

THE CASE FOR COMPOSITIONALITY

Zoltán Gendler Szabó
Yale University

A standard, theory-neutral way to state the principle of compositionality is as follows:

(C₀) The meaning of a complex expression is a function of the meanings of its constituents and the way they are combined.

The principle is always understood as involving a tacit ‘that’s all’ clause – it states that the meaning of a complex expression is a function of the meanings of its constituents, the way they are combined, and nothing else besides.¹ (C₀) tacitly quantifies over expressions of a language.² It holds for most formal languages, which is no great surprise given the extent to which it facilitates meta-linguistic proofs. The question is whether we can find this nice feature in languages we did not design. Stated in full generality, (C₀) is taken to be the claim that the meaning of any expression in any natural language is a function of the meanings of the constituents of that expression in that language, and the way they are combined in that language.

Conventional wisdom regarding (C₀) is that it is reasonably clear, fairly trivial, and more or less indispensable in semantic theorizing.³ I think conventional wisdom is wrong on all three accounts. First, there are significant ambiguities in (C₀). Moreover, no matter how the ambiguities are resolved, there is no familiar argument that adequately supports the truth of the principle. The usual considerations speak in favor of a considerably weaker thesis only. Finally, the compositionality principles discussed in

¹ The principle presupposes that the expressions it quantifies over are unambiguous. There are several ways to extend the principle to expressions with multiple meanings. For the sake of simplicity, I will set this issue aside.

² ‘Language’ can be understood broadly, encompassing systems of symbols that permit the construction of complex symbols. Talk of compositionality in the realm of thoughts presupposes that thoughts make up a language of some sort.

³ Some philosophers regard compositionality as analytic. While this strikes me as a rather implausible claim, for the sake of this paper I propose to set it aside. What matters is not whether compositionality is analytic, but whether it is self-evident. It clearly is not. Someone who assented to ‘Some bachelors are married’ is arguably confused about the meaning of ‘bachelor’ or the meaning of ‘married’, someone who assented to ‘Some natural languages are not compositional’ is not plausibly accused of linguistic incompetence.

philosophy, linguistics, and psychology tend to differ from one other and they are all significantly stronger than (C_0). Since our grounds for accepting (C_0) are already shaky, the stronger principles have at best the standing of methodological assumptions. In other words, despite widespread enthusiasm, the case for compositionality is surprisingly weak.

This is not to say that we should give up on compositionality. It is an interesting hypothesis and it has led to some important discoveries in semantics. My recommendation is only that we should acknowledge that it is on a par with other bold hypotheses, like that all syntactic operations are binary, that lexical categories are universal, that logical form and syntactic structure are intimately related, and so on.

1. Ambiguities

One complaint often voiced against (C_0) is that short of an explicit theory of meaning and a detailed set of constraints on modes of composition it is hopelessly vague. As Barbara Partee puts it:

...if the syntax is sufficiently unconstrained and the meanings sufficiently rich, there seems no doubt that natural languages can be described compositionally. Challenges to the principle generally involve either explicit or implicit arguments to the effect that it conflicts with other well-motivated constraints on syntax and/or on the mapping from syntax to meaning.⁴

Because of this, (C_0) is typically considered but a rough statement of compositionality. Textbooks tend not to dwell on it – rather they proceed quickly with more specific principles which incorporate a number of assumptions about meaning and grammar.

Partee's point can be illustrated with a classic example regarding quantifier scope. Sentences involving multiple quantifiers are often intuitively ambiguous: 'An ambassador was sent to every country' could mean that an ambassador is such that she

⁴ Partee (1984): 153. Compositionality is a demonstrably empty principle if one completely neglects syntactic and lexical constraints. Janssen (1983) has a proof that we can turn an arbitrary meaning assignment on a recursively enumerable set of expressions into a compositional one, as long as we are allowed to replace the syntactic operations with different ones. Zadrozny (1994) has shown that this can also be done by replacing the old meanings with new ones from which they are uniformly recoverable. For critical discussion of these triviality results, see Westerståhl (1998) and Dever (1999).

was sent to every country or that every country is such that an ambassador was sent to it. At the same, the sentence appears to have a single grammatical structure: its subject is ‘an ambassador’, its predicate ‘was sent to every country’, the object within the predicate ‘every country’. There is no independent evidence for significant lexical ambiguity either – whatever dictionaries tell us about the words within this sentence seems irrelevant to the ambiguity under discussion. The sentence presents us with a clear *prima facie* counterexample to (C₀): the meaning of ‘An ambassador was sent to every country’ does not seem to be a function of merely the meanings of its constituents and the way they are combined. Something else makes a difference.

Most of us these days are ready to give up on at least one of these appearances. Some deny that the sentence is ambiguous: they think it has a unitary meaning underspecified with respect to scope relations.⁵ Some deny that it has a unique grammatical structure: they think that the reading where the object takes scope over the subject is generated through a movement that remains invisible on the surface.⁶ And some deny that the sentence is free of lexical ambiguities: they account for the scope possibilities by assigning multiple (but systematically related) meanings of different types to both the subject and the object.⁷ If we are willing to bracket some pre-theoretical intuitions, (C₀) can be smoothly accommodated.⁸

On the other hand, if one has set views on meaning and grammar one might find the challenge scope raises for compositionality insurmountable. The early Russell was, for example, strongly attracted to the view that the meaning of a sentence is literally built up from the meanings of the words within the sentence, and that the way simpler meanings compose into more complex ones tracks the grammatical structure of the sentence. In the face of scope ambiguities he gave up on (C₀), and opted instead for the substantially weaker (C₀’):

⁵ Cf. Kempson and Cormack (1981) and Bach (1982).

⁶ Cf. May (1977) and Heim & Kratzer (1998).

⁷ Cf. Cooper (1975) and Hendriks (1993).

⁸ Some appeals to type-shifting rules are not comfortably described as postulations of lexical or grammatical ambiguities. Still, if there are multiple ways of deriving the surface structure of a sentence due to different applications of type shifting rules and if these different applications lead to different meanings, then it remains natural to say that there is more than one way to build up the meaning of this sentence. So, (C₀) is accommodated.

(C₀') If a complex expression has only meaningful constituents then its meaning is a function of the meanings of its constituents and of the way they are combined.

Russell famously denied that quantifying phrases have meaning, so for him (C₀') came out as vacuously true for sentences involving such expressions.⁹ One may reasonably wonder how sentences containing meaningless expressions end up meaning something. Russell's response to this challenge was the obvious one: not all meaninglessness is the same. A quantifying phrase, like 'an ambassador' is associated with a semantic rule that specifies what larger expressions mean that contain it as a constituent. When we have more than one quantifying phrase, there is a question as to what order the rules are to be applied, and depending on this order we may end up with different meanings. This is how scope ambiguities arise.¹⁰

Russell abandoned compositionality only because he was wedded to his assumptions about meaning and grammar; had he been more flexible he could have preserved compositionality by following one of the paths taken by present day semanticists. But the fact that (C₀) has bite only in conjunction with further commitments does not mean that the principle is in need of further clarification. Consider a claim that is formally analogous to (C₀):

(P₀) The physical properties of an ordinary object are a function of the physical properties of its parts and of the way they are combined.

One may reasonably complain that the notion of a physical property is vague and that there are different views about how parts of ordinary objects are combined. Still, it would be a mistake to declare (P₀) obscure on this basis. We may not know exactly which properties are physical or what laws underlie material composition but we still have a grip on these notions. (We certainly know that being ectoplasm is not a physical property and that elementary particles don't compose larger objects by falling in love with one another.) Similarly, despite our disagreements about meaning and structure we know well

⁹ If we replace the quantifying phrases in 'An ambassador was sent to every country' with logically proper names (expressions that are guaranteed to have meaning) we get sentences for which (C') is no longer vacuous. According to Russell, such sentences don't give rise to scope ambiguity.

¹⁰ Cf. Russell (1905).

enough what these are. (We certainly know that the ‘cat’ and ‘not’ have different kinds of meanings and that ‘John kissed’ is not a constituent of ‘John kissed Mary’.) In fact, we know enough to see that scope ambiguities are a genuine (albeit not insurmountable) challenge to (C₀). Whether we opt for scope-neutral meanings, quantifier-movement, or type-shifting rules to respond to it, we have to acknowledge that such maneuvers go against the appearances, and have to justify them accordingly.

So, I don’t think the open-endedness of the notions of meaning and structure should count as a legitimate complaint against (C₀). Still, there are other reasons for maintaining that the principle is unclear. (C₀) contains three crucial ambiguities that should be brought to light. The source of the first is the word ‘function’, the source of the second is the phrase ‘the meanings of its constituents’, and the source of the third is the pronoun ‘they’. They are all ordinary lexical or structural ambiguities, not the contrivances of philosophers. I will discuss them in order.

1.1. ‘... is a function of...’

Among the various meanings of the word ‘function’ dictionaries tend to distinguish between one that indicates a dependency relation and another – the mathematical sense – that does not. The construction ‘...is a function of ...’ is most naturally taken in the former sense but can, perhaps, be understood in the latter as well. When we say that height is a function of age (among other things), we tend to read this as something stronger than the bland claim that there is a function from age (and other things) to height. So, we need to distinguish between the following two principles:

The meaning of a complex expression is *determined by* of the meanings of its constituents and of the way they are combined.

There is a function to the meaning of complex expressions from the meanings of their constituents and of the way they are combined.

The first of these entails the second, but not the other way around. There is a long tradition of reading (C₀) in the weaker fashion. But it is unclear whether the official reading always conforms to our intuitions about what compositionality should demand.

One way to see this is to consider the possibility of there being two languages that are almost identical: they have the same expressions, the same syntax, and the same meanings for all expressions except for an arbitrarily sentence *S*. In other words, *S* has different meanings in these languages, even though it has the same structure and all its constituents mean the same. The weaker reading is not in conflict with this possibility: there may be different functions in the different languages that map the meanings of *S*'s constituents and the way they combined to *S*'s meaning. The stronger reading, however, rules this out. The meaning of *S* is not determined bottom up, for it *can* mean different things even if we hold the meanings of all its constituents and the way they are combined fixed.¹¹

Another way to see the contrast between these two readings of (C₀) is to consider linguistic change. Suppose we have a language that undergoes some very limited meaning change. The change is manifested in one obvious way: sentence *S* that once meant one thing now means something completely different. At the same time, suppose we have no evidence of any change in grammar. Under these conditions, I submit, we would assume that some constituent or other within *S* also changed its meaning. Why? One might say that because we tend to think that meaning is compositional, and it remains compositional as meaning changes. This is exactly what the stronger reading would allow us to say. But if natural languages are compositional only in the weaker sense the explanation is illegitimate: *S* could change its meaning in accordance with compositionality even its constituents and the way they are combined remains the same. This, I think, is a fairly good indication that the weak reading – despite being the standard one employed in semantics textbooks – is unreasonably weak.

¹¹ In Szabó (2000b), I argued that stronger reading of (C₀) could be spelled out as the claim that there is a single function *across all possible human languages* to the meaning of complex expressions from the meanings of their constituents and of the way they are combined. Thus understood, the stronger reading entails that at most one of the two languages mentioned above is a possible human language. One of them is not learnable in the ordinary fashion, as a first language. It would have to be acquired by acquiring the other language first and then learning a special translation rule for *S*.

1.2. ‘...the meanings of its constituents...’

Plural definite descriptions exhibit a collective/distributive ambiguity and ‘the meanings of its constituents’ within (C_0) is no exception. Consider the sentence ‘The wealth of a country is a function of the wealth of its citizens.’ In its distributive reading this sentence makes a rather controversial claim, namely, that the individual wealth of citizens fixes the wealth of nations. The collective reading permits the wealth of nations to be a function of what its citizens own collectively (including what they own as singleton groups, i.e. what they own individually). If there are roads or parks jointly owned by entire communities they are allowed to influence the wealth of nations on the second, but not on the first reading. We have thus the following two principles:

The meaning of a complex expression is a function of the meanings its constituents have *individually* and of the way they are combined.

The meaning of a complex expression is a function of the meanings its constituents have *collectively* and of the way they are combined.

Opting for the weaker principle allows the possibility of non-structurally encoded semantic relations in a compositional language. Consider, for example, the classic contrast between ‘Cicero is Cicero’ and ‘Cicero is Tully’. The semantic literature has been torn by such examples: on the one hand it is very plausible to think that these two sentences are not synonyms (for the former seems analytic but the latter does not), and on the other hand it is also very plausible that the meaning of proper names is nothing but their referent (for the point of having proper names in a language does not seem to go beyond the labeling of objects). The stronger reading forces us to give up one of these views: otherwise we would have a difference in meaning between the sentences without a difference in the individual meanings of constituents or a difference in the way they are combined. But if we think that the meanings the constituents have collectively in these sentences depend not only on what ‘Cicero’ and ‘Tully’ mean but also on the presence or

absence of *semantically encoded coreference* between the names, we can explain how ‘Cicero is Cicero’ can mean something other than ‘Cicero is Tully’.¹²

Another way to think of this proposal is to bring in the notion of indices. Sameness of indices encodes coreference. The difference between ‘Cicero is Cicero’ and ‘Cicero is Tully’ is that of indexing: the former is represented as ‘Cicero₁ is Cicero₁’, the latter as ‘Cicero₁ is Tully₂’.¹³ If the meanings of these sentences depend on the individual meanings of constituents and the way they are combined, *as well as on indexing*, the difference in meaning between these sentences can be accounted for. If we think of the indexing as a feature of the collective meaning of the constituents the account is compatible with compositionality, understood in the weaker fashion.

1.3. ‘...they...’

The pronoun ‘they’ in (C₀) can have two different antecedents: ‘its constituents’ or ‘the meanings of its constituents.’ The two readings can be paraphrased as follows:

The meaning of a complex expression is a function of the meanings of its constituents and of the way *those constituents* are combined.

The meaning of a complex expression is a function of the meanings of its constituents and of the way *those meanings* are combined.

The difference is subtle but significant. It is syntactic structure that encodes how the constituents of a complex expression are combined, so according to the first reading, complex meaning is a function of constituent meanings and syntax. The second, on the other hand, permits the existence of non-synonymous complex expressions with identical syntactic structure and pairwise synonymous constituents, as long as we have different semantic rules associated with the same syntactic rules.

Here is an example to illustrate the difference between the two readings. Assume that meaning is truth-conditional content, that ‘every man’ has a single syntactic

¹² For details, see Fine (2006). He argues that natural languages are compositional only in the collective reading of the principle.

¹³ This sort of proposal can be found in Fiengo and May (1994).

structure, that ‘every’ and ‘man’ each have a single determinate truth-conditional content, and that the truth-conditional content of ‘every man’ includes some contextual restriction on the domain of quantification. Given the first reading, we have a violation of compositionality. But the second yields no such verdict: if the syntactic rule which combines the quantifier and the noun is associated with a context-sensitive semantic rule, we can still maintain that in every context the meaning of ‘every man’ is a function of the meanings of its constituents and the way those meanings are combined in that context.¹⁴

The first reading of (C₀) is rather restrictive. Compare, for example the expressions ‘red apple’ and ‘pink grapefruit’. As competent speakers know, a red apple is red on the outside, while a pink grapefruit is pink on the inside. Here we seem to have complex expressions whose meanings are combined in different ways from the meanings of their lexical constituents, even though the syntactic mode of composition is the very same. Defenders of the first reading of (C₀) would have to argue for a hidden constituent in the structure of these expressions, and providing independent motivation for such a hidden constituent might be a tall order. Alternatively, they might try to deny that our knowledge of which part of a red apple is red and which part of a pink grapefruit is pink is not part of our linguistic competence.¹⁵ This might be more promising, but when carried to the extreme in responding to great many similar examples, it eventually leads to an uncomfortably narrow conception of meaning.

By contrast, the second reading of (C₀) is rather permissive. Instead of a semantic rule that looks at the context in determining the meaning of a quantifier phrase in order to associate with it a domain, we could have one that says: “Flip a coin – if it lands on tails then the domain of the quantifier phrase is the set of blue things; otherwise it is the set of red things.” If languages that allow such rules count as compositional one surely loses sight of all intuitive motivation behind adopting the principle. I am not suggesting that those who announce the second reading as their way of understanding compositionality would want to allow such rules: they all have explicit or tacit restrictions on what counts

¹⁴ See Stanley and Szabó (2000) and Pelletier (2003). The former subscribe to the stronger reading, the latter to the weaker one.

¹⁵ If I don’t know which part of a red apple must be red I am probably unable to tell which apples are red. But – barring verificationism – there is no quick argument from this premise to the conclusion that I do not know what ‘red apple’ means.

as a legitimate way of combining meanings. It is those restrictions that carry the theoretical weight.¹⁶

1.4. Fixing meaning

The three ambiguities in (C₀) give us eight possible readings. What do these all have in common, other than the fact that they are all specifications of the common formulation? Consider one of the eight readings:

- (C) The meaning of a complex expression is determined by the meanings its constituents have individually and of the way those constituents are combined.

What this says is that once you fix the meanings of the constituents of an expression and its syntactic structure, you have fixed what the expression means. If we assume that the *constituent-of* relation is irreflexive, asymmetric and transitive, and that every expression has but finitely many constituents the claim can be further strengthened: once you fix the meanings of lexical items within an expression and its syntax, you have no more leeway as to what the expression means. We all know that meaning of complex expressions can depend on the lexicon and on syntax – what (C) says is that it does not depend on anything else.

It is important that (C) does not rule out that the meaning of phrases, clauses, sentences, paragraphs, entire novels, etc. should depend on all sorts of odd things. All it says is that *if* the meaning of a complex expression depends on, say, the price of copper on the stock market, then so does *either* the meaning of some lexical item within that expression *or* the syntactic structure of that expression. (C) poses absolutely no restriction on what lexical meanings could be.¹⁷

All other readings of (C₀) besides (C) permit non-lexical non-syntactic sources for multiplicity of meaning in complex expressions. If we construe ‘is a function of’ as

¹⁶ Compare: ‘The wealth of a country is a function of the wealth of its citizens and the way the wealth of the citizens is combined.’ If this is to be a substantive claim there should be constraints on what counts as a way of combining individual wealth.

¹⁷ This is in conflict with the main thrust of a number of papers in Fodor and Lepore (2002). For more a more extensive criticism of the idea that compositionality is a substantive constraint on lexical meaning see Szabó (2004).

merely requiring the existence of a function we allow differences in meaning in complex expressions across languages or stages of developments of languages. If we construe ‘the meanings of its constituents’ collectively we allow differences in meaning in complex expressions due to semantic relations among constituents. If we construe ‘they’ as anaphoric to ‘the meanings of its constituents’ we allow differences in meaning in complex expressions different due to different semantic operations associated with the same syntactic operations.

(C) is both the strongest and the most natural reading of (C₀). It deserves to be called the principle of compositionality. If it turns out to be too string, we have the other versions to fall back upon.

2. Arguments

There are three more or less traditional considerations for compositionality. The first two are rarely stated explicitly – to some extent I am reading between the lines when I call them traditional. They are also obviously inconclusive. The third – the argument from understanding – is frequently evoked and is widely accepted to be sound. This argument comes in two flavors, one emphasizing the productivity, the other the systematicity of understanding. Each argument has significant weaknesses.

2.1. Intrinsicness

Consider a golden ring and its particular mass. Could an object have a different mass if it is made up of the same sort of particles arranged in the same manner as this ring? There is a strong intuition that it could not. Could an object made up of the same particles in the same manner as this ring have a different price? Clearly yes – if such a ring had once belonged to Nefertiti it would surely be worth a fortune. Whence the difference between mass and price? Taking the same sort of particles as the ones that a particular ring comprises and arranging them in the manner they are arranged in that ring amounts to a

duplication of the ring. Properties that duplicates share are intrinsic, and mass, unlike price, strikes us as an intrinsic property.¹⁸

Perhaps one of the reasons we are drawn to compositionality is that we are prone to view meaning as somehow intrinsic to its bearer. Why? The idea is that if e and e' are duplicate expressions then they must be tokens of the same expression type, and (setting aside indexicals) tokens of the same expression type are synonyms (if they have meaning at all).¹⁹ So, meaningful non-indexical duplicate expressions must share their meanings.²⁰ The intuition behind the intrinsicness of meaning can be brought out by considering pictures. A perfect copy of a picture shares its representational properties with the original: if the Mona Lisa depicts a woman with a mysterious smile, so do all its countless reproductions. Of course, the original relates to its object in a way the copies do not, but this is supposed to be a non-representational difference. To the extent that meaning is anything like pictorial representation, the analogy carries some weight. Let's call this the *intrinsicness argument*.

The argument is, of course, deeply problematic. We know that the sort of intuition that underlies it is sometimes deceptive – physics has taught us that we must distinguish between inertial and gravitational mass, and it may well turn out that neither is genuinely intrinsic. More importantly, we know that meaning isn't intrinsic *tout court* – clearly, some meaning is conventional and conventions are external to bearers of meaning. These are important caveats, but I don't think they fully undermine the underlying intuitions. Perhaps mass is an extrinsic property of elementary particles (e.g. because mass depends on frame of reference); still, it remains intuitively plausible (and physically uncontested) that duplicates of objects whose elementary particles have the same mass themselves have the same mass. That is, mass remains intrinsic *modulo* elementary mass. Similarly, although meaning is an extrinsic feature of lexical items (e.g. because meaning depends

¹⁸ The precise distinction between intrinsic and extrinsic properties is an elusive matter. For a good survey of the problems and the headway already made, see Weatherson (2006).

¹⁹ The claim that linguistic expressions are individuated semantically is fairly widely held among philosophers. This entails that linguistic expressions cannot change their meanings. What we ordinarily call meaning change is a (perhaps gradual) replacement of one expression with another that sounds just the same. Those who think that the proposition that 'Snow is white' means that snow is white is a necessary truth are all committed to this.

²⁰ If by meaning we mean something like Kaplan's notion of character then indexicals of the same type turn out to be synonyms across contexts of use. If, on the other hand, by meaning we mean something like Kaplan's notion of content, this does not hold.

on social conventions); still, it remains intuitively plausible (and undefeated by mere platitudes) that duplicates of sentences whose lexical components have the same meaning themselves have the same meaning. That is, for all we know meaning is intrinsic *modulo* lexical meaning. Still, this is no solid ground for accepting any form of compositionality.

2.2. Induction

Convincing counterexamples to compositionality are hard to come by.²¹ In linguistic semantics, compositionality has been a central assumption for at least half a century. Accordingly, a great many putative counterexamples had been offered: adjectives, negative polarity items, propositional attitude contexts, cross-sentential anaphora, conditionals, etc. Some of these are more convincing than others, but none has convinced many. For each of these phenomena, several compositional accounts have been proposed. As things stand, linguists tend to stick with a hypothesis that has worked. Let's call this the *inductive argument*.

The strength of the inductive argument should not be overestimated. For one thing we don't really have satisfactory semantic theories for anything more than small fragments of ordinary languages. The fixes semanticists come up with when faced with putative counterexamples to compositionality are often complicated and lack independent motivation.

What would be a clear violation of compositionality for, say, English? Imagine if the sentence 'Grass is green' were to mean that grass is green during the week, and that grass is *not* green on weekends. Then surely the meaning of this sentence would depend on something other than the meanings of its constituent words and the way those words are combined – what day of the week it is would make a difference. Except, of course, if the change in meaning is coupled with a corresponding change in the lexicon or in syntax. If on weekends 'green' means what 'not green' does on weekdays or an

²¹ Actually, quoted expressions are plausible counterexamples to compositionality. The English words 'woodchuck' and 'groundhog' are synonyms but the English phrases 'the word "woodchuck"' and 'the word "groundhog"' are not. It seems entirely *ad hoc* to blame the meaning differences on different modes of composition. However, this example can be dismissed by claiming that quotation devices that can be iterated are the invention of logicians, not a genuine part of any natural language.

unpronounced negation pops into the structure of ‘Grass is green’ we have no violation of compositionality. To make sure that any given putative counterexample to compositionality is genuine, one must make sure that one has a pair of complexes with different meanings that are built up from synonymous simples in a parallel fashion. Given the rudimentary state of our knowledge of meaning and structure, this is a tall order.

2.3. Understanding

Here is the argument most of us would fall back on in defense of compositionality: the meanings of complexes must be determined by the meanings of their constituents and the way they are combined, since we in fact understand them by understanding their parts and their structure. Call this the *argument from understanding*.

There are two problems with this, as it stands. First, the argument presupposes something rather controversial, namely, that understanding is a matter of grasping the meaning of that which is understood. This is certainly not right in general: to understand a problem, an idea, or a proof is not the same as grasping the meaning of the problem, the meaning of the idea, or the meaning of the proof (whatever *those* might be). Of course, one might try to stipulate that the required equivalence holds in the case of language – that to understand a *linguistic expression* is nothing more or less than grasping what it means. But such a stipulation is by no means innocent. There are many conceptions of meaning for which it is a rather bad fit. It is certainly false, in general, that understanding a linguistic expression is a matter of grasping its extension – so there is no argument from understanding in favor of the claim that extensional semantic theories should be compositional. But richer notions of meaning employed in semantics are also problematic in this context. Take for example the standard view according to which the meaning of a declarative sentence is the set of possible worlds where the sentence is true. Should proponents of this standard view be taken to embrace the idea that understanding a declarative sentence requires that we grasp a set of possible worlds? What sort of grasping would that be? Does it require the ability to tell of any particular possible world whether it is a member of the set? Or take the view according to which the meaning of a proper name is its bearer. Should someone who thinks this also accept that understanding

a proper name is a matter of grasping its bearer? What does that amount to? Does it require that one be able to single out the name-bearer? Many semanticists recoil at these suggestions – they would prefer to stay neutral on what are arguably substantive psychological questions. But without taking a stand on the relationship between meaning and understanding, the argument from understanding gives no support to the principle of compositionality.

Let us then assume that understanding a linguistic expression is grasping of meaning. Even so, there remains the second problem with the argument from understanding. Its central premise – that we understand complex expressions by understanding their parts and their structure – is not obvious. *How* we understand phrases and clauses is not open to reflection – if it were, psycholinguistics would presumably be a far less perplexing field of inquiry. Convincing arguments for compositionality must focus on some feature of language understanding we are certain of. There are two such features that have been emphasized – productivity and systematicity.

2.3.1. Productivity

The argument from productivity goes as follows. It is a fact that competent speakers can understand complex expressions they never encountered before.²² There must thus be something competent speakers know (perhaps tacitly) on the basis of which they can determine what those complex expressions mean. What can this knowledge be? It seems that the only thing we can plausibly point at is knowledge of the structure of the complex expression and knowledge of the meanings of its lexical constituents. Having this knowledge must be sufficient to understanding a complex expression, which means that the structure of the expression, together with the meanings of its simple constituents must determine what the expression means.²³

²² This should be fairly uncontroversial – everyone can come up with sentences they have never heard before. Another way to see that this must be so is to call attention to the fact that we can understand infinitely many distinct complex expressions whose meanings all differ from one another. A plain example of this is the sequence of sentences ‘Ann’s father is bold’, ‘Ann’s father’s father is bold’, ‘Ann’s father’s father’s father is bold’ ... etc.

²³ “... the possibility of our understanding sentences which we have never heard before rests evidently on this, that we can construct the sense of a sentence out of parts that correspond to words.” (Frege 1914?: 79)

An important caveat must be added: the argument is unable to screen out isolated counterexamples. The counterexamples are usually labeled *idioms* and swept under the rug. There is nothing wrong with this, as long as we have an independent characterization of what it is to be an idiom. And – since idioms must presumably be acquired one-by-one – we cannot allow for more than finitely many of them in any given natural language.

A more important problem with the productivity argument is that there are infinitely many complex expressions *nobody* has ever heard. How does the alleged fact that syntax plus lexicon determines the meanings of complex expressions we do hear support the claim that syntax plus lexicon determines the meanings of all the other complex expressions as well? This is not just the general Humean worry about induction – we have good reasons to doubt that the complex expressions we hear form an unbiased sample of all complex expressions. For if there were complex expressions whose meaning we could not determine in the usual way we would presumably want to stay away from them.

Another worry concerns the claim that we *already* understand certain expressions we have never heard before. What is the evidence for this? The fact that when we hear them we understand them shows nothing more than that the information necessary to determine what they mean is available to us immediately *after* they have been uttered. If there are features of the context of utterance we can invariably rely on, those features may well play a role in interpreting novel complex expressions. And there may well be such features. For example, consider a case when a speaker utters a sentence that means one thing in the language of the conversation, and another thing in another language that the hearer also knows. The hearer will then rely on her knowledge of which language is being spoken in interpreting the sentence, even though this information isn't part of the lexicon or the syntax of that language.

One can also wonder whether we always rely on syntax in interpreting novel expressions. In the case of numerals one might reasonably doubt this. It is plausible that syntax writes numerals right-to-left – i.e. that the syntactic constituents of 'five hundred eighty two' include 'eighty two' but not 'fifty eight'. At the same time, by far the simplest algorithm for identifying the referent of numerals (something that may well be crucial for grasping the meaning of these expressions) reads the numerals left-to-right –

i.e. ‘five hundred eighty two’ refers to ten times what ‘fifty eight’ refers to plus what ‘two’ refers to, etc. If in interpreting numerals we do rely on this simple algorithm then we must discern in these expressions a structure distinct from their syntactic structure.²⁴

In sum, I think the argument from productivity supports at best the cautious claim that by and large the meanings of complex expressions are determined by the meanings of their simple constituents, by some familiar pattern into which we can arrange those meanings (which may or may not be the syntactic structure of the complex expression), and, perhaps, by familiar and predictable features of the context in which the complex expression is used. This is a far cry from (C).

2.3.2. Systematicity

The argument from systematicity states that anyone who understands a number of complex expressions e_1, \dots, e_n understands all other complex expressions that can be built up from the constituents of e_1, \dots, e_n using syntactic rules employed in building up their structures. Since this is so, there must be something competent speakers know (perhaps tacitly) on the basis of which they can determine what the complex expressions built through such recombination mean. What can that be? The only plausible assumption seems to be that it is the structure of the original complex expressions e_1, \dots, e_n and the meanings of their simple constituents. Having this knowledge must be sufficient to understanding the recombined complex expression, which means that its structure together with the meanings of its simple constituents must determine what it means.

Like the argument from productivity, this is an argument to the best explanation. It too needs a caveat to dismiss idioms and is also subject to the worry whether it is syntactic structure (as opposed to some other structure) that we rely on in interpreting complex expressions. In addition, it has problems of its own.

First, it is not entirely clear whether the phenomenon it seeks to explain is real. Although it is fairly obvious that anyone who understands ‘brown dog’ and ‘black cat’ also understands ‘brown cat’ and ‘black dog’, the intuition becomes rather weak once we start considering more complex cases. Is the fact that someone understands ‘red car’ and

²⁴ For further discussion of this example, see Szabó (2000a): 77 – 80

‘tall building’ enough to show that he must understand ‘red building’ and ‘tall car’? One might argue that to understand ‘red building’ one has to know which parts of a red building are supposed to be red, and to understand ‘tall car’ one would need to know how to compare cars in terms of height. Neither of these is something one must obviously know in order to understand ‘red car’ and ‘tall building’.

There is also the problem that the argument shows less than it claims to. If we run the argument for the pair of sentences ‘apples are red’ and ‘bananas are yellow’ we can conclude that the meanings of ‘apples’, ‘bananas’, ‘are red’ and ‘are yellow’ plus predication determine the meaning of ‘Bananas are red’. It does *not* follow that the meanings of ‘bananas’ and ‘are red’ plus predication do that.

It is also worth emphasizing that the systematicity argument seeks to prove not only compositionality but also its converse. For the best explanation for recombination is supposed to be that we can decompose complex expressions into their simple constituents and then use the material to compose new ones. If we know how to do this in general then, presumably, the meanings of complex expressions are not only *determined by* but also *determine* their structure and the meanings of their simple constituents.²⁵

3. Strengthenings

Debates in philosophy, linguistics and psychology focus on related but significantly stronger claims. Whatever interest there is in compositionality derives from the fact that it is the common core of the stronger claims. I will identify these claims and call them the philosopher’s principle, the linguists’ principle, and the psychologist’s principle. In so doing, I do not wish to suggest that they widely accepted within those respective fields.

3.1. The philosopher’s principle

Philosophers are not in the business of designing semantic theories for natural languages, although they are certainly interested in general questions about what shape such theories

²⁵ Cf. Fodor and Lepore (2002): 59, Pagin (2003): 292.

could take. Philosophers are interested in *what* expressions mean instrumentally (to clarify various philosophical problems), but they are first and foremost interested in *why* expressions mean what they do. To address this latter problem they need to figure out what is explanatorily prior, the meanings of simple expressions or the meanings of complex ones. A strengthening of (C) offers an answer to this:

- (Φ) Complex expressions have their meanings in virtue the meanings of their constituents have individually and in virtue of the way those constituents are combined.

It is clear that (Φ) entails (C) – if X holds in virtue of Y and Z then X is determined by Y and Z. The converse does not hold because explanatory priority is asymmetric, while determination need not be. (According to Newton’s second law of motion, for example, given the mass of an object, the applied force and the acceleration mutually determine each other.) When philosophers state compositionality as the principle according to which the meanings of complex expressions are *defined in terms of* or *derived from* the meanings of their constituents, or when they say that complex expressions *get their meaning* from the meanings of their constituents they most likely have (Φ) and not (C) in mind. They commit themselves to the idea that the wholes have meaning because their parts do, and not the other way around.

In section 60 of the *Foundations of Arithmetic* Frege declares “it is enough if the sentence as whole has meaning; thereby also its parts obtain their meanings”.²⁶ This has come to be called the *context principle*. The idea is presumably that we should think of the meanings of constituent expressions as abstractions from the entire sentence meaning, which in turn can be taken to mean that words and phrases mean what they do *because* what the sentences in which they occur as constituents mean. On pain of circularity in the explanatory account of meaning one cannot hold on to both (Φ) and the context thesis.

Many philosophers believe that the meaning of sentences is explanatorily prior to the meanings of all other expressions.²⁷ One might hold this because one thinks that

²⁶ Frege (1884), section 60. The translation in the standard English edition is misleading; the translation here is mine.

²⁷ See Davis (2003): 175, fn. 16 for a representative list of defenders of this view. Davis himself rejects sentential primacy.

linguistic expressions get their meanings from mental states (paradigmatically, beliefs) that they are used to express, or from speech acts (paradigmatically assertions) that they are used to perform, and because one thinks that those relevant mental states or speech-act must have propositional content. But there are all sorts of mental states and speech-acts that are *prima facie* non-propositional – to imagine a unicorn, to hope for rain, to refer to an object, or to alert someone are but a few obvious examples. Proponents of sentential primacy must argue either that these are not as central in grounding linguistic meaning as propositional beliefs and assertions, or that despite appearances to the contrary these too have propositional content.²⁸

3.2. The linguist's principle

Linguists, at least the ones I tend to meet, don't have strong views on what makes words and sentence meaningful. They tend to focus on the relationship between structure and meaning. Even in lexical semantics, most of the action is in identifying the structural features that are supposed to underlie lexical categorization, not in trying to explain the meaning-differences within categories. (C) does not make a particularly strong claim about how structure constrains meaning; the principle that holds real interest for linguists is (Λ):

(Λ) The meaning of a complex expression is determined by its immediate structure and the meanings of its immediate constituents.

(C) says that the meaning of a complex expression is determined by its *entire* structure and the meanings of *all* its constituents. In assigning meaning to a complex expression (C) allows us to look at the meanings of constituents deep down the tree representing its syntactic structure, while (Λ) permits looking down only one level.

It is important to note that it is this stronger principle that gets often formalized and stated in textbooks as “the principle of compositionality.” When we demand of compositional meaning assignments that they establish a homomorphism between the

²⁸ One might even challenge the assumption that belief and assertion always have propositional content. For the former, see Szabó (2003), for the latter Buchanan and Ostertag (2005).

syntactic and semantic algebras, we demand that (Λ) should hold for the language in question.²⁹ (Λ) pushes theorists towards rich conceptions of meaning. If we are not allowed to look deep inside complex expressions to determine what they mean we better make sure that whatever semantic information is carried by an expression is projected to larger expressions in which they occur as constituents. There is nothing in the traditional arguments in favor of compositionality that yields support to (Λ) .³⁰

3.3. The psychologist's principle

From the psychological perspective what matters is how we actually understand complex expressions. (Ψ) formulates a rather straightforward hypothesis about this matter:

(Ψ) We understand a complex expression by understanding its structure and its constituents.

(C) guarantees that one could – cognitive limitations aside – determine the meanings of complex expressions on the basis of their structure and the meanings of their constituents. According to (Ψ) , this is in fact how we always proceed.³¹

The fact that we understand complex expressions we never heard before is a reason to think that *sometimes* we do understand complex expressions in this way. But there is no reason to assume that this is how we *always* proceed. In the case of complex expressions we hear all the time we may not have to go through all that. They may be understood as if they were idioms, even if they are not. (Ψ) seems to be an implausibly strong hypothesis that receives no support from the traditional arguments for compositionality.

²⁹ For detailed presentations of such formal statements, see Montague (1970), Janssen (1983) and Westerståhl (1998).

³⁰ An example of a semantics that violates this locality constraint is the treatment of propositional attitudes in Carnap (1947).

³¹ Horwich (1998): 155 thinks that something like (Ψ) is true by definition. This strikes me as a rather unfortunate stipulation in that it seems neither to accord with the ordinary meaning of ‘understanding’, nor to create a theoretically useful novel term.

4. Summary

I have argued for three main claims. The first one is that the usual statement of the compositionality principle is massively ambiguous. One of the eight available readings rules out all sources of multiplicity in meaning in complex expressions besides the lexicon and the syntax. Others are more permissive – how much more is not always clear. The second claim is that traditional considerations in favor of compositionality are less powerful than is often assumed: the intrinsicness argument and the inductive argument are inconclusive; the two arguments based on facts of language understanding have a number of shortcomings. In the end, compositionality is best construed as an empirical hypothesis on meanings expressed in natural languages. Finally, the third claim is that even if compositionality is true, most of the debates in philosophy, linguistics, and psychology will remain open. These debates tend to be about theses that are significantly stronger than compositionality.

Bibliography

- Bach, K. (1982) ‘Semantic Non-Specificity and Mixed Quantifiers.’ *Linguistics and Philosophy* **4**: 593 – 605.
- Buchanan, R. and G. Ostretag (2005) ‘Has the Problem of Incompleteness Rested on a Mistake?’ *Mind* **114**: 889-913.
- Carnap, R. (1947) *Meaning and Necessity*. Chicago: University of Chicago Press.
- Cooper, R. (1975) *Montague’s Semantic Theory and Transformational Syntax*. PhD Dissertation, University of Massachusetts, Amherst, MA.
- Davis, W. (2003) *Meaning, Expression, and Thought*. Cambridge: Cambridge University Press.
- Dever, J. (1999) ‘Compositionality as methodology.’ *Linguistics and Philosophy* **22**: 311 – 326.
- Fine, K. (2007) *Semantic Relationism*. Oxford: Blackwell.
- Fodor, J. and E. Lepore (2002) *The Compositionality Papers*. Oxford: Oxford University Press.
- Frege, G. (1884) *The Foundations of Arithmetic*. Evanston, IL: Northwestern University Press, 1980.
- (1914?) ‘Letter to Jourdain.’ In G. Gabriel et al. eds., *Philosophical and Mathematical Correspondence*. Chicago: Chicago University Press, 1980: 78–80.
- Heim, I. and A. Kratzer (1998) *Semantic Interpretation in Generative Grammar*. Malden, MA: Blackwell.
- Hendriks, H. (1993) *Studied Flexibility*. PhD. Dissertation, University of Amsterdam, Amsterdam.
- Horwich, P. (1998) *Meaning*. Oxford: Clarendon Press.
- Janssen, T. M. V. (1983) *Foundations and Applications of Montague Grammar*. Amsterdam: Mathematisch Centrum.
- Kempson, R. and A. Cormack (1981) ‘Ambiguity and Quantification.’ *Linguistics and Philosophy* **4**: 259 – 309.

- May, R. (1977) *The Grammar of Quantification*. PhD Dissertation, MIT, Cambridge, MA.
- Montague, R. (1970) ‘Universal Grammar.’ Reprinted in R. Thomason ed., *Formal Philosophy*. New Haven : Yale University Press, 1974: 222 - 246.
- Pagin, P. (2003) ‘Communication and strong compositionality.’ *Journal of Philosophical Logic* **32**: 287–322.
- Partee, B. (1984) ‘Compositionality.’ In F. Landman and F. Veltman eds., *Varieties of Formal Semantics*. Dordrecht: Foris: 281–312.
- Pelletier, F. J. (2003) ‘Context-dependence and compositionality.’ *Mind and Language* **18**: 148–161.
- Russell, B. (1905) ‘On Denoting.’ *Mind*, **14**. Reprinted in A. P. Martinich ed., *The Philosophy of Language*. Fourth edition. Oxford: Oxford University Press, 2001, pp. 212 – 220.
- Stanley, J. and Z. G. Szabó (2000) ‘On quantifier domain restriction.’ *Mind and Language* **15**: 219–261.
- Szabó, Z. G. (2000a) *Problems of Compositionality*. New York: Garland.
- (2000b) ‘Compositionality as supervenience.’ *Linguistics and Philosophy* **23**: 475–505.
- (2003) ‘Believing in Things.’ *Philosophy and Phenomenological Research* **66**: 584 – 611.
- (2004) ‘Review of J. Fodor and E. Lepore *The Compositionality Papers*.’ *Mind* **113**: 340–344.
- Weatherson, B. (2006) ‘Intrinsic and extrinsic properties.’ *Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.), URL = [<http://plato.stanford.edu/entries/intrinsic-extrinsic/>](http://plato.stanford.edu/entries/intrinsic-extrinsic/).
- Westerståhl, D. (1998) ‘On mathematical proofs of the vacuity of compositionality.’ *Linguistics and Philosophy* **21**: 635 – 6
- Zadrozny, W. (1994) ‘From compositional to systematic semantics.’ *Linguistics and Philosophy* **17**: 329–342.