pigs fly".⁷ The issue here is *not* which of these views is correct.

On the contrary, so far as a direct meaning-theory's ability to deliver the wanted theorems is concerned, it doesn't matter which of them is correct. That's what I mean when I say that direct meaning-theories are insensitive to claims about the structures of sentences. There is thus a gap between a direct meaning-theory's *issuing in derivations* of such claims as that "dogs bark" means dogs bark and its *explaining* these facts. Consider again the two theories given above concerning the sentence "dogs bark". Although these theories attribute different structures to the sentence, this difference does not affect the theories' ability to deliver the required theorems: The derivation of the claim that "dogs bark" means dogs bark proceeds unhindered, whatever we take its structure to be. Surely, however, not both theories can count as *explaining* why "dogs bark" means dogs bark. But on what ground can we ascribe one of the two theories greater explanatory force than the other? It seems to me that if either counts as explaining why "dogs bark" means dogs bark, then both do. So neither does.

The reply, however, will surely be that we cannot regard a direct meaning-theory as explaining why a sentence means what it does unless it ascribes the right structure to the sentence in question. Note first, however, that, if it is a substantial question what the structure of a certain sentence actually is, then it is a good deal less easy to extend Horwich's theory to one for the whole of English than he seems to suppose: To explain how the meaning of a given sentence is determined

⁷ Of course, this is massively over-simplified, and there are further questions about the structure of the complement, in any event. But we need not pause over such matters here.

The contemporary view is well-supported. To get a sense for the reasons in favor of it, note that the complementizer "that" moves with the clause: We say, "That pigs fly is something John believes", not "Pigs fly is something John believes that".

by the meanings of its parts one must, presumably, figure out what the structure of that sentence actually is. Horwich's explanation of how the meaning of "dogs bark" depends upon the meanings of its parts, and how they are put together, therefore cannot be regarded as obviously correct: It will be correct only if "dogs bark" is of the form "Ns-V", and that claim can be questioned. Indeed, it is obviously false.

But, the reply will continue, though it is indeed hard to say what the structures of certain sorts of sentences are, whatever their structures may be, there *has* to be a direct meaning-theory that attributes those structures—which are, by hypothesis, the right structures—to those sentences. So it is obvious that there is a direct meaning-theory that assigns the right structures to the sentences and allows us to derive facts about their meanings from facts about the meanings of their parts and how they are combined, even if we don't yet know which direct meaning-theory that is. If so, isn't it also obvious that some direct meaning-theory will explain how the meanings of English sentences are determined by the meanings of their parts?

The answer, or so I shall argue, is "No": Even a direct meaning-theory that ascribes the right structures to sentences fails to explain, e.g., how the meaning of "dogs bark" is determined by the meanings of its parts and how they are combined. The problem is that, since any structure will do, direct meaning-theories have no grip on the notion of "right structure".

There are many different notions of a sentence's structure: Surface structure, deep structure, phonological structure, and so forth, not to mention orthographic structure and grammatical structure in the ordinary sense. Horwich cannot simply stipulate the notion of structure he wishes to employ. Suppose he did so—call that notion "H-form"—and rewrite his direct meaning-theory thus:

- "dog" means dog;
- "bark" means bark;
- A sentence that has the H-form Ns-V means:
 $\text{Comp}_{Ns-V}(\text{the meaning of N, the meaning of V})$;
- "dogs bark" has the H-form Ns-V.

Of course, I can just re-write my theory thus:

- "dogs" means dogs;

- “bark” means bark;
- A sentence that has the grammatical form N-V means:
 $\text{Comp}_{N-V}(\text{the meaning of N, the meaning of V})$;
- “dogs bark” has the grammatical form N-V.

Both theories assign a correct structure to “dogs bark”, by hypothesis, and there will be many more besides. What reason can there be, then, to regard one rather than another of these theories as explaining why “dogs bark” means dogs bark? The answer, presumably, is that, while sentences do have many different sorts of structure, there is no reason to suppose that all of these structures are ones it would be appropriate to use in stating a meaning-theory. The problem with the ‘grammatical’ theory, then, is that it uses an inappropriate notion of structure.

Before we can even begin to formulate a meaning-theory, then—before we can even begin to answer the question what ‘the’ structure of a given sentence might be—we need to know what notion of structure is supposed to be relevant. How, then, would Horwich have us identify the notion of structure relevant to questions about compositionality? I have heard it suggested that Horwich should defer to syntax, so that the relevant notion of structure is just whatever the syntacticians say it is. But syntacticians themselves standardly distinguish various levels of representation, and compositionality plays a crucial role in identifying the level at which semantic interpretation is supposed to occur. The emergence of LF (for “logical form”) in the 1970s was driven in large part by the need to identify a level of representation at which compositional accounts of the meanings of sentences—in particular, of quantifiers—could be given. Since the scopes of quantifiers are not represented at surface structure, no compositional explanation can be given at that level of why, say, “everyone read most of the books” means what it does, rather than what “most of the books were read by everyone” means. So another level, LF, is needed.⁸

It has thus long been the standard view, not just in philosophy, but also in linguistics, that assignments of logical form to sentences must

⁸ There are alternatives, to be sure, and, even within the Chomskyan tradition, syntactic theory has been in flux in recent years, but the points made here still stand. In so-called Direct Compositionality accounts, for example, the structure is relatively simple, but the rules for semantic interpretation become much more complicated than Horwich’s picture would allow.

support a compositional semantics for them.⁹ But Horwich cannot accept such an account of LF's importance. It is far too easy to extend his theory to encompass any new construction one would like, for example, adverbs:

- “loudly” means loudly;
- A sentence of the form S-Adv means:
 $\text{Comp}_{S-\text{Adv}}(\text{the meaning of S, the meaning of Adv})$;
- “dogs bark loudly” is a sentence of the form S-Adv.

Now we can derive that “dogs bark loudly” means dogs bark loudly. Or again: One might have thought that it was only with Frege's introduction of the notion of quantification—and more importantly, of the notion of scope—that we were in any position at all to understand why “Everyone loves someone” means what it does. But there is an obvious direct meaning-theory that makes no use of the notion of scope, and that treats “Everyone loves someone” as being of the same form as “John loves Sue”, and yet allows us to derive that “Everyone loves someone” means everyone loves someone. It should, furthermore, be obvious that structural ambiguities will pose no difficulty. If we want to explain the ambiguity of “everyone loves someone”, we need only identify two possible structures for it, say, “ $Q_1 - V - Q_2$ ” and “ $Q_1 - V^* - Q_2$ ”, and give appropriate compositional principles for such structures. As long as there are enough structures to go around, it won't matter what they are.

So, as said, Horwich cannot just stipulate which notion of structure direct meaning-theories are supposed to employ. He cannot, e.g., simply insist that grade school grammar books will tell us all we need to know. We need to be told *why* one notion of structure, rather than some other, is the one appropriate for use in explanations of the semantic properties of expressions: why, to put the point in Horwich's own terms, a sentence's meaning what it does consists in its being composed in *that* way from parts that mean what they do, rather than in some other way. And what I have just been arguing is that Horwich simply co-opt the standard answer to this sort of question, since that answer depends upon a conception of the explanatory burden imposed by compositionality that

⁹ And it is then a substantial question how, in this theoretical context, compositionality should best be understood. This is the sort of problem I mentioned towards the beginning of the paper.

he rejects. And I know of no other answer. But without one, there is no principled way to choose among the far too many direct meaning-theories we can construct: to distinguish the one that is supposed to be explanatorily adequate from the great majority whose explanatory vacuity is obvious.

III

It's easy to think there has to be something wrong with the foregoing. Let us, therefore, consider directly the question why direct meaning-theories are so plentiful. As we shall see, it is because they are almost devoid of content.

It will simplify the exposition if we resolve, at this point, to treat terms of the form “dogs”—which are supposed to mean something like “the meaning of the word ‘dogs’”—as semantically complex. Horwich says that such terms are “primitive” (524/175), and this seems to be his official view. But he ought not to have said this: To insist that “dogs” is primitive is to undermine our understanding of the principle COMP, that is, of how meanings are supposed to compose, on Horwich’s view. Recall that COMP, as Horwich states it, is:

The meaning of the result of applying combinatorial procedure P to a sequence of primitives = The result of applying P to the sequence of the meanings of those primitives.

Horwich never states this principle using the quasi-formal machinery he employs elsewhere, and for good reason. If we attempt to formalize the principle, the best we can do would seem to be something like this:

$$\text{Comp}_{N-V}(\underline{x}, \underline{y}) = \underline{N-V(x,y)}.$$

Here, ‘ x ’ and ‘ y ’ are variables ranging over “primitives”, that is, atomic expressions; ‘ $N-V(x,y)$ ’ denotes a sentence having x as its subject and y as its predicate. So the principle *seems* to say that the result of applying Comp_{N-V} to the meanings of the primitives x and y is the meaning of the result of applying the corresponding mode of combination to the expressions x and y . But, so formulated, the principle makes no sense. If terms of the form “dogs” are primitive, then quantification into underlining is unintelligible. To put the point differently, when Horwich speaks, in his formulation of COMP, both of primitives and of their

meanings, he speaks in a way that *requires* us to construe terms of the form “dogs” as semantically complex.¹⁰

So how do matters stand if we do treat terms like “dogs” as semantically complex? Well, first of all, such terms would better be written “‘dogs’”, for it is the meaning of “dogs” that this term is supposed to denote, not the meaning of dogs (whatever that may be). Just so we can be clear about this, then, let us write such terms as: $\mu(\text{‘dogs’})$. My version of Horwich’s theory then takes the form:

- “dogs” means $\mu(\text{‘dogs’})$;
- “bark” means $\mu(\text{‘bark’})$;
- a sentence of the form N-V means: $\text{Comp}_{N-V}(\mu(N), \mu(V))$;
- “dogs bark” is a sentence of the form N-V;
- $\mu(N - V) = \text{Comp}_{N-V}(\mu(N), \mu(V))$.

Note that the last principle—which we may call a *composition principle*—takes the place of the relevant version of COMP.¹¹

Let us look closely at the content of the composition principle. What does it actually assert? The principle says that the meaning of a sentence of a certain form is a certain function of the meanings of its parts, and it gives us a name for that function, namely, “ Comp_{N-V} ”. But neither the composition principle nor the theory as a whole tells us

¹⁰ Horwich himself suggests a way of construing his theory on which terms like “dogs” are treated as semantically complex: Formulating his theory in this way allows him to finitely axiomatize it, and he knows some will think that desirable (524, fn.15/176-7). My point is not that Horwich has overlooked the possibility of such a construal. It is, rather, that this sort of construal is not just an option available to fetishists about finite axiomatizability. On the contrary, it is required if the compositional principle at the core of his theory is to be so much as intelligible.

It may be that Horwich intends COMP to be understood as a schema. I think it cannot be so understood, but we need not pursue this issue, since Horwich appears willing to accept the emendation. Moreover, although it is worth being clear about the points just made, nothing in the arguments I am about to give depends upon them: It will be clear that the arguments apply both to theories that treat “dogs” as primitive and to theories that do not.

¹¹ If one wanted to say that this theory is too trivial to be of any interest at all—because all that is now being said is that “dog” means whatever “dog” means, and so forth—I would be sympathetic. But one needs to be careful with the details of this criticism. Horwich considers what looks to be a similar objection and seems unmoved by it (506-7/157, fn. 5).

anything about this function, other than that the meanings of expressions and the function Comp_{N-V} should together satisfy the composition principle. That is, if we strip away the unexplained quasi-formal machinery and focus simply on what direct meaning-theories actually *say*, we see that their whole content amounts to this:

1. Sentences are composed of parts (in a sense the theory specifies).
2. The parts have meanings.
3. For each “mode of composition”, there is a function that yields the meaning of a sentence so composed when it is given the meanings of its parts as arguments.

No more content can be extracted from such a theory.

That is the essential point: Beyond its claim that sentences are ‘composed’ of ‘parts’ and that these ‘parts’ have ‘meaning’, the only thing a direct meaning-theory tells us—even one built upon a *correct* account of sentential structure—is that, for each ‘mode of composition’, *there is* a function that yields the meaning of a sentence of the appropriate form when given the meanings of its parts as arguments, that is, that *there is* a function satisfying the appropriate composition principle. Insofar as such a theory purports to explain how sentence-meaning is determined by word-meaning, it can therefore do so only in terms of the *existence* of such a function, for that is all the theory asserts. But simply to assert the existence of such a function is not to *explain how* the meaning of a sentence is determined by the meanings of its parts: It is just to say, in fancy language, that the meaning of a sentence *is* determined by the meanings of its parts.¹²

That there is a function that yields the meaning of a sentence when given the meanings of its parts as arguments implies that the meaning of a sentence is determined by the meanings of its parts only in a very weak sense: a purely mathematical sense. And functions are cheap. *Very* cheap. In fact, the bare existence of a function satisfying a direct meaning-theory’s composition principle follows from minimal assumptions that have almost nothing to do with meaning and sentential structure. The following three conditions, in particular, will suffice:

1. Sentences *in some sense* consist of ‘parts’.

¹² Horwich considers such an objection and summarily dismisses it (516-7/167-8). But there is more to the objection than he allows, as I shall now argue.

2. No two sentences consist of the same ‘parts’.
3. Distinct expressions have distinct ‘meanings’.

No more is required: In particular, there is no reason that the ‘parts’ we identify should be semantically relevant parts of the sentences of which they are parts; nor need the ‘meanings’ associated with expressions in any plausible sense be *meanings*.¹³ The sense in which direct meaning-theories are insensitive to claims about sentential structure is thus not just a proof-theoretic one: It is not just that one can formulate many different such theories all of which deliver the wanted theorems. On the contrary, so long as we are concerned only with the purely mathematical sense of “determination”, it will be *true* to say that the meaning of a sentence is determined by the meanings of its parts *however* we identify its parts—that is, whatever structure we take the sentence to have—so long as we require of the ‘meanings’ and ‘parts’ only that they satisfy the three conditions just mentioned.

The point can be made more precise.

Consider the following view, sometimes ascribed to Frege. There is a substantial, independent conception of the meanings of sentences (a truth-conditional one, or whatever), but the only legitimate notion of the meaning of a word (or other sub-sentential constituent) is insubstantial: The meaning of a word simply consists in whatever ‘contribution’ it makes to the meanings of sentences in which it is contained. On this sort of view, no matter what structures we might assign to sentences, there will be a way of assigning meanings to their primitive parts so that there is a function from the meanings of the parts to the meanings of wholes: Simply assign different meanings to all the primitives (unless, if you like, they are intersubstitutable *salva significatione*); the existence of the required function is then a mathematical consequence of the fact that no two different sentences consist of the same parts put together in the same way. If we lack any independent conception of the

¹³ If this does not seem obvious, consider the following (which constitutes a proof). Suppose someone said that the meaning of an expression is just that very expression and that the structure of a sentence is just its obvious, orthographic structure (whence the ‘parts’ of a sentence are just the words of which it is composed, in the usual sense). Such a person could accept the sort of theory we’ve been discussing by interpreting μ as the identity function and Comp_{N-V} as concatenation, and taking “means” to mean “is identical with”. That is: Every direct meaning-theory can be truly interpreted as a theory concerning not how the meanings of sentences are determined by the meanings of their parts but how the identity of a sentence is determined by the identities of its parts.

meanings of the parts—if our only conception of what a part means is that its meaning is its ‘contribution’ to the meanings of sentences—we can have no independent “check” on whether the assignments of meanings to parts is correct, and we cannot object to any such assignment nor to any conception of structure.¹⁴

Similarly, suppose that we take ourselves to have a substantial, independent conception of the meanings of words (or of what constitutes their meaning what they do), but regard as unnecessary any independent conception of the meanings of sentences: A sentence’s meaning what it does simply consists in its being composed in the way that it is of words that mean what they do. Then again, no matter what structures we might think sentences have, there is going to be a way of assigning meanings to them so that there is a function from the meanings of the parts to the meanings of wholes: Simply assign different meanings to all of the sentences; the existence of the required function will then be a mathematical consequence, once again, of the fact that no two sentences consist of the same parts put together in the same way.

To see how ridiculous the situation can become, consider the following direct-meaning theory:

1. ‘a’ means a
2. ‘b’ means b
- ...
26. ‘z’ means z
27. ‘ ’ means (there is a space there)
28. An expression of the form $\xi\eta$ means: $\text{Comp}_c(\text{the meaning of } \xi, \text{ the meaning of } \eta)$.

(The operation here is, of course, concatenation.)

29. The meaning of the result of applying combinatorial procedure C to a sequence of primitives = The result of applying C to the sequence of the meanings of those primitives.

It is easy to derive from these axioms the theorem: “dogs bark” means dogs bark. Thus we show how the meaning of a sentence is determined

¹⁴ This point is due in essentials to Michael Dummett (1991, ch. 6).

by the meanings of the letters (and spaces) contained in it and how they are put together.

Of course, the theory shows no such thing. But why not? It is *not* because this theory takes sentences to be constructed from ‘parts’ that don’t mean anything. For one thing, every dictionary I’ve ever seen lists the letters, often as names of themselves, but sometimes with other meanings, too (e.g., “a” can mean: shaped like an “a”). Moreover, even if every letter isn’t, as things are, an independently meaningful expression, it is obvious to me that the above theory would still be unacceptable even if every string of English letters *were* a meaningful expression. Nor can one content oneself with the thought that the meaning of a sentence just *isn’t* determined by the meanings of the letters of which it is composed and how they are put together: In the purely mathematical sense, it is (modulo ambiguity and the like, which we, together with Horwich, are ignoring).

What the foregoing shows is thus this. Suppose we have a direct meaning-theory that operates with an insubstantial, or “deflationary”, conception *either* of the meanings of words *or* of the meanings of sentences. Then, given any conception of the structures of sentences, the existence of functions satisfying the theory’s composition principles is mathematically guaranteed. In particular, these functions will exist whether or not sentences have the structures the theory assigns to them in *any* plausible sense. The theory might, for example, regard the word “catatonic” as composed of “cat”, “a”, and “tonic”. It makes no difference.

IV

One might worry that I must have made too little of Horwich’s constitutive thesis. Perhaps the full force of this thesis is not captured by direct meaning-theories, at least, not as I have formulated them. Let me begin to address this worry, then, by discussing a quite different sort of example, one Horwich himself uses to explain the force of the constitutive thesis.

The way a complex expression *sounds*, Horwich remarks, is determined by the way its component parts sound and how they are combined: The way “birds sing” sounds is determined by the way “birds” sounds, the way “sing” sounds, and how they are combined in this sentence. If we want to explain this fact, we can do so in a way reminiscent of direct meaning-theories: We can say that the fact that a sentence sounds the way it does simply consists in its being composed in a certain

way from words that themselves sound certain ways; for example, we can say that the fact that “birds sing” sounds the way it does simply consists in its being composed as it is from words that sound the way that “birds” and “sing” do. Given this assumption, we can explain why “birds sing” sounds the way it does by deriving the fact that it sounds σ (“birds sing”) from the facts that “birds” sounds σ (“birds”), that “sing” sounds σ (“sing”), and that “birds sing” is composed of those words in a certain way (525-6/176). Doesn’t that seem like a good explanation? If so, the obvious analogy between ‘direct sounding-theories’ and direct meaning-theories should lead us to suspect that there is something wrong with the argument that direct meaning-theories are explanatorily vacuous.

My response here is as above: If these ‘direct sounding-theories’ are supposed to operate with an insubstantial, “deflationary” conception of how sentences sound, they too explain nothing at all. The only thing such a theory will assert is that there is a function from the way the parts sound to the way that the whole sounds. The existence of such a function is not only independent of any facts about how sentences are actually structured—the argument given above applies here, too—it is independent of how sentences in fact *sound*. That a sentence’s sounding the way it does consists in its being composed in a certain way from words that sound the way they do implies nothing at all about how the sentence actually sounds: As far as the constitutive thesis is concerned, “birds sing” could sound like “egg leg”. To think otherwise is, most likely, to confuse the constitutive claim with a different one, namely, that “birds sing” sounds how “birds” sounds *followed by* how “sing” sounds. That claim has empirical content—it’s false—but it is no longer just a constitutive claim, and the conception of how a complex expression sounds is no longer insubstantial. So long as direct sounding-theories *do* operate with an insubstantial conception of how complex expressions sound, however, they will fail to engage the empirical facts. They therefore cannot possibly provide any explanation of those facts.

Similarly, Horwich’s constitutive claim regarding meaning—that for a sentence to mean what it does is just for it to be composed in a certain way from words with certain meanings—implies nothing at all about what any sentence actually means. The constitutive thesis is, for example, consistent with the claim that “dogs bark” means not that dogs bark but that Oregon is in Ohio. To think otherwise is simply a mistake. So direct meaning-theories, understood as operating with an insubstantial conception of the meanings of sentences, fail to engage the empirical facts. They therefore cannot provide any explanation of those facts.

Let me emphasize that the foregoing does not show that Horwich's constitutive claim is false. What it shows is that the constitutive claim, on its own, does not yield an explanation of compositionality. Consider a weaker thesis: that the meaning of a sentence supervenes on how it is composed of parts and what those parts mean. It is hard to imagine that anyone would think that we can explain why "dogs bark" means what it does simply by noting that its doing so supervenes on its structure and the meanings of its parts. But if someone did want to take this line, the arguments above would suffice to show that supervenience is too weak a relation to support such an explanation (that being a familiar sort of point, in any event). That does not mean that the supervenience claim is false. Surely it is true.¹⁵ But the supervenience thesis only tells us *that* the meaning of a sentence is determined by the meanings of its parts, whereas the question semantic theories purport to answer is *how* the meaning of a sentence is determined by the meanings of its parts.

One might wonder whether the fact that Horwich speaks of constitution, rather than supervenience, ought to make a difference. But I just do not see how it can do so: The whole point of the constitutive thesis is supposed to be that it commits us, essentially, to nothing. The foregoing amounts to little more than verification of its vacuity.

Horwich thus writes:

[T]here is nothing more to meaning what we do by 'Socrates is wise' than using the words 'Socrates' and 'wise', *and the operation of predication*, in the ways that are constitutive of their meanings, and in appreciating how the sentence is constructed from those three elements. (Horwich, 1998, p. 37, my emphasis)

But what is it to use the operation of predication in the way constitutive of its meaning? What, indeed, is the meaning of predication? As Higginbotham (1999, pp. 677–8) has emphasized, we are told nothing about such matters, and that is no accident. Standardly, predication is taken to be an operation that yields a sentence that is true if, and only if, the denotation of the subject-term (which may of course be a variable) is in the extension of the predicate (or has the property the predicate expresses, or whatever). That is precisely not Horwich's view.

¹⁵ It is another question, obviously, whether compositionality is itself to be *characterized* in terms of supervenience (Szabó, 2000). My own view, for what it's worth, is that compositionality is better characterized in epistemic terms (Heck, 2004).

Lacking any substantial conception of what sentences mean—that is, of the *result* of predication—Horwich *cannot* have a substantial conception of the semantic significance of predication. All he can say about predication is that, operating on words with certain meanings, it yields a sentence whose meaning what it does consists in its being constructed, by predication, from words with those meanings: That is, the sentence will mean whatever sentences mean that are constructed by predication from words that mean whatever those words mean. That cannot fail to be correct, but it is the most boring of tautologies. Horwich’s constitutive thesis thus adds no explanatory force to direct meaning-theories.

V

I conclude that direct meaning-theories have no explanatory force. If not, then no “trivial” explanation of how a sentence’s meaning is determined by the meanings of its parts appears to be available. Indeed, we may conclude more: In order for a non-trivial explanation to be possible, we must operate with non-deflationary conceptions of the meanings *both* of sentences *and* of their parts. Then things are quite different. Consider, for example, a broadly truth-theoretic semantic theory according to which the semantic value of a predicate is its extension; that of a sentence, its truth-value. In the context of such a theory, the claim that adverbs are predicate modifiers implies that there is a function taking us from the extension of “ate broccoli” to the extension of “ate broccoli passionately”. It is *not* mathematically provable that there is any such function, which is a good thing, since there is no such function. If there were such a function, then, if “ate broccoli” and “ate cauliflower” were co-extensional, then “Jane ate broccoli passionately” and “Jane ate cauliflower passionately” would have to be materially equivalent. But they need not be, even if “ate broccoli” and “ate cauliflower” happen to be co-extensional.¹⁶

¹⁶ Thanks to Michael Glanzberg, Bob Hale, Jim Higginbotham, Paul Pietroski, Michael Rescorla, and Jason Stanley, for helpful discussion and reaction to earlier drafts of this material. Talks based upon it were given at the University of Glasgow and Vassar College. Thanks to the audiences at both places for their comments.

This paper was originally written in 1999, but I never published it. Material from it found its way into another paper on compositionality (Heck, 2004), but that one too has remained unpublished. So I am pleased finally to see it published here.

References

- Davidson, D. (1984). 'Truth and meaning', in *Inquiries Into Truth and Interpretation*. Oxford, Clarendon Press, 17–36.
- Dummett, M. (1991). *The Logical Basis of Metaphysics*. Cambridge MA, Harvard University Press.
- Heck, R. G. (2004). 'What is compositionality?', *Manuscript*.
- Higginbotham, J. (1999). 'A perspective on truth and meaning', in L. Hahn (ed.), *The Philosophy of Donald Davidson*. Chicago, Open Court, 671–686.
- Horwich, P. (1997). 'The composition of meanings', *Philosophical Review* 106: 503–532.
- (1998). *Meaning*. Oxford, Clarendon Press.
- Johnston, M. (1998). 'The end of the theory of meaning', *Mind and Language* 3: 28–42.
- Parsons, T. (1996). 'Fregean theories of truth and meaning', in M. Schirn (ed.), *Frege: Importance and Legacy*. New York, de Gruyter, 371–409.
- Szabó, Z. G. (2000). 'Compositionality as supervenience', *Linguistics and Philosophy* 23: 475–505.