Semantic Compositionality
The Argument from Synonymy

Francis Jeffry Pelletier

The Principle of Semantic Compositionality is the principle that the meaning of an expression is a function of, and only of, the meanings of its parts together with the method by which those parts are combined. As stated, the Principle is vague or under specified at a number of points such as "what counts as a part", "what is a meaning", "what kind of function is allowed" and the like. But this hasn't stopped some people from treating it as an obviously true principle, true almost by definition, nor has it stopped some others from attacking it both on "empirical grounds" and on theoretico-methodological grounds. It seems to me that many of these discussions fail because of a lack of precision on the above mentioned points and that other discussions are best described as "how compositionality can/cannot be accommodated within theory X" rather than whether the Principle is or is not true.

That is, the majority of the arguments against the Principle rely on having some specific grammatical framework in which to work and consist of arguing that some phenomenon cannot be compositionally treated within this theory. Some arguments rely on assumptions whose truth seems even more questionable than that of the Principle's, as for instance an argument which asserts that mental states are (or are not) compositional in nature and that since language is the mirror of our mental states therefore the Principle is (or is not) true. There are not, in fact, very many arguments in the literature in *favour* of The Principle, probably because almost all theorists – especially philosophically-oriented theorists – have a warm and fuzzy feeling about the Principle. They cannot even imagine what it might mean for it to be false and so they do not bother arguing in its favor but rather merely shake their heads sadly and knowingly at anyone whose theory supposes that the Principle is false.

So, although there may be various theory-internal grounds for adopting or denying The Principle, I claim that there are only four theory-independent arguments concerning it. There are two in its favor and two opposed. They are:

In favour	1	The argument from learnability
	2	The argument from understandability
Opposed	1	The argument from synonymy
	2	The argument from ambiguity.

The argument from learnability is that if a language lacked compositionality it would be un*learn*able. If the meaning of the whole were not a function of the meaning of its parts, this argument says, then we would not be able to *learn* the language. The only way we can learn an infinite (or even extremely large) set of sentences would be for us to learn a finite base and learn a finite number of ways of combining items. But if the Principle weren't true then this manner of learning would not help us in knowing what the constructed sentences meant, and therefore we wouldn't really have learned the language. The argument from learnability poses a choice between the Principle on the one hand, versus being able to learn only a small portion of language.

The argument from understandability is in a way the converse of the argument from learnability. Here it is argued that the Principle is the only explanation of how a finite mechanism (such as the human brain/mind) can *understand* an infinite set of sentences.[1] How else, this argument asks us, are we to be able to figure out the meaning of an arbitrary, new, novel sentence – if it isn't by the fact that we know some finite number of parts and finite number of ways of putting them together? How is it that we can understand a novel sentence, except by predicting its meaning from our understanding of the meaning of its parts and mode of combination? And, the argument concludes, this is all that the Principle says.

The argument from synonymy is the argument that if the Principle were true generally then there could be no synonymy, whether of sentences, of phrases, or of lexical items. And since the existence of these phenomena is so clear, the argument concludes that the Principle is incorrect.

The argument from ambiguity alleges that it is impossible for there to be cases of identical surface structure but distinct meanings, if the Principle is true. There could be no such example as *Every linguist knows two languages* being simultaneously ambiguous between the two obvious meanings while at the same time that sentence having only one surface structure. (There being only one structure, the Principle would claim that this maps onto only one meaning). The argument affirms that there is certainly this type of ambiguity, where there is one surface structure and more than one meaning. Therefore the Principle is wrong.

In this paper I wish to concentrate on just one of these arguments, the argument from synonymy. All four of the arguments are interesting and I hope to investigate each one at some time or other,[2] but for now I would

---

[1] These arguments need not be stated in terms of an infinite set of sentences. As Grandy has pointed out, they seem to work even if we just presume a hugely large but finite set of sentences.

[2] In "The Principle of Semantic Compositionality", I make superficial remarks about all four. Although the majority of my attention there is spent looking at what I think of

---

like to concentrate attention on this argument.[3]

The argument from synonymy is *against* The Principle. Thus it needs to prove that there is no *function* from meanings of parts and modes of combination to meanings of wholes that will preserve synonymy. This means I must display some parts-and-modes meanings which are non-functional, that is, ones which do not yield at most one meaning-whole. This is the only way to show there to be no such function as required by the Principle – find an example where one and the same group of parts-and-modes yields two or more meaning-wholes. But how shall we convince ourselves that we have more than one meaning? Here is a sufficient condition:

[A_1] If Φ and Ψ have the same meaning, then they must have the same truth value.

I do not wish to enter into discussion of whether meaning is to be identified with truth conditions or truth-in-all-possible-worlds, or whether a theory of truth is *eo ipso* a theory of meaning. I only wish to insist on the very weak assumption [A_1], giving a minimal relation between meaning and truth.[4] [A_1] is called "the most certain principle" in Bäuerle & Cresswell 1988.

There are many theories of meaning – too many, some would say. Some theorists would identify meaning with a set of possible worlds, others with certain intentions of speakers, still others with a function from possible worlds and contexts to truth values, yet others with a speech act potential, and there are even some who think it is an expression of some other language. Not wishing to choose amongst such theories, I would ask only that they accept that there are some sentences – at least two – for which assumption [A_2] holds, regardless of the theory of meaning and the theory of syntax involved.

[A_2] There are syntactically distinct sentences Φ and Ψ which mean the same.

There is, in fact, at least one theory of meaning according to which this apparently innocuous assumption is false: the theory of "complete structured

---

as the unconvincing arguments, I do spend some time discussing especially the argument from ambiguity and in making some remarks that might provide a different light on the arguments from learnability and understandability.

[3] The argument will be recognized as being implicit in numerous discussions in semantics and in philosophy of language; but it has not, I believe, previously been explicitly deployed against The Principle. See for example Church's "translation argument" in his 1950, Bigelow's 1978, and the works of Cresswell cited in the bibliography.

[4] In fact there is a theory of meaning according to which [A_1] is false: distinguish between meaning and what is expressed in a context. According to such theories, although a sentence has a meaning in isolation, it only expresses a proposition in a context, and is therefore only true or false in a context. Thus two sentences might have the same meaning and yet not have any truth value at all... except in a context. (For example, indexical sentences like 'We are there' and 'I am here'). To such theorists, I ask that they interpret [A_1] in such a way that Φ and Ψ are in a context, and therefore able to have truth values. With such a modified understanding, I think these theorists should accept [A_1].

meanings" which has sometimes been associated with Lewis 1971 and Cresswell 1988. But even these authors believe this "complete structure" theory to be too strong. The identification of meaning with syntactic structure plus associated functions on possible worlds gives, as they say, a much too fine-grained approach to meaning. Cresswell's response is to give a theory in which the amount of syntactic structure that is part of the meaning *varies* from occasion to occasion in which it is used. One might wonder whether this, all by itself, isn't a non-compositional theory. (Cresswell says it *is* compositional.) But such discussions are beyond the scope of this paper. I would prefer here to just put it as a challenge: I believe that no theory which has the consequence of denying our [A_2] can stand much of a chance of capturing our pre-theoretic intuitions about meaning and sameness of meaning.

To find counterexamples to the Principle, we need to know when complex expressions are put together by the same method of combination. Different theories will naturally have differing ways of describing "combinations of parts". For the purposes of the present discussion we needn't choose amongst such theories; all we need to assume is that in any one theory the relevant sentences are generated or analyzed in the same way. So, the third assumption is:

[A_3] In each syntactic theory, there is only one syntactic rule (or only one sequence of syntactic rules) which creates or analyzes sentences of the form:

$$Kim + believes + that + \text{Sentence}$$

from the four component parts.

As I said before, I don't wish to take a stand on what the exact syntactic structure of such sentences is, since different theories might assign different structures and I intend my argument to hold against them all. Nor do I want to rule out the possibility that a given syntactic theory might in fact allow there to be more than one rule or rule-sequence that generates that kind of sentence. I state it this way merely for convenience of exposition. The crucial point is that I want to have to consider only one syntactic rule (or rule sequence) which creates all these types of sentences (in any one syntactic theory), so that, given two applications of this rule, if the corresponding parts mean the same then the whole will mean the same (assuming the Principle to be true). If the syntactic theory allowed more than one analysis of such sentences we could still generate the argument below, but it would introduce needless complexity into having to keep track of when a sentence is analyzed by which of the sequences of rules. Finally here I should remark that although [A_3] is stated using 'believes' (or 'believes that'), we could have used any of a number of so-called opacity-creating verbs, for example 'sincerely claims' (or 'sincerely claims that'). This point will re-emerge in the discussion of [A_4].

Finally I claim that for any pair of syntactically distinct sentences it is always possible that a person not believe they mean the same – even if the sentences really do mean the same.[5]

[A_4] If Φ and Ψ are syntactically distinct sentences, then it is possible that exactly one of (i) and (ii) is true:

(i) *Kim believes that* Φ

(ii) *Kim believes that* Ψ

Of course, there are certain philosophical theories according to which [A_4] cannot be true: if a person believes that Φ, then he must also believe that Ψ, if Φ and Ψ mean the same. One such theory would claim that *believes* is always "transparent". Another related theory would say that the evaluation of Φ picks out a certain class of possible worlds – by hypothesis, the same class that Ψ picks out. So if Kim believes Φ, Kim must also believe Ψ: after all, Φ and Ψ are the same! (Same class of possible worlds, that is.) Bäuerle & Cresswell ("Propositional Attitudes", pp. 493ff.) consider such theories, and charitably attribute to them the view that "for purposes of a viable semantics we must treat [such sentences] as if they had the same truth value". This is not the place to debate such theories, and instead I will state my case by remarking that one *could* evade the thrust of my anti-compositional argument by denying [A_4]. But notice that such counter theories have to say (implausibly) that the person "*really* believes" Φ despite his protestations to the contrary. And had I replaced *believes that* in both [A_3] and [A_4] by *sincerely claims that*, then these counter-theories would be in the unpalatable position of having to claim that the person "*really* sincerely claims that Φ", despite his (sincere!) protestations to the contrary.

Now for the (brief) argument. By [A_2] there are two syntactically distinct sentences – call them S_1 and S_2 – which mean the same thing. By [A_3] it is the same rule (or sequence of rules) which, from these identically-meaning S_1 and S_2, forms or analyzes both of

(1) Kim believes that S_1

(2) Kim believes that S_2

So according to the Principle, (1) and (2) must mean the same thing. But by [A_4] it is always possible that one of (1) and (2) is true and the other is false. And so it follows then by [A_1], that sentences (1) and (2) have different meanings. So, [A_1]–[A_4] are inconsistent with the Principle.

A variant of this argument can be wielded against anyone who admits that there are pairs of distinct phrases (as opposed to sentences) which have

---

[5][A_4] needn't hold for all sentences – it need only assert that, for each syntactic theory there be some pair or other for which it is true, and that this pair satisfy [A_2] and [A_3]. (Arguably, Mary kissed John and John was kissed by Mary mean the same and that no one believes one without believing the other. If this be so, then choose some other pair of sentences which will satisfy the postulates.)

the same meaning. Consider, for example, the possibility that *a locus of all points on a plane equidistant from a given point* means the same as *a circle*. By analogy with [A₃], we would further assume that there is only one sequence of syntactic rules which forms or analyzes both of (3) and (4) from their parts.

(3)   A circle is a circle

(4)   A circle is a locus of all points in a plane equidistant from a given point

from its parts. Since (3) and (4) were formed by the same rule (sequence) from parts that by hypothesis have the same meaning, The Principle says that (3) and (4) mean the same. But (5) and (6) are, according to principle [A₃], formed or analyzed from (3) and (4) by using the same rule (sequence).

(5)   Kim believes that a circle is a circle

(6)   Kim believes that a circle is a locus of all points on a plane equidistant from a given point

Since the same rule is applied to items which mean the same, The Principle says that (5) and (6) mean the same. But [A₄] says that one of them can be true and the other false; and by [A₁] it then follows that (5) and (6) do *not* mean the same. So here the Principle is shown to be incompatible with synonymy of phrases.

The argument can be further extended to show the incompatibility of The Principle with word synonymy. This extension is of some additional interest to those theorists attracted to "structured meanings". It is tempting to adopt some variant of this "structured" theory of meaning, both for the reasons outlined in Lewis 1972 and Cresswell 1980 and for the fact that the theory of structured meanings appears to evade the preceding argument by denying identity of meaning of the sentences mentioned in [A₂] and of the phrases mentioned in (3) and (4). I think it will appear to be much less tempting when it is pointed out that the theory must also deny the possibility of lexical synonymy. For, if there were two words which meant the same, for example maybe *attorney* and *lawyer*, then we could mimic the preceding argument. By analogy with [A₃], we assume that there is just one sequence of syntactic rules which forms or analyzes both of

(7)   All lawyers are scoundrels

(8)   All attorneys are scoundrels

By hypothesis this one sequence is operating on parts (ultimately, meanings of words) which are identical. Therefore, according to the Principle, (7) and (8) mean the same thing. But once again [A₃] analyzes both (9) and (10) by the same rule sequence:

(9)   Kim believes that all lawyers are scoundrels

(10)   Kim believes that all attorneys are scoundrels

Again by hypothesis, (9) and (10) have parts that mean the same, namely, (7) and (8), and therefore (9) and (10) must mean the same. Yet according to [A₄] one of them can be true and the other false; and by [A₁] it follows that (9) and (10) must mean different things – thereby generating the contradiction once again.

Thus the Principle is incompatible with [A₁] – [A₄], and also with the extended notions of phrasal and lexical synonymy. Which should be rejected? I think there is no question that [A₁] and [A₃] must be maintained. So the choice is amongst: (a) the Principle, (b) that different strings of symbols can mean the same ([A₂]), and (c) that a person might believe something yet reject a synonymous claim ([A₄]). Of these three, surely the Principle is the most suspect. Who would ever wish to deny that there is *any* synonymy – neither word nor phrasal nor sentence synonymy (i.e., deny [A₂])? And who really thinks that the manifest facts concerning belief should be overthrown by something so theoretically-motivated as the Principle?[6]

References

Bäuerle, R. & M. Cresswell 1988 "Propositional Attitudes", in D. Gabbay & F. Guenthner *Handbook of Philosophical logic* 4, Dordrecht: Kluwer, pp. 491–512.

Bigelow, J. 1978 "Believing in Semantics" *Journal of Philosophical Logic* 9, 101–144.

Church, A. 1950 "On Carnap's Analysis of Statements of Assertion and Belief", *Analysis* 10, 97–99.

Cresswell, M. 1980 "Quotational Theories of Propositional Attitudes", *Journal of Philosophical Logic* 9, 17–40.

Cresswell, M. 1985 *Structured Meanings*, Cambridge, MA: MIT Press.

Lewis, D. 1972 "General Semantics", in D. Davidson & G. Harman (eds.), *Semantics of Natural Language*, Dordrecht: Reidel, pp. 169–218.

Pelletier, F.J. 1994 "The Principle of Semantic Compositionality", *Topoi* 13, 11–24.