Wilfrid Hodges

COMPOSITIONALITY IS NOT THE PROBLEM

Abstract. The paper analyses what is said and what is presupposed by the Principle of Compositionality for semantics, as it is commonly stated. The Principle of Compositionality is an axiom which some semantics satisfy and some don't. It says essentially that if two expressions have the same meaning then they make the same contribution to the meanings of expressions containing them. This is a sensible axiom only if one combines it with (a) a converse, that if two expressions make the same contribution to the meanings of (say) sentences containing them, then they have the same meaning; and (b) some assumption that two expressions which can't meaningfully be substituted for each other have different meanings. (The paper formalises (a) as a full abstraction principle, and (b) as 'Husserl's principle'.) Moreover the Principle of Compositionality speaks only about when two expressions have the same meaning; it adds nothing whatever about what that meaning might be (the 'representation problem'). Some recent discussions by writers in linguistics and logic are assessed. The paper finishes by reviewing the history of the notion of compositionality.

CONTENTS

1. A principle and a property

Being only a nuisance logician and not a linguist, I write this paper with some misgivings. Several of the writings that I comment on contain deep and searching discussions of issues on which I have nothing at all to offer. The paper flows from my own problems in understanding the concept of compositionality. Perhaps they are not such problems for people working closer to the coal face. But if you don't ask questions you don't get answers.

We can conveniently start with the definition of the 'principle of compositionality' given by Partee, ter Meulen and Wall ([29], p. 318). It runs as follows:

> The meaning of a complex expression is a function of the meanings of its parts and of the syntactic rules by which they are combined.

Call this sentence PoC. It has a reasonably clear formal content, which I extract as the paper proceeds. (See [20] for a more mathematical discussion.)

The sentence PoC uses three undefined primitive terms:

MEANING; EXPRESSION; SYNTACTIC RULE.

Any full-blown semantic theory is likely to contain or assume interpretations for these terms. But let us go a little slower. Suppose we have interpretations for MEANING, EXPRESSION and SYNTACTIC RULE. Then we can ask two questions:

1. With these interpretations of MEANING, EXPRESSION and SYNTACTIC RULE, does PoC form a meaningful statement?

2. If Yes, is this statement true?

If the answer to Question 1 is Yes, then I shall say that these interpretations form a *tidy semantics*, and that they meet the *presuppositions* of PoC. If the answer to Question 2 is Yes, then I shall say that the tidy semantics *satisfies* PoC, and that it is *compositional*. Compositionality is a property of semantics. (In practice writers usually regard the interpretations of EXPRESSION

and SYNTACTIC RULE as given beforehand, so that compositionality is a property of the interpretation of MEANING. This is harmless.)

In a moment we shall sketch the main presuppositions of PoC, i.e. the main conditions that a semantics has to meet in order to be tidy. For future reference, note that the presuppositions of PoC are also presuppositions of its negation:

> The meaning of a complex expression is *not* a function of the meanings of its parts and of the syntactic rules by which they are combined.

So an untidy semantics is as much a challenge for the negation of PoC as it is for PoC.

The Principle of Compositionality states that some (undetermined) semantics is compositional. Partee comments ([28], p. 281f):

> I have several goals in this paper. One is to emphasize how many versions of the principle there can be, since I think some arguments about it are clouded by assumptions that it is more clearcut than it is .... Given the extreme theory-dependence of the compositionality principle and the diversity of existing (pieces of) theories, it would be hopeless to try to enumerate all its possible versions.

Yes and no. We can choose any semantics to put into the Principle and get a different statement each time. Even restricting to tidy ones, the variety is endless, though most of these semantics will be linguistically uninteresting. But any ambiguity in the *property* of compositionality is much more limited.

The presuppositions are not too hard to spell out.

> The interpretation of EXPRESSION is a set, whose elements we call *expressions*.

> The interpretation of SYNTACTIC RULE is a set, whose elements we call *syntactic rules*. Each syntactic rule α is a function whose domain is a set of n-tuples of expressions, for some fixed n depending on α; if the n-tuple $(e_1, \ldots, e_n)$ is in the domain of α then $\alpha(e_1, \ldots, e_n)$ is an expression.

> (Single-valuedness of meaning) The interpretation of MEANING is a function whose domain is a set of expressions.

We also need some conditions on the expressions and syntactic rules in order to give sense to the words 'part' and 'complex'. For example an expression e is complex if and only if it has the form $\alpha(e_1, \ldots, e_n)$ for some syntactic rule α; if it does have this form, then α, e_1, ..., e_n are uniquely determined by e, and we call these occurrences of e_1, ..., e_n the *immediate constituents* of e. And so on; nothing in this paper rests on any subtlety in these requirements. As it is stated, PoC seems to assume that every expression has a unique structural analysis. One easily rephrases it to avoid this, but for present concerns there is no need.

PoC also presupposes something about the existence of meanings. I think it doesn't presuppose that every expression is meaningful (i.e. has a meaning), but it certainly presupposes something of the form

> If an expression e is meaningful, then so are all the relevant constituents of e.

This will have to do for the moment, until we have worked out what object is meant by the plural noun phrase

> the meanings of the parts of e and of the syntactic rules by which they are combined.

Finally there is the phrase 'is a function of'. Some writers claim that we need to say what class of functions we have in mind. Not so; to narrow down the class of functions is to add to PoC, not to explain it. However, the phrase 'is a function of' does have several readings, some of them with causal or epistemological overtones. Writers who formalise the property of compositionality (for example Dowty *et al.* [10, p. 42], Janssen [23, p. 447ff]) agree in taking it in the most straightforward mathematical sense, and I shall follow suit. We can express the result as follows. (We say that two expressions are *synonymous* if they have the same meaning.)

> (Functionality) Suppose a complex expression e is meaningful; suppose that e_1, ..., e_n are distinct non-overlapping meaningful constituents of e, and f is a meaningful expression which comes from e by replacing each e_i by an expression f_i; suppose also that for each i, e_i is synonymous with f_i, and that certain structural conditions are met. Then e and f are synonymous.

We shall have to spell out what the 'certain structural conditions' are. But that will also take care of the plural noun phrase above and the problem about which constituents of a meaningful expression are meaningful.

2. Synonymy

It springs to the eye that the statement of Functionality doesn't mention meanings directly; it mentions only meaningfulness and synonymity. More precisely, if a semantics has a meaning function μ, then it also has an associated synonymity relation $\equiv_\mu$ defined as follows.

> If e and f are expressions, then $e \equiv_\mu f$ if and only if e and f are both meaningful and $\mu(e) = \mu(f)$.

Let us define a *synonymy* to consist of interpretations for EXPRESSION, SYNTACTIC RULE and SYNONYMOUS. Then each semantics gives rise to a synonymy, and the property of compositionality can be paraphrased as a property of the synonymy rather than the semantics. Again we call the synonymy *tidy* if this paraphrase is meaningful. For example, in a tidy synonymy the relation 'synonymous' is an equivalence relation on a set of expressions.

Would anything be lost if we regarded compositionality as a property of synonymies rather than of semantics? Yes, just one thing: a synonymy doesn't assign meanings. From a purely formal point of view this is no problem. If we have a tidy synonymy, we can construct a tidy semantics from it by taking the meaning of an expression to be its equivalence class under the synonymity relation; then the associated synonymy of this semantics is exactly the synonymy that we started from. Let us call this semantics the *equivalence-class semantics* associated with the synonymy. (There is a set-theoretic problem that the equivalence classes might be proper classes; but there are set-theoretic solutions too. No linguist should lose a night's sleep over this.)

Of course, given any synonymy Σ, the equivalence-class semantics is only one of the possible semantics that have Σ as their associated synonymy. A good semantic theory will rightly look for good representations of the equivalence classes, and 'good' here includes virtues like 'easy to handle' or 'of explanatory value'. Let me pick out this important part of the endeavour as the *representation problem*. Let me note at the same time that PoC says *nothing whatever* about the representation problem.

3. Problems for (not-)compositionality

Many of the items commonly listed as 'problems for compositionality' consist of semantics which are untidy. For most of these, the issue is not the content

of the theory but the question which of the outputs of the theory should be labelled 'meaning'.

For example, the literature contains many semantics that don't deliver their output in the form 'the meaning of expression e is M'. One instance is Tarski's famous semantics in [36]. It yields results in the style

> Assignment v satisfies sentential function x.

So it is neither a tidy semantics nor a tidy synonymy, and we need to translate it into our terminology. What is an expression, what is a syntactic rule and what is synonymy? There may be semantics for which this is a real problem. But at least for Tarski's semantics two of the answers are easy and obvious:

> 'Expression' = 'sentential function';
> 'e is synonymous with f' = 'e and f are satisfied by the same assignments'.

(I leave syntactic rules to the reader.)

The reasoning behind this translation of 'synonymous' runs as follows: *two expressions should be synonymous if and only if they are indistinguishable for semantic purposes.* I shall call this the *no-difference principle.* If we want a semantics rather than a synonymy, we can pass to the equivalence-class semantics as in the previous section. But logicians younger than Tarski have often used other solutions of the representation problem; for example many have followed Frege and Peirce by taking the meaning of all true sentences to be Truth.

The semantic theory of Tarski's paper [36] is highly abstract. Probably the simplest way of adapting it to model theory, tense logic or Montague grammars, for example, is to put some appropriate restrictions on the assignments to variables. Then in tense logic the appropriate notion of synonymity might be:

> 'e is synonymous with f' = 'for every time t and every assignment
> v, v satisfies e at t if and only if v satisfies f at t.'

Likewise for other logics. In all cases we count two expressions as synonymous if they are semantically indistinguishable; the no-difference principle applies. A standard solution to the representation problem in cases like these is to take meanings as functions whose domain is, say, the set of pairs of the form (point of time, assignment).

Note that nothing in this section so far has anything to do with the presence of variables. Even for a purely propositional logic, Tarski's semantics speaks in the style

> Sentence s is true.

So we need a translation, just as before.

In all these cases, logicians are apt to see the various translations as one and the same semantics in different notations, and they are amazed to find Tarski's semantics described as a problem for compositionality. No doubt it is right and proper that the no-difference principle should be spelt out somewhere. But it solves a problem of communication across cultures, not a problem for compositionality.

The no-difference principle also applies to those semantics which give the same expression more than one meaning according to the physical or linguistic context in which the expression is used. This includes semantics for indexical expressions such as demonstrative pronouns. A standard solution to the representation problem for indexicals is to make the meaning a function from contexts of use to references.

The same principle also applies to the curious examples studied by Pelletier [30] where there is no lexical or structural ambiguity but the sentence still has more than one possible reading. More generally it applies to semantics which give an expression several possible meanings, relying on a filter to choose between them at some higher level in the sentence or context of use. In these cases one possible solution to the representation problem is to take the 'meaning' to be a set of alternative readings. Katz and Fodor [25] suggest using this kind of representation systematically, and Cooper storage [8] is an example of the same idea. Some writers have suggested that a meaning should be a set of possibilities with probability weightings attached.

Many of these examples involve ingenious and hard-won solutions to the representation problem. For example Montague semantics, published mostly in the 1970s, was the fruit of work begun in 1959 ([27], p. 96). But to repeat, the representation problem has nothing to do with compositionality; there is no hint of it in PoC.

I take one more example which is interesting for historical reasons. Montague several times referred to what he called *Frege's functionality principle*, or more briefly *Frege's principle*. He stated it as follows ([27] p. 128, see also p. 101, 159):

> [...] Frege's functionality principle applies fully to our notion of extension: the extension of a formula is a function of the extensions (ordinary extensions) of those of its parts not standing within indirect contexts [...], together with the intensions [...] of those parts that do stand within indirect contexts.

In brief, Frege said at the end of his essay 'On sense and reference' [12] that the reference (Bedeutung) of an expression in certain linguistic contexts is not its customary reference but its sense (Sinn). Here the no-difference principle tells us that we should count two expressions as synonymous in Frege's semantics if they have the same reference and the same sense. An easy representation would be to take the meaning of an expression to be the ordered pair of its reference and its sense. I doubt that Frege would have cared one way or another about our choosing to tack this terminology onto his theory, provided that we undertook never to use the word 'meaning' in any other sense.

Note that what Montague calls 'Frege's principle' is, on the surface at least, a denial of PoC. This should ring some warning bells about the historical origins of compositionality. I come back to this in section 8 below.

Theo Janssen very kindly sent me his thesis [22] and pointed out a discussion for which the no-difference principle would be no help. In his Chapter V.3 he offers some examples to illustrate the 'heuristic value of the principle of compositionality'. Sections 3.2 and 3.3 contain a critique of a proposed semantics. In fact Janssen makes three criticisms: (i) The semantics is not everywhere defined, since one of the recursive meaning rules gives no defined output for some of the possible inputs. (ii) The semantics is 'semantically incorrect' (i.e. in some cases it gives an output different from the intended one). (iii) The functionality is not expressed in terms of polynomials.

Certainly (i) and (ii) are faults in a semantic theory. But the first is a failure of tidiness and the second has nothing to do with compositionality. What fails in the third is a special requirement added on top of compositionality. (These are my comments, not Janssen's.)

4. The source of Frege's Principle

The argument which led Frege to his Principle had two parts. First, he had examples where the reference of a sentence was changed by changing the indirect reference (or Sinn) of a constituent while keeping its customary reference fixed. And second, he had a general rule ([12]):

> If we now replace one word of the sentence by another having the same reference [...], this can have no bearing upon the reference of the sentence.

(If I were Montague, which indeed I am not, I would have used the phrase 'Frege's functionality principle' for this quotation rather than for the conclusion of the same essay.) Presumably the intuition behind this rule is that the reference of an expression should include *at least* whatever it is that the expression contributes to the references of complex expressions containing it. But rather than dig into the mind of Frege, let us translate his rule into a condition not on reference but on meaning. Namely:

> If e is a meaningful expression and f [is a meaningful expression which] comes from e by replacing a constituent e_1 by an expression f_1 with the same meaning as e_1, then e and f have the same meaning.

This is a functionality condition of very much the kind that appears in the PoC.

The square brackets mark a question that Frege seems not to have considered. If e in the rule is meaningful but f is not, does this imply that e_1 and f_1 have different meanings? Let us say that two expressions e_1 and f_1 *have the same category* if whenever e_1 occurs in a meaningful expression, we still get a meaningful expression if we replace e_1 by f_1, and vice versa. This is an equivalence relation, so that the definition makes sense. The question is whether two expressions with the same meaning must also have the same category.

Tarski discusses the matter in his paper on the concept of truth ([36], p. 215ff). By implication he adopts the answer Yes, and suggests that in some form one can trace this answer back to Husserl. For English one can test one's intuitions on pairs of words like 'asleep' and 'sleeping' (the participle, not the verbal noun). The two words have different categories, as one sees from the pairs

　　　He was fast asleep.
⋆　　He was fast sleeping.
　　　He was sleeping off a hangover.
⋆　　He was asleep off a hangover.

It is not easy to find a sentence where either word will do but the choice affects the meaning of the whole sentence.

I have no wish to legislate; I just want to understand what the possible principles are. Frege's version of functionality, without the words in square brackets, is equivalent to the conjunction of the two following principles:

> (Husserl's principle) If two expressions have the same meaning then they have the same category.

> (Strong functionality) If e is a meaningful expression and f is a meaningful expression which comes from e by replacing a constituent e_1 by an expression f_1 with the same meaning as e_1, then e and f have the same meaning.

Strong functionality is Frege's rule with the words in square brackets included.

Now there are two ways of applying strong functionality to a particular case. The first is where one already knows that e_1 and f_1 have the same meaning and that e and f are meaningful, and one deduces that e and f have the same meaning; this is working bottom-up. In the second case one already knows that e and f have different meanings and that e_1 and f_1 have meanings, and one deduces that the meanings of e_1 and f_1 are different; this is reasoning top-down. In [12] Frege did both.

In both cases we already know something about which expressions are synonymous and which are not, and we apply strong functionality to guide us in extending the notion of meaning or synonymity to a wider class of expressions. I shall call this the *Extension Problem*: given a semantics or synonymy which doesn't cover all expressions, to extend the notion of meaning or synonymity so as to cover more expressions.

5. The Composition of the Principle

Now we can come back to the ambiguities of PoC. Strong functionality is evidently a good candidate for the functionality statement; but there is no reason to limit to one replacement at a time. There are two versions of functionality, one that implies Husserl's principle and one that doesn't; the square brackets below mark where one has to make the choice. For either version to make sense, we need to know that the constituents of a meaningful expression are meaningful. Thus:

Strong compositionality 1

(Domain rule) If an expression is meaningful then so are its constituents.

(Functionality) Suppose e is a meaningful complex expression e; suppose that $e_1, \ldots, e_n$ are distinct non-overlapping constituents of e, and f is [an expression/a meaningful expression] which comes from e by replacing each e_i by an expression f_i; suppose also that for each i, e_i is synonymous with f_i. Then e and f are synonymous.

Given the usual assumptions about syntactic rules, the version without Husserl's principle is equivalent to the following, where we write $\mu(e)$ for the meaning of e:

Strong compositionality 2

(Domain rule) If an expression is meaningful then so are its constituents.

(Functionality) There is a function ϕ such that for every meaningful complex expression e of the form $\alpha(e_1, \ldots, e_n)$,

$$\mu(e) = \phi(\alpha, \mu(e_1), \ldots, \mu(e_n)).$$

(The proof of equivalence is not deep. See [20].) This is the 'rule-by-rule' compositionality of Bach [1].

Strong compositionality 2 is sometimes paraphrased in terms like 'the meaning is defined by recursion on the syntax'. This is a useful shorthand, but it has the power to mislead. The definition would still be by recursion on the syntax if the condition on ϕ read:

$$\mu(e) = \phi(\alpha, \mu(e_1), \ldots, \mu(e_n), e_1, \ldots, e_n).$$

But this doesn't entail any recognisable form of PoC. Cooper [8, p. 11] states the point neatly: 'The issue [of compositionality] has arisen, I think, not because people really wish to deny that semantics can be defined recursively on the syntax, but because they have conceived of the principle of compositionality as saying something more than this.' (However, the 'more' that he adds seems to be the tidiness assumption, not functionality.)

Janssen [23, p 454] gives an illuminating example of a semantics which is not strongly compositional. The expressions are finite strings of numerals $0, \ldots, 9$. There is just one rule for generating complex expressions: add a numeral at the lefthand end of an expression. So Husserl's principle holds vacuously, since all expressions have the same category. The meaning of an expression is its natural number value in Arabic notation. For example 4

and 004 both have the meaning 4. But 3004 and 34 have different meanings, so strong compositionality fails. (This example is particularly pleasing since it belongs to what is virtually a natural language.)

The no-difference principle makes short shrift of this example. No two expressions make the same contribution to the meanings of expressions that contain them; so synonymity is identity. But there is another way of looking at the example: when we replace an expression by a different expression with the same meaning, we alter the syntactic structure of the containing expression. So Janssen's semantics does satisfy the following weaker condition, which is again a way of reading the Functionality clause in section 1:

Weak compositionality 1

(Domain rule) If an expression is meaningful then so are its constituents.

(Functionality) Suppose e is a meaningful complex expression e; suppose that $e_1, \ldots, e_n$ are distinct non-overlapping constituents of e, and f is [an expression/a meaningful expression] which comes from e by replacing each e_i by an expression f_i; suppose also that for each i, e_i is synonymous with f_i and has the same syntactic analysis. Then e and f are synonymous.

Again given the usual assumptions about syntactic rules, this statement in the non-Husserl form is equivalent to the following (writing $\mu(e)$ for the meaning of e again):

Weak compositionality 2

(Domain rule) If an expression is meaningful then so are its constituents.

(Functionality) There is a function ϕ such that for every meaningful complex expression e of the form $\tau(e_1, \ldots, e_n)$, where τ is a term in the syntactic rules and $e_1, \ldots, e_n$ are the atomic expressions in e,

$$\mu(e) = \phi(\tau, \mu(e_1), \ldots, \mu(e_n)).$$

Weak compositionality 2 still makes sense if we replace the domain rule by the weaker statement: If an expression is meaningful then so are all its atomic constituents (i.e. the words).

The weak compositionality property is extremely weak. Any counterexample would have to consist of two sentences of exactly the same grammatical structure which have different meanings although the words in one have exactly the same meanings as the corresponding words in the other. Wouldn't this be strong evidence that some of the words were being not just used but quoted?

The 'original notion of compositionality' in Seuren [34, p. 19]:

> A compound expression is semantically compositional just in case its meaning is a function of the meanings of each constituent part and from the position of these parts in the structure at hand.

seems to be exactly equivalent to weak compositionality 2. Seuren himself slightly muddies the issue by suggesting at once that 'the original notion of compositionality' means that meanings can be computed by recursion on the syntax.

6. The extension problem

If it is correct that Frege's functionality rests on the idea that the meaning of an expression includes *at least* what the expression contributes to the meanings of expressions containing it, then we have an obvious candidate for a principle in the other direction. Viz. the meaning of an expression includes *at most* what the expression contributes etc. etc. Or using the same style as the functionality principles above,

> If e_1 and f_1 are expressions which [have the same category but] don't have the same meaning, then there are an expression e and an expression f which comes from e by replacing an occurrence of e_1 by f_1, such that e and f don't have the same meaning.

(Again in the presence of Husserl's principle we can leave out the words in square brackets.) One can see from its form that this rule is not going to be much use in the bottom-up direction, failing any clues about where to look for e and f. But top-down it is quite powerful, if we already have a semantics for a class K of expressions which is in some sense typical for the whole language, and we want to extend the semantics to constituents of these expressions.

Namely, if K is a class of sentences, the principle reads:

> (Full abstraction relative to K) Let e_1 and f_1 be expressions with the property that if e is any expression in K which has e_1 as a

> constituent, and f comes from e by replacing e_1 by f_1, then e
> and f have the same meaning. Then e_1 and f_1 have the same
> meaning.

(The name 'full abstraction' comes from an analogous notion studied in computer science by Robin Milner [26] and Gordon Plotkin [31]). If you believe, for example, that the meaning of an expression is the contribution that it makes to the meanings of *sentences* containing it, then you might well adopt full abstraction relative to sentences, side by side with strong functionality, as a way of expressing this.

Full abstraction relative to K gives a special role to the class K of expressions. This makes full abstraction relative to K unlike all the other principles that we have considered, including PoC.

While we are discussing extension principles, we should distinguish between two kinds of extension problem which are formally similar, though in practice one knows that they are completely different. The first problem is the one handled by full abstraction: we have a class of 'controlling' expressions whose meanings are known, and we use this class to guide the extension of meaning to other expressions.

The second problem is where we already have a semantics which gives meanings to all expressions of a language E, and we extend the language to a larger language F by adding some new words — or possibly some new expressions containing old words. In this case a new expression f can't be a constituent of an expression already in E, so there is no question of using full abstraction in some class of expressions of E to guide the choice of meaning for f. Some genuine new semantic input will be needed. But then the new semantic information may clash with what we already know about E.

No new expression can be a constituent of an old expression. So the problem can only arise top-down, going from expressions e, f at least one of which is new, to constituents e_1, f_1 which are old expressions. (As before, f comes from e by replacing an occurrence of e_1 by f_1.) If e and f have the same meaning, nothing follows from any principle that we have stated. But if e and f have different meanings, strong functionality implies that e_1 and f_1 must have different meanings too. Yes, but we may already have decided for E that e_1 and e_2 have the same meaning; so the new notion of meaning fails to be an extension of the old one.

This is precisely the point that Frege reached in [12] when he considered what would happen if one broke free of the mathematical vocabulary that he had studied earlier, and added the words 'believes' and 'after' to the

language. He gave examples to show that a person can believe a sentence and disbelieve another sentence with the same truth-value but a different Sinn. So he rightly inferred that a suitable notion of meaning for the expanded language has to notice difference of Sinn as well as difference of truth-value. (His solution for 'after' was similar but more complicated.) Today an author might make the same kind of point about the results of adding the word 'typical', or of introducing a device for forming comparatives of adjectives.

I mention in passing, since I have nothing new to say about it, that you can cripple a semantics not only by adding new concepts to the language, but also by using the original language in more elaborate ways. For example it is often argued that for everyday use of language we need to know not only what is red and what is not red, but also what is *paradigm* red. Fodor [11] Chapter 5 protests that there are going to be serious problems in finding a compositional semantics that meets this condition, bearing in mind that if you start from paradigm red and paradigm fire engines, there is no reason why you should finish up with paradigm fire-engine red. (See also Kamp and Partee [24].)

If you believe both strong compositionality and, say, full abstraction in sentences, then certain things follow. Given two expressions e and f which have the same meaning in your language E, you have to allow that either (a) there could be a larger language in which these same expressions e and f have different meanings, or (b) there couldn't. Each possibility leads to problems. The problem for (a) is that one thinks of the meaning of an expression as something about that expression and not about the language as a whole. How could we learn the language if we had to know the meaning of every expression in order to be sure of knowing the meaning of a single expression?

The problem for (b) is that it is hard to legislate for what might in principle be added to a language. Suppose we didn't have the concept of belief. Then it might simply not occur to us that one might add expressions to the language which distinguish one Sinn from another.

The paper [20] contains two theorems that are relevant to this section. The first says the following:

> (Extension theorem) Suppose a semantics μ is given for a class K of expressions which is large enough that every expression of the language occurs as a constituent of some expression in K. Suppose also that μ obeys Husserl's principle and a kind of

strong compositionality within K. Then there is a fully abstract strongly compositional semantics ν for the whole language, which extends μ, is fully abstract with respect to K and obeys Husserl's principle. Any two such semantics ν have the same associated synonymy.

There is probably a version of this theorem without Husserl's principle, but without that principle the condition of strong compositionality within K seems impossibly complicated.

The second theorem describes the situation discussed above, where the language is extended, and gives sufficient conditions under which the fully abstract semantics for the smaller and the larger language agree with each other.

7. Problems for compositionality

The 'problems for compositionality' that we considered earlier were not problems about compositionality at all; they were mostly problems about how to translate a semantics into a certain format. But now we have two reasonably precise versions of the compositionality property, and we can turn to problems about semantics which don't have one or other version of this property. In fact discussion usually centres on strong compositionality, because weak compositionality barely has any content.

One might expect to see problems of the form: here is a semantics that seems to express what we wanted of a semantics, but it isn't (strongly) compositional. In fact the literature contains barely any problems at all of this kind. The example of Janssen in section 5 was a rare exception. To show what happens more typically, I summarise two examples from opposite ends of the debate.

First example: West Greenland Eskimo. We look at a paper of Maria Bittner [2] on quantification in West Greenland Eskimo. Bittner described various scenarios to several Eskimos, and asked them to say which of certain sentences of Eskimo were true in which scenario. She is satisfied that the results gave her reliable translations of these sentences into English. (For example one sentence meant 'Almost all the boys who got a balloon broke it within ten minutes'.) The root meanings of the Eskimo words are not in doubt, but most words carry an array of suffixes whose role in fixing the meaning of the whole sentence is not entirely clear. Bittner begins her closing section with the words:

> By this point, it will be a truism that Greenlandic Eskimo poses some formidable challenges for compositional semantic theory.

She then states a form of strong compositionality 'for natural language semantics'.

Reading this, one might suppose that Bittner has given a semantics which describes a meaning for each expression, including the suffixes; but unfortunately her semantics is not compositional, and she sets it as a challenge to find one that is. But unless I have seriously misunderstood the paper, this is not the case. Bittner has only a partial description of the meanings of expressions which are not sentences, and the challenge is to give a complete description. As things stand, there is not yet a semantics to test for functionality.

Second example: independence-friendly logic. We turn to a formal language designed by Jaakko Hintikka and Gabriel Sandu [17]. This language has the syntax of ordinary first-order logic, except that quantifiers (Qx) can be replaced by slash quantifiers $(Qx/yz\dots)$ where $yz\dots$ is a string of variables distinct from x, and there are some restrictions on the binding of these new variables. The intended sense of $(\exists x/y)$ is 'there exists an x independent of y'. The semantics runs as follows. For each sentence we define a game by induction on the construction of the sentence. Slashed variables represent places in the game where players have imperfect information. The sentence is true if player $\exists$ has a winning strategy for the corresponding game. The Hintikka-Sandu language extends first-order logic, and on sentences of first-order logic it gives exactly the same truth-values as the familiar Tarski semantics. But unlike Tarski's semantics, the game-theoretic semantics gives no interpretation for formulas which are not sentences. Hence again the semantics is not compositional — not because functionality fails, but because the domain rule fails.

This problem was easier to crack than Bittner's. We now have a strongly compositional semantics in the Tarski style which interprets every formula of Hintikka-Sandu logic, agrees with their semantics on sentences and is fully abstract with respect to sentences ([18], [19]).

In both these examples the problem that needs to be solved is the extension problem. The semantics that we have in hand violates the domain rule, not the functionality principle. Now one might say:

> Yes, but the problem is not just to find *any* meaning for the expressions which haven't yet been given one — otherwise we

> might as well follow Frege and make them all mean the moon.
> The new meanings have to be *correct*. The functional part of
> compositionality says what correctness amounts to.

This defence is fine except for the last sentence.

Certainly one reasonable requirement on a semantics is that two expressions have the same meaning if and only if they make the same contribution to the meanings of sentences containing them. We have seen that this is two requirements. Strong compositionality is 'only if'; 'if' is full abstractness, which is not any part of PoC. (The paper [20] gives an example where an extension of a semantics is strongly compositional and not fully abstract.)

There are other reasonable requirements. For example an important requirement often made is that it should be possible to *come to understand* a sentence by first understanding its constituents, starting from the smallest ones and working upwards. Let me call this the requirement of *compositional understanding*. It is certainly not any part of the functionality condition in PoC, because (as we saw) that condition can be paraphrased in terms of the associated synonymy, and it would obviously be hopeless to try to express compositional understanding as a condition on the associated synonymy.

Another very reasonable requirement is that it should be possible to explain the meaning of a word in terms which use only that word (and perhaps a few others); the meaning of a word should not be something which changes when new words are added to the language. (In the talk in Żegań on which this paper is based, I asked the audience to consider what influence it would have on the meaning of the sentence

> Jestem bardzo podobny do ojca.

if the following Polish words were abolished:

> administracja, balet, chmura, doskonale, entuzjazm, fajka,
> grubas, herbata, ...

You can make up your own example.) The extension theorem mentioned in section 6 gives no guarantee that this is possible. For example the condition holds for the semantics for Hintikka-Sandu logic given in [18] and [19], but not for the equivalence-class semantics got by applying the extension theorem to the same problem, although these two semantics have the same associated synonymy. This is one major reason why the semantics in [18] and [19] is better than the other.

8. Historical notes

8.1. The name 'compositionality'

Although the -al ending has been creative in English for centuries, the word 'compositional' seems to be recent; the 1933 Shorter Oxford English Dictionary [35] shows no knowledge of it. To the best of my knowledge, both 'compositional' and 'compositionality' came into the semantic literature through the paper [25] of Katz and Fodor. In their discussion of lexical meaning, for example, they say

> As a rule, the meaning of a word is a compositional function of the meanings of its parts, and we would like to be able to capture this compositionality.

They also discuss sentence meaning, and here they introduce what they call 'type 1 and type 2 projection rules'. A semantics using only type 1 projection rules is strongly compositional in our sense, though it is not entirely clear what features Katz and Postal meant to capture by their use of the word. Earlier in the same essay they say

> Since the set of sentences is infinite and each sentence is a different concatenation of morphemes, the fact that a speaker can understand any sentence must mean that the way he understands sentences he has never previously encountered is compositional: on the basis of his knowledge of the grammatical properties and the meanings of the morphemes of the language, the rules the speaker knows enable him to determine the meaning of a novel sentence in terms of the manner in which the parts of the sentence are composed to form the whole.

This points not to PoC but to the notion of compositional understanding (section 7 above). Possibly both these passages from Katz and Fodor owe something to the remark of Chomsky [7, p. 15], that in order to produce or understand indefinitely many new sentences, a speaker needs to *project* (Chomsky's italics) from the finite corpus of observed utterances.

8.2. The phrase 'function of'

Before writing this section I downloaded and examined a page-worth of examples of the phrase 'is a function of' from the internet, courtesy of Alta-Vista. The reader might find it illuminating to do the same.

I am guessing that the phrase 'function of' in PoC traces back to Montague's use of the same phrase in his statements of 'Frege's principle' (and then in turn to Carnap, cf. section 8.3 below). We have already seen that Montague's 'Frege's principle' is not PoC — in fact it heads off in a quite different direction — but that one of Frege's supporting arguments for it is arguably a statement of the functionality part of PoC.

The relevant usages can be paraphrased in the form

> The X of an E is a function of the Y of the E.

As I noted earlier, there can be causal overtones here:

> We can change $X(E)$ by changing E so as to change $Y(E)$.

Or epistemological:

> If you want to know about the X of an E, the first place to look
> is its Y.

The mathematician's rendering that we used in section 1 has neither of these overtones. We can render it as follows:

> If $Y(E) = Y(F)$ then $X(E) = X(F)$.

Montague's use of the phrase is closer to the causal version. For example on [27, p. 228] he argues that 'if the interpretation of a compound is always to be a function of the interpretations of its components', then anything needed to fix the interpretation of a component will in general be needed to fix the interpretation of the compound too. The mathematical reading makes nonsense of this argument. What we need is something more like:

> If $Y(E) \neq Y(F)$ then in general $X(E) \neq X(F)$.

A rough paraphrase of Montague's version is that the X of an E is a mathematical function of *at least* the Y of the E; in other words, it is *not* a mathematical function (in the earlier sense) of something less than Y. (Montague's sense is not far from an example that AltaVista gave me: 'The quality of your essay is a function of the effort that you put into it'. This does *not* mean that if two of you put the same effort into your essays, the essays will have the same quality!)

In short, Montague's 'function of' is not the 'function of' in PoC as standardly formalised. I do strongly suspect that some interference from

Montague's usage lies behind the bid to describe the items in section 3 above as 'problems for compositionality'.

8.3. The idea of compositional semantics

Katz and Fodor were certainly not the first people to suggest a compositional semantics. Probably that honour should go to George Boole [3, p. 49ff], if one counts his translation from logic to sets as a kind of semantics.

Frege's role is confused. His [12] contains one of the clearest statements of the functionality part of PoC, in terms of reference (Bedeutung). Nevertheless in that same paper he chose to present his semantics in a style which is 'untidy' (in our sense above). If he had given the details, it might well have been a triviality (as it is with Tarski) to paraphrase his theory into a strongly compositional style; but there is no evidence that this task would have interested Frege.

In 1947 Rudolf Carnap [6] generalised Frege's statement of functionality for reference, and called the result Frege's Principle. Montague's Frege's Principle is a revision of Carnap's.

Seuren [34, p. 18] claims to find the notion of weak compositionality in Frege [13, p. 303]. Here Frege says that in a well-constructed definition the sense of the definiens is determined by the senses of the signs of which it is composed. The idea of weak compositionality is certainly there, but as a property of well-formed definitions, not as a property of semantics.

After Frege the next major contribution comes from Alfred Tarski [36], whose truth-definition is (apart from the style of the presentation) very clearly compositional in the strong sense, and very clearly intended to be. Here is Tarski's own formulation ([36], p. 214):

> I have pointed out that in drawing up a correct definition of the concept of satisfaction use can be made of recursive definition. For this purpose it suffices — recalling the recursive definition of sentential function and bearing in mind the intuitive sense of the primitive sentential functions and the fundamental operations on expressions — to establish two facts: (1) which sequences satisfy the fundamental functions, and (2) how the concept of satisfaction behaves under the application of any of the fundamental operations (or to put it more exactly: which sequences satisfy the sentential functions which are obtained from given sentential functions by means of one of the fundamental operations, assuming that it has already been established which sequences satisfy the sentential functions to which the operation is applied).

Tarski's paper also contains a section (p. 209ff) discussing sufficient conditions for a formal language to have a strongly compositional semantics. This is the section in which Tarski refers back to Husserl's notion of semantic category. (Tarski gives no reference, but I take the following from Ajdukiewicz: [21], pp. 294f, 305–312, 316–321, 326–342.)

Tarski's student Montague [27] adapted Tarski's truth-definition to fragments of English, and many linguists first became aware of Tarski's idea through Montague. Montague's semantic framework for fragments of English rested partly on what I called the no-difference principle, but also on a general methodological principle that expressions which are semantically comparable should have meanings of similar forms. He inherited this from Tarski; a typical example is that Tarski and Vaught [37] give the meaning of a formula in terms of a set of assignments to *all* variables, not just those which are free in the formula. This is relevant to Montague's use of the notion 'function of'.

Meanwhile in 1967 Donald Davidson [9, pp. 17ff] proposed a theory of meaning to which philosophers have attached the name 'compositional'. The accounts known to me are too problematic to allow any straightforward comparison with the strong and weak properties of compositionality. As a fairly typical example:

> a *compositional (truth-theoretic) semantics* for a language L is a finitely
> statable theory that ascribes properties to, and defines recursive con-
> ditions on, the finitely many vocabulary items in L in such a way that
> for each of the infinitely many sentences of L that can (in principle)
> be used to make truth-evaluable utterances, there is some condition
> (or set of conditions) such that the theory entails that an utterance of
> that sentence is true iff that condition (or a certain member of the set)
> obtains. (Schiffer [33], p. 178)

We note: (1) There is no mention of functionality, except perhaps in the word 'recursive'. (Davidson's own use of the word 'recursive' is so vague that one can't easily tell whether the root idea is strong compositionality, computability, definition by recursion on the syntax, or something quite different; see for example [9], pp. 21, 30 footnote, 57f.) (2) A key part of the definition is a requirement that the semantical theory should entail certain sentences. But any inconsistent theory meets the requirement, because it entails all sentences. One supposes that Schiffer meant to imply something about the theory *not* generating some other sentences; but since Schiffer seems to have missed this point, it may be idle to speculate on what he

would have said had he faced it. (To be fair to Schiffer, there may be a similar problem about Tarski's choice of metatheory.)

8.4. Internal structures of meanings

One has to mention a passage from Gottlob Frege ([14], p. 36):

> It's amazing how fertile language is. With a few syllables it expresses unimaginably many thoughts; it finds a way of expressing a thought which some person has had for the first time in the history of the world, so that another person can understand it although it is completely new. This would be impossible if we were not able to separate thoughts into parts corresponding to the parts of a sentence, so that the construction of the sentence pictures the construction of the thought.

Also from [15]:

> The possibility of our understanding propositions which we have never heard before rests evidently on this, that we construct the sense of a proposition out of parts that correspond to the words.

It seems entirely reasonable to label Frege's idea in these two passages as 'compositional', since it is about how meanings are related to the composition of sentences. But I no longer know why I ever thought it was connected with PoC. PoC is about how the meaning of an expression is related to the meanings of its constituents. Frege's statement is not about that at all; it is about how the meaning of a sentence is a structured object, with a structure corresponding to the syntax of the sentence. As we have seen, both strong and weak compositionality can be expressed in terms of synonymy, and synonymy has nothing whatever to do with internal structure of meanings. Frege doesn't even say that the meaning structure of a complex expression must be build up from the meaning structures of its parts.

It was almost a cliché of idealist logic from the 1830s to the beginning of this century that a 'judgement' is built up out of 'ideas' corresponding to the words used to express it. In fact logicians of that school saw it as one of the central problems of logic to describe the glue that joins the ideas together in a judgement. So Frege's conclusion was not new. Nor was his argument from the infinity of possible meaningful sentences. In 1900 Husserl ([21], p. 326–333) had already remarked that there are 'Unendlichkeiten' of possible forms of sentences, and that infinitely many of them make sense to us and infinitely many don't. He argued that we must have a procedure

for distinguishing the meaningful from the meaningless by analysing a sentence into its semantic components; his theory of semantic categories was his attempt to describe this procedure. This seems to me a more perceptive conclusion than Frege's. (Frege's [14] is from 1923. Frege's [15] is undated, but Jourdain was born in 1879, the earliest writing of his in logic that I know of was in 1904, and he was certainly corresponding with Frege in 1913.)

In modern times Cann [5, p. 4] seems to recall Frege when he suggests that in a compositional semantics, the meaning of an expression 'maintains' the meanings of its constituents. (In fact he builds this into his definition of 'The principle of compositionality'.) I found this obscure. For example in propositional logic the truth-value of a proposition hardly 'maintains' the truth-values of its subformulas. One suspects that Cann has unwittingly generalised from an important and commercially valuable notion which occurs elsewhere in his book, namely *compositional translation*. See for example [32], [4].

9. Conclusions

First, the alleged ambiguity of the Principle of Compositionality may be rather overblown. The Principle states an axiom; some semantics satisfy the axiom and some don't. In the same way the mathematical definition of 'group' is an axiom; groups are those structures which satisfy it. Nobody counts this as an ambiguity in the definition of 'group'.

Second, there are indeed some ambiguities in the property of compositionality. Besides the well-known split between rule-by-rule functionality and a much weaker property, there are also the domain question ('Which expressions get meanings at all?') and some subtler points connected with Husserl's principle. These latter issues are hardly mentioned in recent discussions of compositionality

Third, many alleged problem cases for compositionality are not about compositionality at all, but about whether the semantics has been presented in a form where one could directly raise the question of compositionality. Some of these cases are genuine and taxing problems, but the difficulty lies in solving the representation problem, not in ensuring compositionality.

Fourth, most of the remaining problem cases for compositionality are about finding a semantics which obeys the domain rule; the usual problem is to extend a semantics for sentences to a semantics which gives meanings to constituents of sentences. Of course the problem is not to find just *any*

assignment of meanings; the assignment has got to be in some sense correct and informative. But the functional part of compositionality is only one of the requirements. In practice full abstractness and solution of the representation problem are equally important.

Fifth, a whole range of other notions have become attached to the Principle of Compositionality, although there is no way of reading them into the literal statement of it. Two notable examples are Frege's requirement that meanings should have internal structures corresponding to the syntax, and the requirement that it should be possible to understand sentences by working upwards recursively on their syntax.

Acknowledgements. My warmest thanks to Heinrich Wansing, Andrzej Wiśniewski and the other organisers of the meeting at the delightful Talleyrand Palace in Żegań in March 1998. I also thank Jeff Pelletier for pointing out to me Carnap's role, and Theo M. V. Janssen for various comments.

References

[1] Bach, E., "An extension of classical transformational grammar", in *Problems of Linguistic Metatheory*, Proc. 1976 Conference, Michigan State University.

[2] Bittner, M., "Quantification in Eskimo: A challenge for compositional semantics", in *Quantification in Natural Languages*, ed. E. Bach *et al.*, Kluwer, Dordrecht 1995.

[3] Boole, G., *The Mathematical Analysis of Logic*, Macmillan, Barclay and Macmillan, Cambridge 1847.

[4] Bos, J., *et al.*, *Compositional Semantics in Verbmobil*, Proceedings of COLING '96.

[5] Cann, R., *Formal Semantics*, Cambridge Unviersity Press, Cambridge 1993.

[6] Carnap, R., *Meaning and Necessity*, University of Chicago Press, Chicago 1947.

[7] Chomsky, N., *Syntactic Structures*, Mouton and Co., 'S-Gravenhage 1957.

[8] Cooper, R., *Quantification and Syntactic Theory*, Reidel, Dordrecht 1983.

[9] Davidson, D., *Inquiries into Truth and Interpretation*, Clarendon Press, Oxford 1984.

[10] Dowty, D. R., R. E. Wall, and P. Peters, *Introduction to Montague semantics*, Reidel, Dordrecht 1981.

[11] Fodor, J. A., *Concepts*, Clarendon Press, Oxford 1998.

[12] Frege, G., "Über Sinn und Bedeutung", Zeitschrift für Philosophie und philoso-
     phische Kritik 100 (1892), 25–50. Translated as "On sense and reference" in
     Meaning and Reference, ed. A. W. Moore, Oxford University Press, Oxford
     1993, p. 23–45.

[13] Frege, G., "Über die Grundlagen der Geometrie", Jahresbericht der Deutschen
     Mathematiker-Vereinigung 15 (1906), 293–309.

[14] Frege, G., "Gedankengefüge", Beiträge zur Philosophie des deutschen Idealis-
     mus III (1923), 36–51.

[15] Frege, G., "Letter to Jourdain", in *Meaning and Reference*, ed. A. W. Moore,
     Oxford University Press, Oxford 1993, p. 43–45.

[16] Hintikka, J., *The Principles of Mathematics Revisited*, Cambridge University
     Press, Cambridge 1996.

[17] Hintikka, J., and G. Sandu, "Game-theoretical semantics", *Handbook of Logic
     and Language*, ed. J. van Benthem, A. ter Meulen, Elsevier 1996, p. 361–410.

[18] Hodges, W., "Compositional semantics for a language of imperfect informa-
     tion", Logic Journal of the IGPL 5 (1997), 539–563.

[19] Hodges, W., "Some strange quantifiers", in *Structures in Logic and Computer
     Science*, ed. J. Mycielski *et al.*, Lecture Notes in Computer Science 1261,
     Springer, Berlin 1997, p. 51–65.

[20] Hodges, W., "Formal features of compositionality", preprint 1998,
     `ftp.maths.qmw.ac.uk/pub/preprints/hodges/formcomp.tex`.

[21] Husserl, E., *Logische Untersuchungen. Vol. ii, part 1*, Max Niemeyer Verlag,
     Tübingen 1993 (original 1900).

[22] Janssen, T. M. V., *Foundations and Applications of Montague Grammar*,
     Mathematisch Centrum, Amsterdam 1983.

[23] Janssen, T. M. V., "Compositionality", *Handbook of Logic and Language*, ed.
     J. van Benthem, A. ter Meulen, Elsevier 1996, p. 417–473.

[24] Kamp, H., and B. H. Partee, "Prototype theory and compositionality", Cog-
     nition 57 (1995), 129–191.

[25] Katz, J. J., and J. A. Fodor, "The structure of a semantic theory", Language
     39 (1963), 170–210.

[26] Milner, R., "Fully abstract models of the typed lambda-calculus", Theoretical
     Comp. Sci. 4 (1977).

[27] Montague, R., *Formal Philosophy: Selected Papers of Richard Montague*, Ed.
     R. Thomason, Yale University Press, New Haven 1974.

[28] Partee, B. H., "Compositionality", in *Varieties of Formal Semantics*, ed.
     F. Landman and F. Veltman, Foris, Dordrecht 1984, p. 281–312.

[29] Partee, B. H., A. ter Meulen and R. E. Wall, *Mathematical Methods in Linguistics*, Kluwer, Dordrecht 1990.

[30] Pelletier, F. J., "The principle of semantic compositionality", Topoi 13 (1994), 11–24.

[31] Plotkin, G. D., "LCF considered as a programming language", Theoretical Comp. Sci. 5 (1977).

[32] Rosetta, M. T., ed., *Compositional Translation*, Kluwer, Dordrecht 1994.

[33] Schiffer, S., *Remnants of Meaning*, MIT Press, Cambridge Mass. 1989.

[34] Seuren, P. A. M., *Semantic Syntax*, Blackwell, Oxford 1996.

[35] Little, W., *et al.*, *The Shorter Oxford English Dictionary*, Clarendon Press, Oxford 1933.

[36] Tarski, A., "The concept of truth in formalized languages", in *Logic, Semantics, Metamathematics*, ed. J. Corcoran, Hackett Publishing Co., Indianapolis 1983.

[37] Tarski, A. and R. L. Vaught, "Arithmetical extensions of relational systems", Comp. Math. 12 (1957), 81–102.

Wilfrid Hodges
e-mail: `W.Hodges@qmw.ac.uk`