Zoltán Gendler Szabó
Yale University

Abstract: This paper is a defense of substantive explanations in semantics. I begin by offering a diagnosis of why the view that semantic theories are merely descriptive has been widely accepted in philosophy and I suggest that these grounds are not compelling. Then I argue that semantic explanations don't have a uniform direction – upwards or downwards the syntactic tree. There is an explanatory division within the lexicon: the meanings of content words (nouns, verbs, adjectives, and adverbs) are semantically fundamental, while the meanings of function words (auxiliaries, connectives, copulas, derivational morphemes, determiners, expletives, prepositions, etc.) are derivative.

1. Descriptive and foundational semantics

Why do linguistic expressions mean what they do? Why, for example, does the Polish sentence 'Śnieg jest biały' mean that snow is white, rather than, say, that grass is green? In fact, why does it mean anything at all, why isn't it just empty noise? One might reasonably expect that a theory of meaning would have to address such basic explanatory questions. Yet, semanticists rarely claim that they have or even seek answers to them.

The standard view in philosophy is that this is as it should be. What a working semanticist does is try to build a theory that tells us, among other things, that the Polish sentence 'Śnieg jest biały' means that snow is white; what a speculative semanticist adds to the theory are ideas about what features of the mental life and social behavior of Polish speakers might make this semantic theory correct. The working semanticist is in the business of building what various authors have dubbed a *descriptive* semantics; the speculative semanticist is in the business of thinking about *foundational* questions of meaning. As David Lewis said, "only confusion comes of mixing these two topics."[1]

---

[1] Lewis (1970): 19. See also Stalnaker (1997): 535 and Kaplan (1989): 573-4.

Here is an analogy that brings out the purported contrast:[2]

> Imagine an anthropologist specializing in table manners sent out to observe a distant tribe. One task the anthropologist clearly might undertake is to simply describe the table manners of that tribe – to describe the different categories into which members of the tribe place actions at the table, and to say which sorts of actions fall into which categories. This would be analogous to the task of the philosopher of language interested in semantics; her job is say what different sorts of meanings expressions of a given language have, and which expressions have which meanings. But our anthropologist might also become interested in the nature of manners; he might wonder how, in general, one set of rules of table manners comes to be the system of etiquette governing a particular group. Since presumably the fact that a group obeys one system of etiquette rather than another is traceable to something about that group, the anthropologist might put his new question by asking, 'In virtue of what facts about a person or group does that person or group come to be governed by a particular system of etiquette, rather than another?' Our anthropologist would then have embarked upon the analogue of the construction of a foundational theory of meaning: he would then be interested, not in which etiquette-related properties particular action types have in a certain group, but rather the question of how action-types can, in any group, come to acquire properties of this sort. Our anthropologist might well be interested in both sorts of questions about table manners; but they are, pretty clearly, different questions. Just so, semantic theories and foundational theories of meaning are, pretty clearly, different sorts of theories.

This is a clean division of labor but it isn't one that should make working semanticists feel comfortable. Cataloguing table manners can hardly be considered as building a theory – even if it is systematic and comprehensive, a catalogue remains silent about the reasons behind the classification it employs.

Casting semantics as painstaking gathering of recalcitrant data for future theories in psychology or sociology fits poorly with the self-conception of most semanticists. It is true that linguists sometimes call their work descriptive, but they don't see it as non-explanatory. The intended contrast is with *prescriptive* theories, which most of us would happily relegate to the realms of cultural criticism and political punditry. But refraining from taking a stand on how people should use language is quite different from refraining to account for why they use it as they do. Eschewing explanation would make semantics embarrassingly trivial. Most of the work semanticists do is on well-documented languages, and while there are plenty of disagreements around the edges, the core judgments are hardly ever called into question. When it comes to linguistic meaning, the main difficulty is not in how to access the relevant facts but in how to make them intelligible by building theories on their basis.

---

[2] Speaks (2010).

It is one thing to distinguish *what* and the *why* questions in the theory of meaning, and quite another to insist that semantic theories address exclusively the first kind. We can all agree that there is a difference between describing and explaining the motion of the planets, but that does not mean that Kepler's Laws are purely descriptive and Newton's purely explanatory. Kepler predicted future positions of planets on the basis of earlier ones, and in so doing he provided genuine explanations. These explanations were not as deep as the ones subsequently given by Newton, but that doesn't mean that they did not go beyond cataloguing observational data. Kepler's kinematics and Newton's dynamics both describe *and* explain planetary motion – what makes the latter more fundamental than the former is the greater generality of laws it relies on.[3] I think it is natural to see the relationship between existing semantic theories and future psychological or sociological theories along similar lines.

The straightforward way to articulate the aim of semantic explanations is to tie them to the principle of compositionality, the thesis that the meaning of complex expressions is a function of the meanings of their constituents and the way those constituents are combined. The idea is that while semantic theories cannot explain *all* there is to linguistic meaning, they do *partially* explain the meaning of *some* linguistic expressions by deriving them from the meanings of others. Here is Robert Stalnaker's take on the task of semantics:[4]

> A descriptive semantic theory is a theory that says what the semantics for the language is without saying what it is about the practice of using that language that explains why that semantics is the right one. A descriptive-semantic theory assigns semantic values to the expressions of the language, and explains how the semantic values of the complex expressions are a function of the semantic values of their parts.

Like Lewis, Stalnaker thinks that semantic theories consist of systematic assignment of semantic values to expressions of a particular language. Like Lewis, Stalnaker insists that semantic theories won't tell us what it is about speakers of that language that fixes the right meaning-assignment. The one difference – and it is a significant one – is that unlike Lewis, Stalnaker uses

---

[3] Should we say that Kepler's explanations are non-causal? I don't think so. Prior location of a planet is as causally relevant to its current location as is its mass, and what counts as *the* cause of current location depends on the explanatory context. At most, we can say that kinematic explanations abstract from *more* casual factors than dynamic ones.
[4] Stalnaker (1997): 535.

the word "explain" to characterize what these theories do: they tell us *how* the semantic values of complex expressions are functionally determined by the semantic values of their parts, thus providing a limited explanation of *why* they mean what they do. This, I think, is a *prima facie* attractive picture of the explanatory ambitions of semantics.[5]

To see what a concrete example of the sort of compositional explanation Stalnaker has in mind might look like, here is a toy semantic theory about the English sentence 'Snow is white':

(1)   a.
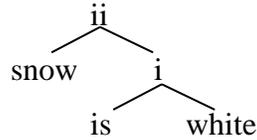
       b. $[\![\text{snow}]\!] = \lambda w_s. snow(w)$

                $[\![\text{white}]\!] = \lambda w_s \lambda x_e. white(x, w)$

                $[\![\text{is}]\!] = \lambda w_s \lambda P_{\langle s, \langle e, t \rangle \rangle}. P(w)$

       c.   IFA: $\lambda w_s. \varphi(w)[\lambda w_s. \psi(w)] = \lambda w_s. \varphi(w)(\psi(w))$

       d. $[\![\text{i}]\!] = \lambda w_s \lambda P_{\langle s, \langle e, t \rangle \rangle}. P(w)[\lambda w_s \lambda x_e. white(x, w)] =$

                $\lambda w_s (\lambda P_{\langle s, \langle e, t \rangle \rangle}. P(w))(\lambda x_e. white(x, w)) = \lambda w_s \lambda x_e. white(x, w)$

                $[\![\text{ii}]\!] = \lambda w_s \lambda x_e. white(x, w)[\, \lambda w_s. snow(w)] =$

                $\lambda w_s (\lambda x_e. white(x, w))(snow(w)) = \lambda w_s. white(snow(w), w)$

Under (1a) I offered a simplified syntactic structure of this sentence. (It is simplified because it involves no decisions about syntactic categorization, because it ignores tense and aspect, because it makes the controversial assumption that the subject and the predicate of this sentence have no hidden syntactic complexity.) Under (1b) I gave the semantic values of the lexical items that

---

[5] Lewis says that "[a] concise grammar for a big language – for instance, a finite grammar for an infinite language like ours – had better work on the compositionality principle." Lewis (1980): 25. But his commitment to compositionality was always instrumental: non-compositional theories may be messy and sometimes intractable but they would not be any more or less explanatory. Lewis defines a language as a set of ordered pairs of strings and meaning and he says that he knows of "no promising way to make objective sense of the assertion that a grammar Γ is used by a population *P* whereas another grammar Γ′, which generates the same language as Γ, is not." Lewis (1975): 176. His account of why linguistic expressions mean what they do bypasses compositional semantics altogether. For Lewis, the explanation of why any expression *e* of English has its meaning *m* is, roughly, that English speakers are party to a specific convention of truthfulness and trust which links them to an infinite set of pairs, which contains ⟨*e*, *m*⟩ as a member.

occur within 'Snow is white'. The indices on the first occurrences of variables indicate semantic type; the type system is the standard one with e, t, and s as basic types. Semantic values are intensions, i.e. functions from possible worlds to extensions. The intension of 'snow' is a function from possible worlds to some entity at that world that is snow. (This would perhaps be the mereological sum of all the snow in that world.) The intension of 'white' is a function from possible worlds to functions from entities to truth-values that assigns True to an entity at a world just in case the entity is white at that world. (In other words, a function that assigns to possible worlds the characteristic function of the set of white entities at that world.) Finally, the intension of 'is' is a function that maps possible worlds to a function from predicate-intensions to the corresponding predicate-extension in those worlds. The syntactic tree consists of two branchings, both interpreted as intensional function application spelled out under (1c). The rule says that the result of applying the intension φ to the intension ψ is an intension whose values at a world w can be calculated by applying $\varphi(w)$ to $\psi(w)$. (1d) specifies how intensional function applications yield intensions for the expressions occupying the higher nodes of the tree. The intension of the entire sentence is the function from possible worlds to truth-values that assigns True to those worlds where the entity that is snow in that world is white in that world. This represents the truth-conditions of the sentence.

How is this theory supposed to explain what 'Snow is white' means? Semantic theories, like scientific theories in general, abstract from certain aspects of their subject matter. Besides determining extension relative to possible worlds, meaning gives rise to various discourse effects, ground multiple implications, and triggers countless suggestions. Intensional semantics remains silent about all of these. But when it comes to the aspects of meaning the theory does aim to capture, the theory provides explanations tracing the syntactic tree. The theory leaves the intensions of the expressions at the leaves unexplained and accounts for the intensions of expressions at higher nodes in terms of them. Thus, $\lambda w_s \lambda x_e. white(x, w)$ is the intension of 'is white' *because* it is the intension of 'white' and *because* the intension of 'is' is such that when applied to any intension it yields the very same intension back, and $\lambda w_s. white(snow(w), w)$ is the intension of 'Snow is white' *because* it is the result we get when we apply the intension of 'is white' to the intension of 'snow'. Of course, the explanation isn't really a consequence of the

theory – there is no 'because' anywhere in the semantics – it is just a natural way to understand the significance of the derivation under (1d).

It's worth stressing that acceptance of compositionality, or any specific compositional semantic theory, does not dictate a unique explanatory pattern. What compositionality demands is that the meaning of any complex expression should be a *function* of the meanings of its constituents and the way those constituents are combined. This entails that complex expressions built up in the same way from synonymous parts have the same meaning. This suggests but does not necessitate a uniformly bottom-up explanatory order. Functions are cheap. We can safely assume that no two people with the same name have the same pattern of fingerprints but that does not mean that names can be explained by patterns of fingerprints. There is a function from patterns of fingerprints to the names of people with that pattern of fingerprints simply because sameness of fingerprints is exceedingly rare. Similarly, it may be that lexical synonymy is rare or even nonexistent, in which case compositionality would be a trivial matter giving us no reason to expect explanations of meaning to track syntactic construction. Those who deny that semantic theories explain anything don't necessarily reject those theories – rather, they tend to think the theories are approximately correct but the bottom-up explanations we are inclined to tag onto them are wholly mistaken.

This paper is a defense of two main claims: that, contra Lewis, semantic theories provide genuine, deep, albeit partial explanations of meaning and that, contra Stalnaker, these explanations don't always go from simpler to more complex expressions. I will start by discussing the reasons why skepticism about semantic explanations has been such a strong undercurrent in philosophy of language and by making a case that the reasons are not compelling. Then I will argue against the idea that semantic explanations have a uniform direction – upwards or downwards the syntactic tree. I claim that there is an explanatory division within the lexicon: the meanings of content words (nouns, verbs, adjectives, and adverbs) are semantically fundamental, while the meanings of function words (auxiliaries, connectives, copulas, derivational morphemes, determiners, expletives, prepositions, etc.) are to be explained in terms of certain expressions in which they occur as constituents. I accept the explanation of the meaning of 'Snow is white' outlined above (modulo empirical objections aimed at the details

of the syntax and semantics) but I see the explanation as incomplete. I think it can be supplanted with an explanation of why 'is' means just what it does in terms of meanings of expressions that are not constituents of this sentence.

2. Against deflationism and cognitivism

Why would philosophers of language be attracted to the view that the theories built by working semanticists are purely descriptive? Some might be skeptical about explanations in the special sciences or even about explanations in general. One might be tempted by an explanatory exclusion argument: semantic facts supervene on non-semantic facts, so for any putative semantic explanation there is an underlying non-semantic one, and unless these are actually the same explanation cashed out in different vocabulary, the latter trumps the former. Of course, if this argument succeeds it also shows that there are no explanations in economics, sociology, psychology, biology, or chemistry.[6] Or one might think that lack of agreement about what meaning really is must foil any attempt to understand its distribution among linguistic expressions. But if this were correct, we should be skeptical of the value of explanations in physics too: we should worry that lack of agreement about what it is to be a law of nature undercuts the value of appealing to them in explaining the phenomena.[7]

I won't address general reasons for doubts about explanation. Here I will simply assume a basic explanatory realism: there are good explanations in the sciences (as well as in ordinary life) and one of the things that make them good is that they track objective dependencies in the world. I also take it for granted that not all objective dependencies are causal – some are constitutive. For example certain properties of a molecule can be explained in terms of certain properties of the atoms constituting it and the way in which those atoms are combined. This is the sort of constitutive explanation compositional theories of meaning can also provide. There may be deep reasons to worry that the constitutive explanations we have can't really do what we think they

---

[6] Kim (1989) defends the thesis that no event can be given more than one complete and independent explanation.
[7] Of course, there are those who embrace this conclusion: "At the basis of the whole modern view of the world lies the illusion that the so-called laws of nature are the explanations of natural phenomena." Wittgenstein (1921): 6.371.

are supposed to but responding to such wide-ranging skepticism is not a task specifically for philosophy of language.

A quite different form of radical skepticism about semantic explanations is rooted in the thought that the sort of semantic theories that tend to be developed in linguistic departments nowadays – that is, broadly truth-conditional theories – miss their mark so badly that they cannot support any serious explanation of meaning. It is often remarked that truth-conditions are too coarse-grained: 'Cicero is Cicero' is trivial while 'Cicero is Tully' is not, which marks a fundamental difference in their meaning with no corresponding difference in truth-conditions. So meaning is in some sense richer than truth-conditions. Some doubt that we can extend our everyday talk of the truth of beliefs or statements to ascriptions of truth to sentences relative to contexts: the truth of what one says in uttering 'Every bottle is empty' depends on one's communicative intentions (because the domain of quantification does), and communicative intentions are not fixed by either the meaning of the expression uttered or the context in which the utterance takes place. So truth-conditions are in some sense richer than meaning. When you put the two criticisms together it's easy to feel that meanings and truth-conditions are entirely different things.

I won't address skepticism about truth-conditional semantics either, except to note that much of the criticism of the project seems to stems from the assumption that it seeks to tell us *what* meanings are. Maybe linguist meaning of a declarative sentence is in some regard more fine-grained and in another regard less fine-grained than its truth-conditions, but that in no way shows that such a meaning cannot be usefully modeled by truth-conditions. Models abstract and idealize their target phenomena and this fact in no way detracts from their explanatory value. Geometrical optics bypasses the question what light is, it models them as rays that propagate in straight lines losing a lot of complexity of the actual physical phenomenon. There are certain things geometrical optics simply cannot account for – such as diffraction and interference. At the same time, it does provide excellent explanations for other phenomena – such as reflection and refraction. Using the same semantic value to model the meanings of 'Cicero is Cicero' and 'Cicero is Tully' will make the semantics of verbs like 'believe' and 'hope' inadequate and using potentially different semantic values for different uses of 'Every bottle is empty' will make the semantics of 'follows' and 'equivalent' overly complicated. But that is not a good reason to think that truth-conditional theories of meaning cannot provide any good explanations.

8

Let us set aside general worries about explanation and about semantics aside, and turn to more specific reasons for doubting the cogency of semantic explanations. I know of two such reasons – one based on the idea that the sorts of explanations truth-conditional theories of meaning can provide are vacuous, the other on the idea that they explain something other than meaning. I will address them in that order.

Frege held that "corresponding to the whole-part relation of a thought and its parts we have, by and large, the same relation for the sentence and its parts."[8] In this regard, Russellian propositions are very much like Fregean thoughts: complex entities whose structure mirrors more or less to the structure of sentences expressing them.[9] The difference is only what the smallest parts are supposed to be – for Frege, they are modes of presentation of objects and functions; for Russell, they are objects, properties, and relations. While there is not much in ordinary thought that would encourage us to think that content is syntactically structured – it is very surprising to hear that Mount Blanc itself or some way of thinking of it is part of what we say in uttering the sentence 'Mount Blanc is covered with snow' – one can certainly construct semantic theories that assign these sorts of entities to linguistic expressions in a systematic fashion. Suppose one were to think, as for example Paul Horwich does, that this isn't just an interesting theoretical option but a matter of conceptual necessity: that the meaning of a complex expression *must* be something that shares its structure.[10] Then we should think that we can give an entirely trivial explanation of why any complex expression of any language means just what it does in terms of the meanings of its simple constituents and its syntactic structure: just arrange those simple meanings into that structure and you get the complex entity that *is* the meaning. Semantic explanations would thus be thoroughly *deflated*.

---

[8] Frege (1919): 255.
[9] It is not entirely clear why Frege said that the structure of thought is "more or less" the same as that of the sentence expressions it, but as far as Russell was concerned the hedge was extremely important. On Russell's view, syntactic form diverges systematically from propositional form, and his theory of description was an elaboration on this theme. As it turns out, Russellian truth-conditions can easily be assigned to sentences containing descriptions without postulating that they have a "hidden" or "deep" or "logical" structure. In fact, not one of the standard reasons for postulating divergence between syntactic and logical form stands up to scrutiny – semantic theories nowadays tend to run off directly of syntactic structure. See Sainsbury (1991): 291 – 2 for a list of alleged differences between the grammatical and logical forms of English sentences.
[10] Horwich (1998), (2005).

Let me illustrate how the deflation works on a particular example. The English sentence 'The boys lifted the piano' is ambiguous: it can mean either that each of the boys lifted the piano or that the boys together lifted the piano. One might think that an explanation of the ambiguity is called for, and traditional semantic theories go searching for its source. Perhaps the first reading arises due to an unpronounced universal quantifier, perhaps the second reading comes about due to an unpronounced collectivizing operator, or perhaps the definite article, the plural morpheme, or the verb itself have multiple meanings. The deflationist takes no stand on which of these alternatives is the correct one, but claims that once the dust settles and the linguists reach an agreement about the possible meanings of the parts and the possible ways they can be arranged to yield a sentence, there will be nothing more to say about the meaning of the whole. In other words, all the substantive explanatory work resides in providing a non-semantic explanation of why the lexical components of the sentence have their meaning (citing, perhaps, the conventions regulating their use) and a non-semantic explanation of the two distinct ways those lexical components can be combined (citing, presumably, syntactic principles and parameters).

I think this picture is fundamentally mistaken: compositionality is not a conceptual truth derivable from truisms about language understanding. The fact that a sentence has a certain syntactic structure and that its lexical components means what they do does not conceptually necessitate *anything* about what the sentence means. Here is my argument. Suppose English is in fact compositional and suppose we were to stipulate that whenever it rains at the location of utterance, the sentence 'Elephants are gray' shall have the meaning of 'Julius Caesar was murdered on the ides of March' while retaining its usual meaning on other days. Let us also stipulate that all other expressions, including the morphemes, the words and the phrases within either of these sentences retain their old meanings whether it rains or shines. Assuming our stipulation hasn't affected the syntax of the language, our new language, Rain-English, is no longer compositional: the meaning of 'Elephants are gray' is no longer a function of the meanings of its constituents and the way they are combined; it also depends on whether it is raining.[11] Those who think compositionality is a conceptual truth must say that there cannot be a

---

[11] Why not say that Rain-English contains two distinct sentences with the same phonological form as the English 'Elephants are gray' – one that means that Julius Caesar was murdered on the ides of March and another that means that elephants are gray? Because that goes against the stipulation I gave: I defined Rain-English as a language that has no more sentences with the phonological form of the English 'Elephants are gray' than English itself.

language like Rain-English – yet, its semantics is perfectly coherent. The language is also easily learnable – I assume you have already learned it, although I concede that if you are inattentive to the weather you may never become a fluent speaker.[12]

The example gives a taste of the power of arbitrary linguistic stipulation. When we say that given lexical meanings and syntax, 'Elephants are gray' *must* mean just what it does, we are ignoring Rain-English and its countless easily conjured cousins. I am not saying that ignoring these possible languages is a mistake, but I am saying that it is akin to our practice of ignoring far-fetched possible worlds – like saying that if it is 4am the sun *must* soon be rising, that if the sailor has scurvy he *must* have vitamin deficiency, or that if the demand for gold is peaking so *must* its price. As in the other cases, what makes ignoring the exceptions legitimate is that they are violations of a *ceteris paribus* law. Compositionality is a plausible example of a general law of human languages, not a conceptual truth.[13] Horwich's attempt to trivialize semantic explanations fails: such explanations are substantive exactly because they limn the contours of what it is to be a natural (as opposed to gerrymandered) language.[14]

The second reason for skepticism about semantic explanations goes as follows. Since Davidson's groundbreaking early papers, many philosophers and linguists have come to view the theory of meaning as the theory of the competent speaker's knowledge of meaning.[15] This is what is called the *cognitivist* perspective on semantics. What led Davidson to this view was the idea that a theory of meaning needs an explanatory target, and the one he found concerned how knowledge of meaning is possible: how is it that finite creatures like us come to know an infinite array of things, to wit, the meanings of infinitely many linguistic expressions? The answer Davidson

---

[12] This is a twist on an example I used in Szabó (2000a). There I devised a similar meaning-switch that preserves a function from meanings of parts and their mode of composition to the meaning of the whole to argue that the intuition behind compositionality demands more than mere functional dependence.

[13] One plausible way to think about semantic laws is to see them as laws of language acquisition; cf. Szabó (2000a). We could easily learn Rain-English, but not as a first language: normal human beings under normal human conditions would presumably have to acquire English first and learn Rain-English on its basis, not the other way around. But my argument does not rely one any specific conception of what may underwrite the laws of language.

[14] One might also protest the idea that semantic explanations would be trivial if they rested entirely on conceptual truths. On many people's views (not mine), mathematical truths are conceptual and nonetheless many people (including me) think mathematical proofs can be explanatory. But this point is incidental to the argument against Horwich: I claim that compositional semantic explanations are substantive in a way explanations in the natural sciences are.

[15] Davidson (1965), (1967).

11

proposed was, roughly, that speakers of a language have this infinite aptitude because they know a semantic theory that derives interpretative T-sentences for all the sentences of the language. According to the cognitivist conception, semantic theories make explicit a body of knowledge any competent speaker already possesses. If this is so then semantic theories can hardly explain why linguistic expressions mean what they do, since this is not the sort of thing that ordinary speakers tend to know. Linguistic competence may well require tacit knowledge of a theory, such a theory may well explain why various expressions mean what they do, but it does not follow that linguistic competence requires tacit knowledge of those explanations.

Consider the following analogy. Ten-year-old children tend to be competent in multiplying two digit numbers. It may well be that the best way to capture their competence is to attribute to them tacit knowledge of a theory about basic arithmetic. Such a theory provides a fully adequate explanation for why $34 \cdot 67 = 2{,}278$. And, yet, it seems rather implausible to suggest that the children know why $34 \cdot 67 = 2{,}278$. To know that they would have to have an unusually good grip of why the rules they follow are the right rules for multiplication. Knowledge of the explanation would require a sort of reflection that is hard to achieve without considering what the right rules of multiplication would look like in a notation that is not base 10.

I think the cognitivist perspective on semantic theories has a significant impact even on those who do not explicitly embrace it. Take David Kaplan's claim that the fact that 'owsnay' means snow in Pig-Latin belongs to semantics, but the reason why this is so does not.[16] I think there is a pretty strong argument that Kaplan is mistaken: there *is* a simple semantic explanation why 'owsnay' means snow in Pig-Latin. It goes as follows. First, there is a semantic fact about Pig-Latin: that its words inherit their meanings from the English words they are derived from. Second, there is a semantic fact about English: that 'snow' means snow. To round out the explanation, we add the fact that 'owsnay' derives from 'snow', which follows from the lexical fact that 'snow' is an English word and the syntactic fact that Pig Latin forms its words by taking

---

[16] Kaplan (1989): 573-4. I disagree with Kaplan about the proper spelling in Pig-Latin. He says 'ohsnay' – I say 'owsnay'. Like Kaplan, I won't introduce a special convention (e.g. capitalization) for linguistic expressions to refer to their own meanings and let context help when this is the intent.

the first consonant cluster of an English word, moving it to the end of the word, and suffixing it with an 'ay'.

Kaplan would presumably object that the correlation in meaning between words of Pig-Latin and words of English is not a semantic one. But why not? This is where I suspect the argument Kaplan relies on presupposes cognitivism. Imagine a child, Ohnnyjay, raised by vicious experimental linguists as a mono-lingual speaker of Pig-Latin. Since Ohnnyjay is a competent Pig-Latin speaker, he must tacitly know the correct semantic theory for Pig-Latin. If that theory could yield the above explanation then we would have to conclude that Ohnnyjay tacitly knows about the correlation of Pig-Latin and English. This sounds preposterous, so the explanation must be non-semantic.

I consider the claim that semantic theories must be bound to a single language a very unpleasant consequence of cognitivism. Human languages are not isolated from one another and cross-linguistic explanations are commonplace is phonology or syntax. Why would we adopt a conception of semantics that rules out such explanations from the get-go?

I do not repudiate the aims of cognitivist semantics – we do want an explanation of how people can make semantic judgments and it is a good hypothesis that a tacitly known theory underlies this ability. Similarly, we want an explanation of how people can make judgments about the path of moving objects, about the intentions of others, or about inferential relations. We might call these theories naïve physics, naïve psychology, and naïve logic, respectively. They should not be confused with physics, psychology, and logic. Analogously, I think we should call the hypothesized internal theory underlying semantic competence, *naïve semantics* and keep investigation into it distinct from semantics proper. It is, presumably, not part of the naïve semantics Ohnnyjay employs that 'owsnay' inherits its meaning from 'snow'. But this is no reason to think that it should not be part of the semantics of Pig-Latin.

Since our naïve semantic judgments are much more extensive and much more reliable than our naïve physical judgments, it is reasonable to think that naïve semantics is a *better* theory than naïve physics. But even if naïve semantics were an infallible guide to linguistic meaning (which

13

it surely isn't, since competent speakers are prone to make semantic mistakes), it wouldn't need to yield much insight into its subject-matter. Consider the following analogy. Suppose I have a device that reads barcodes and suppose – contrary to fact – that barcodes are equipped to encode any price whatsoever. (Of course, to encode astronomical prices barcodes would have to be astronomically long, but the software in the device could decode barcodes of any length.) My device operates by instantiating an algorithm which matches barcodes with prices. There is a theory underlying the algorithm knowledge of which would suffice to figure out the price of an arbitrary item as long as it is properly labeled. Nonetheless, we would not call this theory a *theory of pricing.* The problem isn't that the theory fails to tell us what prices really are. It is, rather, that while it tells us what things cost, it says nothing about why. It doesn't even entail simple things like if item x is a freely detachable part of item y then the price of x won't exceed the price of y.17 A resolutely descriptive theory of meaning, one that captures nothing more than what ordinary speakers know on account of their semantic competence is not that different from such a theory of pricing. The semantic theories philosophers and linguists are trying to build do better in explaining linguistic meaning than the barcode theory does in explaining market pricing. What we need is some idea how much better.

3. The direction of semantic explanation

Semantic explanations track dependences among semantic facts: they tell us that expressions mean what they do because certain other expressions mean what they do. According to explanatory realism, these dependences are objective and it is the job of semantic theories to discover them. The relevant dependencies, I argued, are neither metaphysically necessary, nor knowable *a priori.* They are not metaphysically necessary because, as the example of Rain English illustrates, stipulation can change meanings in a pointwise manner – ascribing a new meaning to a single expression while leaving the meanings of all others the same. The dependencies between semantic facts are not knowable *a priori* because, as the example of Pig

---

17 I know that one-way tickets sometimes cost more than two way tickets. To pull this trick, the airlines tend to make the parts of the two-way ticket non-detachable: the one-way part of the two-way ticket is not for sale.

Latin shows, they may depend on cross-linguistic connections that needn't be available to competent speakers.

A major part of any general inquiry into dependencies among semantic facts is to uncover their correlation with mereological structure. Does semantic explanation go always from parts to whole, always from whole to parts, or sometimes one way and sometimes the other? I call this the question of *the direction of semantic explanation*.

The part-whole structure of a linguistic expression is given by *syntax*: a complex expression has exactly as many parts as there are nodes in its phrase structure.[18] Although 'Snow is white' begins with 'snow is' syntax refuses to regard the latter as part of the former. The official reason is that 'snow is' fails standard constituency tests, such as coordination. Thus, while we have sentences like 'Snow is white and melts in the heat' where 'is white' and 'melts in the heat' are conjoined into a single expression, there is no sentence like 'Snow is and Jill's favorite color is white' indicating that 'Snow is' and 'Jill's favorite color is' cannot be so conjoined. The best explanation for this fact is supposed to be that 'snow is' and 'Jill's favorite color is' aren't linguistic expressions at all – they a mere sequences of words.

There are two crucial ways in which syntactic structure differs from the usual part-whole structures. The first is that syntax permits multiple *occurrences* of the same expression within another. Thus, the verb 'like' is part of the sentence 'People like those who like them' twice over. By distinguishing linguistic expressions and their occurrences we capture the important fact that 'People hate those who like them' and 'People like those who like them' have the same structure, and hence, that one could be obtained from the other via substitution. Analogously, the need to capture the relationship between chemical compounds led chemistry to view oxygen as having multiple occurrences within carbon-dioxide.

The second way in which syntax differs from standard mereologies is that it eschews *unrestricted composition*. There is no linguistic expression all of whose parts overlap either 'the'

---

[18] Alternatively, one might opt for a more fine-grained mereology driven by *morpho-syntax.* On this conception, prefixes, suffixes, and infixes would count as parts of expressions too.

or 'or' – to make a phrase with both as a part we need to add some more words. To be a linguistic expression is to exhibit a certain kind of unity. What makes a linguistic expression *one* are syntactic connections among its parts, just as chemical bonds are responsible for the unity of molecules. What makes it the case that certain parts can be syntactically unified is that they have the matching categories, just as positions within the periodic table account for the possibility of chemical bonds among elements.

Someone who believes that semantic explanations are uniformly top down is committed to the idea that we can explain why the words within 'snow is white' mean what they do by appealing to the meaning of this very sentence (and perhaps less efficiently, by appealing to the meaning of any other sentence that contains this sentence as part). By contrast, someone who takes semantic explanations to be uniformly bottom up thinks that we can use the facts that 'snow', 'is' and 'white' mean what they do to explain why 'snow is white' has its meaning (and also, why any other sentence built form exclusively from these words means what it does). In the next two sections, I will argue that neither of these views is correct: semantic explanations don't have a uniform direction. Nonetheless, I think bottom-up explanations are the rule. So, at the end of the paper I will try to give a principled account of when and why the exceptions arise.

4. Against uniformly top-down explanations

One philosopher who apparently subscribed to the idea that semantic explanations are uniformly top-down is Frege. One of the three methodological principles he lists in the introduction to the *Foundations of Arithmetic* is that "one should ask for the meaning of a word in the context of a sentence, not in isolation".[19] Frege thought adhering to this principle was crucial for a sound theory of numbers because asking what words like 'two' or 'plus' mean out of linguistic context pulls us towards the view that numbers are ideas and arithmetic is about mental operations. Yet, one might think that the siren song of psychologism could be resisted without plugging our ears to perfectly sensible questions about lexical meaning. And indeed, there is more behind Frege's

---

[19] Austin's standard translation uses 'proposition' where I use 'sentence' – but since in contemporary philosophical parlance proposition is not a linguistic entity, this is misleading.

insistence on putting sentences first: he believes that "it is enough if the sentence as whole has meaning; thereby also its parts obtain their meanings."[20] If words *obtain* their meanings for the sentences in which they occur then words mean what they do *because* sentences in which they occur mean what they do. This is sometimes called *context principle* – a commitment to the explanatory priority of sentences over words is semantics.

The context principle is a fairly radical claim, but it is certainly not absurd. It entails that semantic theories must yield infinitely many explanations for why 'snow' means what it does, one for each sentence in which the word occurs. But it does not follow that these explanations are equally good. Presumably, we would want to say that explain the meaning of 'snow' by appeal to the meaning of 'snow is white' is more perspicuous than explaining it by appeal to, say, 'snow is white and snow is white.' The context principle is even compatible with compositionality – if we explain the meanings of words in terms of the meanings of sentences in which they occur, then the latter must functionally determine the former – but that does not mean that the former cannot also functionally determine the latter. Functional determination can be symmetric, even if explanation is not.

I will offer three arguments against the context principle. The first goes after its motivation. Why think sentences are explanatorily prior to their constituent words? Frege thinks most expressions refer to functions – entities that are in some sense "unsaturated" or "incomplete". These expressions are in need of completion and this fact may suggest that they are not independently meaningful. Sentences are different – they refer to objects, to wit, truth-values, which makes them semantically self-standing. Alas, for Frege proper names also refer to objects, so this sort of consideration can hardly support the claim that the meanings of sentences are explanatorily prior to the meanings of the words they contain.

Followers of Frege have tried to replace the metaphysical distinction that can't do the job with a psychological one that might. William Alston says that "a sentence is the smallest linguistic unit that can be used to perform a complete action that is distinctively linguistic."[21] What makes

---

[20] Frege (1884), §60. Again, I replaced Austin's 'proposition' with 'sentence'.
[21] Alston (1964): 33.

sentences special on this view is that they are the smallest expressions that can be used to assert, query, command, and in general, communicate anything. As David Armstrong put it:[22]

> If somebody utters a mere part of a sentence, as a word usually is, what does that communicate? In general, very little. Suppose I just say, "Horse." Has anything much been communicated? In general, a complete sentence is needed to get something across. The *unit of communication* is the sentence. And so a theory of meaning that is based on the communicative function of language must treat the sentence as in some way semantically fundamental.

Note that Armstrong does not say that we cannot make an assertion by uttering 'horse', only that in general we don't. The question of whether non-sentential assertion exists – a question that has been vigorously debated in the last decade – is independent of our main concern.[23] Armstrong's idea is that we bestow meaning on linguistic expressions in the course of the communicative use of language, and since we normally use sentences to communicate, the linguistic expressions upon which meaning is bestowed in the first place must be sentences.

But there is a hole in this argument: one can envision explaining linguistic meaning by appeal to the communicative use of language without appealing to specifically communicative acts. Suppose I utter 'horse' in the course of uttering 'There is a brown horse in the stall' and suppose that in uttering the latter I asserted that there is a brown horse in the stall. Did I perform a speech act in uttering the former? Plausibly, yes. The act was certainly not a communicative act: I did not intend to get a thought across by uttering 'horse'. But when I uttered 'horse' I did refer to horses and I did intend that my audience should recognize this fact. There is a sense in which my referring act was subsidiary to my overall aims – I referred to horses, so as to say in uttering some other words that one of them is in the stall – but this fact does not undermine the status of the act as a genuine speech act.[24]

---

[22] Armstrong (1971): 428.

[23] One can hold up a letter utter the phrase 'from Spain' and thereby assert that the letter in question is from Spain. Whether this is because the phrase uttered really is an elliptical sentence in disguise is a difficult empirical question not to be decided by an appeal to gut feeling. For an extended argument in favor of non-sentential assertion, see Stainton (2006). For some arguments against, see Merchant (2004), Ludlow (2005).

[24] Searle (1969) argues in detail that referring and predicating are speech acts and he usefully contrasts them with illocutionary acts.

Acts like referring and predicating are the sorts of speech acts we perform constantly when using language to communicate our thoughts. These speech acts can bestow meaning on words just as much as communicative acts can bestow meanings on sentences. Even if there were reason to think that communicative acts are explanatorily prior to referring and predicating acts – and I am not convinced that there is – this would have no impact on the question of at hand. It could still be that 'horse' has its meaning because it is used to refer to horses *and* that these acts refer to horses because they are parts of more encompassing illocutionary acts of conveying thoughts about horses. The plausible view that linguistic meaning is ultimately to be explained by appeal to the communicative use of language yields no real support for the context principle.

The second argument concerns the source of context-sensitivity. There are sentences in any natural language that are truth-apt but lack stable truth-conditions. What things must be like for 'You have to read that book' to be true certainly depends on who is being addressed and which book is being demonstrated, and arguably depends also on the time of utterance and the kind of necessity salient in the context. We need to distinguish between two notions of meaning – one that varies and one that stays the same across contexts of use. Following David Kaplan (but eschewing his specific assumptions about how exactly we should think of these two kinds of meaning) we can call the two kinds of meaning *character* and *content.* The fact that a linguistic expression is context-sensitive is a fact about its character, and if the context principle holds we can explain the context-sensitivity of a word via the context-sensitivity of any sentence in which the word occurs. But this claim is false: there are sentences with a constant character containing words with variable ones – 'I am sitting and I am not sitting' is an obvious example.

A defender of the context principle may respond by pointing out that this sentence contains two others whose character is variable, so one could perhaps say that the first occurrence of 'I' gets its character from the first conjunct, and the second occurrence from the second conjunct. However, this concession won't help when it comes to a sentence like 'I exist' which does not contain a further sentence as a constituent.25 The character of this particular sentence *alone* is plainly insufficient basis for explaining the context-sensitivity of the first person singular

---

25 'Could' and 'exist' are also arguably context-sensitive words: the former has a domain of possibilities provides by context, while the latter is inflected for tense.

pronoun. In other words, the meanings of certain sentences cannot account for a crucial feature of the meanings of some of their parts – the context principle fails to hold in full generality.

Indexicals give rise to *submergent* context-sensitivity: as we go down the syntactic tree of certain sentences context-sensitivity may appear even though it wasn't present at a higher level. By contrast, whether there is such a thing as *emergent* context-sensitivity – context sensitivity the pops up at a syntactic node even though it is not present at any node below it – is a highly controversial matter.

Take the example made famous by Charles Travis – the color of a painted leaf.[26] Travis claims that we can say different things in uttering the sentence 'The leaf if green' even if we hold the leaf we are talking about fixed. For suppose that the leaf is painted green but red under the paint; if we use this sentence in a context where we are sorting leafs for decoration we say something true but if we use it when trying to identify what species the leaf belongs to we say something false. This shift in what the speaker said on the two occasions may or may not be attributable to linguistic meaning but for the sake of argument let's assume that what we are detecting here is a shift in the content of the sentence uttered. If there is no shift in the content of any of the words this would be a case of emergent context-sensitivity. But is it plausible that the content of 'green' is the same in the two contexts? Our reason for thinking that the content of 'The leaf is green' differs in the two contexts is the intuition that the speaker said different things in uttering it. But if that is so shouldn't we also concede that the speaker said different things about the same leaf in calling it 'green' in the two scenarios? And shouldn't this be equally good reason to ascribe a shift in the content of the word 'green'?[27]

This style of argument generalizes widely to the sorts of examples that are cited by some as evidence for emergent context-sensitivity: 'This elephant is small', 'The glass is empty', or 'Socrates is ready'. It is, at the very least, a sensible working hypothesis that all context-sensitivity of in complex expressions is traceable to the context-sensitivity of some of their

---

[26] Travis (1985).
[27] For a detailed version of this argument, see Szabó (2010).

constituents.[28] If the hypothesis is true there is no emergent context-sensitivity in natural language even though submergent context-sensitivity is abundant. This isn't just a blow to the context principle; it suggests that an important feature of the meaning of certain words must always be explained in a bottom-up fashion.

My final argument against the context principle begins with the observation that intensional semantics is incompatible with it. There is no way to recover the intensions of 'snow', 'is', and 'white' from the intension of 'snow is white' so we cannot hope to explain that the former have the intensions $\lambda w_s. snow(w)$, $\lambda w_s \lambda P_{\langle s,\langle e,t\rangle\rangle}. P(w)$, and $\lambda w_s \lambda x_e. white(x, w)$, respectively, on the basis of the fact that the latter has the intension $\lambda w_s. white(snow(w), w)$. The problem is not that we cannot uncover the structure of the sentence from its intension: the intension of 'snow is white' together with the structure of this sentence is *still* insufficient to fix the intensions of 'snow', 'is', and 'white'. When we combine semantic values in intensional semantics we lose information: there are infinitely many intensions such that when they are combined via intensional function application they yield the very same intension.

Proponents of the context principle will no doubt reply that the problem is simply with the coarse-grained conception of meaning intensional semantics adopts. We can ensure that semantic values of complexes fully determine the semantic values of their parts by replacing intensional function application (and all other modes of semantic composition that leak information) with something else. The standard line is simply to use ordered pair formation, so instead of IFA, we get OPF:

(2)   a.   IFA: $\lambda w_s. \varphi(w)[\lambda w_s. \psi(w)] = \lambda w_s. \varphi(w)(\psi(w))$

        b.   OPF: $\lambda w_s. \varphi(w)[\lambda w_s. \psi(w)] = \langle \lambda w_s. \varphi(w), \lambda w_s. \psi(w) \rangle$

---

[28] Stanley and Szabó (2000), Stanley (2000), and Szabó (2000b), all defend this thesis. You can think of it as a strong version of the principle of compositionality. It is commonly presupposed in semantic theorizing, although it should be conceded that the usual arguments for compositionality fail to establish its truth.

In this way, semantic values preserve the structure of the expression to which they are assigned – meaning is represented as structured. Thus, the semantic value of 'Snow is white' ends up (3b) rather than (3a):

(3)   a. $\lambda w_s.\,white(snow(w),w)$

        b. $\langle \lambda w_s.\,snow(w), \langle \lambda w_s \lambda P_{\langle s,\langle e,t \rangle \rangle}.\,P(w), \lambda w_s \lambda x_e.\,white(x,w) \rangle \rangle$

Philosophers have been attracted to a structured representation of meaning largely because it useful in handling propositional attitudes.[29] They also like it because they tend to be suspicious of the Fregean idea that most expressions designate functions. It sounds much more intuitive to say that the semantic value of 'snow' is the property of being snow, the semantic value of 'white' is the property of being white, that 'is' is semantically vacuous, and that the semantic value of 'snow is white' is a structured proposition that snow is white.[30] We can even remain neutral on the exact nature of semantic values of lexical items – whatever $[\![snow]\!]$, $[\![is]\!]$, and $[\![white]\!]$ might be, $[\![snow\ is\ white]\!]$ is $\langle [\![snow]\!], \langle [\![is]\!], [\![white]\!] \rangle \rangle$.

Yet, this advantage also reveals a disadvantage. Unlike IFA, OPF puts no constraints on when it can apply: for any two things there is an ordered pair of those things. The simple intensional semantics outlined above for interpreting 'Snow is white' explains why 'snow is' is not meaningful: the intensions don't match – they cannot be combined via IFA at all. It also explains why none of 'snow', 'is', 'white', and 'is white' are truth-evaluable: their intensions are not functions from possible worlds to truth-values, and so, they fail to specify truth-conditions.[31] If we go in for structured meanings we throw all these explanations away.

We do not have to settle the longstanding debate whether structured meanings are worth their price because introducing them won't save the context principle anyway. Ordered pairs, or more

---

[29] The idea goes back to Carnap (1947), who introduced the notion of intensional isomorphism. See also Lewis (1970) and Cresswell (1985).

[30] Depending on one's metaphysics of properties this may help to make the semantic values even more fine-grained – we may be able to distinguish $[\![triangular]\!]$ and $[\![trilateral]\!]$, or $[\![ophthalmologist]\!]$ and $[\![eye\ doctor]\!]$.

[31] Of course, there are limits. Given the simple resources we postulated we cannot explain why 'snow white' fails be meaningful: the intensions of 'snow' and 'white' match – they could be combined via IFA to yield the exact same intension we associated with 'Snow is white.'

generally the sorts of entities one might use as structured semantic values, are all explanatorily posterior to their components. The claim that the semantic value of 'snow' is $[\![snow]\!]$ *because* the semantic value of 'snow is white' is $\langle[\![snow]\!], \langle[\![is]\!], [\![white]\!]\rangle\rangle$ does not bear up to scrutiny. Structured semantic values save a consequence of the the context principle (i.e. the existence of functional determination from the semantic values of sentences to the semantic values of the words within those sentences) by sacrificing the principle itself.

So, my final argument against the context principle is that it is incompatible with both standard unstructured and structured accounts of semantic values. What defenders of the context principle would need is a conception of semantic value which would make the semantic values of words *abstractions* from the semantic values of sentences in which those words appear without making them *arbitrary abstractions*. To use the metaphor Frege liked, they would need a conception of sentence meaning according to which "they can be split up in many ways" – the meaning of '*a* is parallel to *b*' into the meanings of '*a*', 'parallel to' and '*b*', but also to the meanings of 'the direction of *a*', 'identical to' and 'the direction of *b*'.[32] Structured meanings won't give us such flexibility. Mere intensions do, but they would also let us carve up the semantic value of '*a* is parallel to *b*' into the semantic values of '*a*', 'is parallel to or both identical and not identical to', and '*b*' or even to '5', 'plus', '7', 'identical to' and '12'. It is not at all clear that the two desiderata can be jointly met.

5. Against uniformly bottom-up explanations

In the previous section, I presented three arguments against the context thesis. The first – the argument form *speech-acts* – tried to undermine the idea of the explanatory primacy of sentence meaning by pointing out that we do perform speech-acts by uttering mere words and these speech-acts could perhaps be used to explain the meanings of those words. The project of trying to explain linguistic meaning by appeal to the communicative function of language does not force us to embrace the thought that explanations of word meaning must always appeal to the meanings of sentences. The second – the argument from *context-sensitivity* – centered on the fact

---

[32] Frege (1884) §64. The quote is from Frege (1892): 49.

that certain sentences are not context-sensitive even though some words they contain are. This immediately entails that a certain aspect of the meaning of certain words cannot be explained by the meaning arbitrary sentences that contain them. Moreover, the fact that submergent context-sensitivity is abundant and emergent context-sensitivity is dubious suggests that context sensitivity should be explained bottom-up. Finally, the third – the argument from *structure* – points out that none of the semantic values employed in current theories can sustain top-down explanations: unstructured semantic values for complex expressions fail to determine the semantic values of the parts from which those expressions are built, and structured semantic values are explanatorily posterior to their components. I will argue now, similar the considerations used against the idea that semantic explanations are uniformly top-down *also* tell against the much more mainstream idea that semantic explanations are uniformly bottom-up.

Let's start with speech-acts. It makes sense to say that by uttering the noun 'snow' we normally refer to snow or that by uttering the adjective 'white' we normally predicate whiteness. But there are no easily identifiable speech-acts we can link to normal uses of 'is'. The semantic value our mini-theory ascribed to this word was $\lambda w_s \lambda P_{\langle s,\langle e,t\rangle\rangle}. P(w)$, a function that maps possible worlds to a function from predicate-intensions to the corresponding predicate-extension in those worlds. It's easy to imagine semantic theories that assign different semantic values, but it is hard to see how any of them could assign something much less abstract, and even harder to see how such meaning could be associated with a speech act speakers regularly perform.
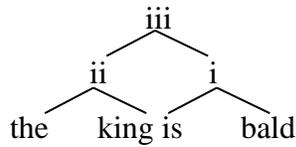
The American Heritage Dictionary defines the meaning of the copular verb 'be' as "to have or show a specified quality of characteristic"; the Oxford English Dictionary defines it as "having a state, quality, identity, nature, role, etc. specified". But how are these things specified? Normally by an expression that combines with the copula to form a predicate, as in the verb phrase 'is white'. The copula seems to be just as Frege thought all words are – indefinable without some reference to a larger expression in which it occurs.

Russell thought that some but not all words work as Frege's context principle demands. *Incomplete* symbols get their meaning through contextual definitions from the meanings of sentences in which they occur; *complete* symbols get theirs on their own. For example, singular

definite descriptions are not directly assigned a meaning in Russell's theory – rather, their meaning is given by specifying the way in which they contribute to the meaning of larger expressions. A sentence of the form 'The F is G' means that something is F, everything that is F is identical to that thing, and everything that is F is G. One may worry that this does not tell us all we need to know about the meaning of the definite article, since not all sentences in which it occurs are of the form 'The F is G.' Russell himself thought that sentences of the form 'The F exists' require a separate clause; he proposed that a sentence of this form means that something is F, and everything that is F is identical to it.[33] This still leaves open the interpretation of a great many sentences containing singular definite descriptions – 'He is the present king of France', 'Every king except for the present king of France is bald', 'The present king of France or the present king of England is bald', 'So incredibly bald is the present king of France that he cannot show himself without a wig', etc. Still, one might reasonably hope that, with sufficient ingenuity, the indirect method of specifying meanings for sentences containing the definite article could be made to work.

Once we have the contextual definition it is routine to turn it into an explicit one if we have the appropriate logical resources. To build on the intesional semantics we used before, the intension of the sentence 'The king is bald' could be determined from the intensions of its constituents as follows:

(4)   a.

$$
\begin{array}{c}
\text{iii} \\
\diagup\quad\diagdown \\
\text{ii}\qquad\text{i} \\
\diagup\diagdown\quad\diagup\diagdown \\
\text{the}\quad\text{king}\ \text{is}\quad\text{bald}
\end{array}
$$

    b. $[\![\text{the}]\!] = \lambda w_s \lambda P_{\langle s,\langle e,t\rangle\rangle} \lambda Q_{\langle s,\langle e,t\rangle\rangle} \exists x_e (\forall y_e (P(y,w) \leftrightarrow x = y) \wedge Q(x,w))$

        $[\![\text{king}]\!] = \lambda w_s \lambda x_e. king(x,w)$

        $[\![\text{is}]\!] = \lambda w_s \lambda P_{\langle s,\langle e,t\rangle\rangle}. P(w)$

        $[\![\text{bald}]\!] = \lambda w_s \lambda x_e. bald(x,w)$

    c.   IFA: $\lambda w_s. \varphi(w)[\lambda w_s. \psi(w)] = \lambda w_s. \varphi(w)(\psi(w))$

    d. $[\![\text{i}]\!] = [\![\text{is}]\!][[\![\text{bald}]\!]] = \lambda w_s \lambda x_e. bald(x,w)$

---

[33] Russell (1905).

$[\![ii]\!] = [\![the]\!][\![king]\!] =$

$\lambda w_s(\lambda Q_{\langle s,\langle e,t\rangle\rangle}\exists x_e(\forall y_e(king(y,w) \leftrightarrow x = y) \land Q(x,w))$

$[\![iii]\!] = [\![the\ king]\!][\![is\ bald]\!] =$

$\lambda w_s(\exists x_e(\forall y_e(king(y,w) \leftrightarrow x = y) \land bald(x,w))$

The question is how such a theory is put to explanatory use. Russell's main insight could be captured by saying that the intension of 'the' turns out to be this particular complicated function *because* the intension of sentences of the form 'The *F* is *G*' is as Russell characterized it: $[\![the]\!]$ was simply reverse-engineered so that when combined with $[\![king]\!]$ and $[\![is\ bald]\!]$ (or any other pair of one-place predicates) we obtain the right result. The same could be said about 'is': $[\![is]\!]$ was designed in order to ensure that $[\![bald]\!] = [\![is\ bald]\!]$ (and in general, combining the copula with a one-place predicate results in a predicate with the same intension). By contrast, 'king' and 'bald' have the intensions they do not in virtue of how they combine with other words, but in virtue of how they are employed in referring and predicating speech-acts. The partial explanatory reversal is welcome, for we would find it hopeless to specify the meaning of the definite article or the copula in isolation. There doesn't seem to be anything we mean by uttering such expressions alone – what we mean to do with them can only be articulated in connection with our uttering other expressions as well. This is the sense in which they can be thought of as incomplete symbols, and this is the reason why we should regard them as expressions whose meaning is to be explained by appeal to the meanings of other expressions.

The next strike against the idea of uniformly bottom-up explanations in semantics concerns context-sensitivity. I argued that the meanings of indexical words are cannot be explained by the meanings of some arbitrary sentence that contains them. It seems that the meanings of these can be explained directly without appeal to any other expression: 'I' has it meaning because speakers tend to use this word to refer to themselves, 'here' has its meaning because speakers tend to use this word to refer to the place they are at, etc. But there are other small items within syntactic trees that are just as context-sensitive as the indexicals but are almost certainly not susceptible to a similar use-based account.

Consider *PRO*, an empty category postulated by many syntactic theories as the subject of a non-finite clause, as in 'Jack$_i$ wants *PRO*$_i$ to go home'. This is a variable whose interpretation is linked to that of its antecedent – in our case 'Jack'. *PRO* is not typically described as a word, since it is unpronounced, but the difference between items that do and items that don't have a phonological form is not a deep one, as far as the organization of grammar is concerned. *PRO* features within sentences and it has no further parts – the straightforward assumption is that it comes from the lexicon just like affixes, roots, and compound words. *PRO* is context-sensitive, albeit not in quite the same way indexicals are: its content depends solely on the specifically linguistic context in which it occurs. In this regard it is somewhat like reflexives, and anaphoric pronouns. Which is why the meaning of *PRO* is not the sort of thing we could explain without relying on other expressions with which it is in construction. And there is definitely no speech-act we could perform whose content is just the content of *PRO* on a given occasion of its use.

Not everyone believes in *PRO*.[34] But almost everyone believes in some unpronounced variables: it is very common to allow some movement in syntax, where the displaced element leaves a trace behind, as in 'Who$_i$ does Jack want to see t_i' or 'There is some book$_i$ each student read t_i'. Semanticists tend to postulate variable-like elements which are sometimes bound and sometimes acquire their value from the context. 'You left the stove on' arguably contains a time variable whose contextual setting allows the sentence to say that you left the stove on a short while ago, 'Every bottle is empty' presumably contains a domain variable which is how it ends up expressing the thought that every bottle at home is empty, 'Jill is tall' possibly contains a standard variable responsible for the fact that it can have the content that Jill's tallness exceeds what is normal for sixth graders. Whether there is good reason to postulate such variables is a matter of debate but it is hard to imagine that semantics will eventually settle on a view that eschews them all.[35] The ones that are real are all counterexamples to the idea that semantic explanations invariably trace syntactic trees upwards.

---

[34] See Hornstein (1999) for arguments that it can be eliminated from syntax.

[35] Jacobson (2002) argues that we should eschew variables in syntax and semantics completely. If variable-free grammars turn out to be better than the currently popular ones (which would mean that our naïve conceptions of what the syntax and semantics of natural languages look like are quite fundamentally mistaken) this particular argument against uniformly bottom-up semantic explanations won't fly.

But couldn't one dig in, deny the need for variables in syntax, reject the idea that lexical items get their meanings from speech-acts, and insist that although we use our intuitions about the truth-conditions of sentences to *discover* the meanings of small words like the definite article or the copula, linguistic meaning is to be ultimately *explained* uniformly bottom-up? Perhaps we could. What is less clear is whether it is worth the effort.

This brings us to the final argument based on consideration of linguistic structure. Consider again the English sentence 'Snow is white.' The Hungarian translation of this sentence is 'A hó fehér'. The Hungarian word for snow is 'hó', for white is 'fehér'. The 'a' at the beginning of the Hungarian sentence is the definite article, which cannot be dropped on pain of ungrammaticality. The copula in Hungarian is 'van', which cannot be added on pain of ungrammaticality. Nonetheless, a good semantic theory of English and a good semantic theory of Hungarian should jointly establish that these two sentences have the same meaning (except for differences we ignore when we model meanings via truth-conditions). Moreover, they should do that without implausibly telling us that the meanings of the nouns and the adjectives in the two sentences differ (except for, again, the differences we ignore).

One way this could be done is by positing unpronounced elements – a copula in the Hungarian sentence and a definite article in English sentence. The other option is to say that the semantic rules used to derive the meaning of 'Snow is white' in English absorb the meaning of the definite article, and the semantic rules used derive the meaning of 'A hó fehér' in Hungarian absorb the meaning of the copula. There may be a bias towards the first approach in contemporary semantics but everyone should agree that there is no *a priori* to assume that the best semantic theories of these languages will go that way. As long as the second option live, we should concede that there may well be a connection between semantic rules available in a language and the meanings of expressions like 'is' and 'the.'

If there are such connections, they could in principle be explained cross-linguistically. We could say, for example, that the definite article means in English what it does is because that is just the meaning we would need to give to an unpronounced element in order compositionally derive the meaning of 'Snow is white' using semantic rules available in other possible human languages,

28

such as Hungarian, where the definite article is actually pronounced in the corresponding sentence. And we could also say that the Hungarian 'van' means what it does is because that is just the meaning we would need to give to an unpronounced element in order compositionally derive the meaning of 'A hó fehér' using semantic rules available in other possible human languages, such as English, where the copula is actually pronounced in the corresponding sentence.

I realize that these cross-linguistic explanations sound highly speculative. Still, the fact that currently we have no conception what an explanation of the meanings of the definite article or the copula might look like should make us weary of dismissing them out of hand. The fact that proponents of uniformly bottom-up explanations is semantics would do that is a strike against their view.

6. The explanatory divide in the lexicon

Let us call an expression whose meaning is not explained in terms of the meanings of any others a *semantic primitive.*[36] I have argued that some words are semantic primitives while others are not. I put 'I' and 'bald' in the former category; 'the' and 'be' in the latter. Given the considerations in the previous two sections, the most natural way to draw the line is to say that lexical items that belong to *open categories* are semantic primitives while expressions in *closed categories* are not.

The distinction between open and closed categories is a good candidate for being a linguistic universal. We don't have a definition of what it is to belong to these categories but we do have some fairly reliable tests. Closed categories, as their names suggest, are relatively stable – languages tend not to gain or lose these expressions easily. We routinely coin nouns, verbs, or adjectives and we frequently borrow them from other languages; not so for articles, determiners,

---

[36] The term is Davidson's, although his definition presupposes certain things I want to leave open: "Let us call an expression a *semantical primitive* provided the rules which give the meaning for the sentences in which it does not appear do not suffice to determine the meaning of the sentences in which it does appear." Davidson (1965): 9.
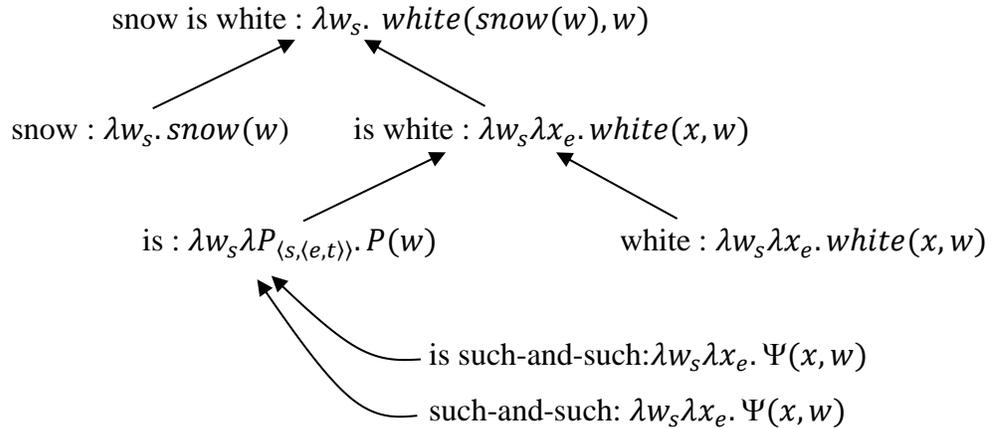
or conjunctions. Relatedly, expressions in closed categories tend to be from the older strata of language and they tend to be short. In addition, closed category items don't participate in derivational morphology. There are suffixes in most languages that allow you to turn a verb into a noun or an adjective into an adverb but there are no ways to turn an auxiliary into a preposition or a determiner into a conjunction. Expressions in closed categories seem to be bound up with grammar in a way in which expressions in open categories are not. This is why traditional grammar and logic viewed the former as *syncategoremata* – function words whose meanings are exhausted in their contribution to the form of expressions. This is closely related to Russell's notion of an *incomplete symbol* – an expression whose meaning is to be explained by how it contributes to the meanings of larger expressions.

Let us look at the toy intensional semantics I gave for 'Snow is white' and the compositional explanation it gives for why this sentence has the intension it does. The explanation bottoms out with the semantic facts that the three words in this sentence have the intensions they do. This is satisfactory regarding 'snow' and 'white.' There is no semantic explanation of why these have their intensions; I conjectured that they get them from speech-acts we perform uttering these words – from acts of referring to snow and acts of predicating whiteness. But 'is', being a closed category expression, is not a semantic primitive. If we wanted to explain why 'is' has the intension it does, we could use an arbitrary adjective and the verb phrase we get combining it with the copula and rely on the fact that these two expression have the same intension. Thus, we could say that the intension of 'is' is $\lambda w_s \lambda P_{\langle s, \langle e, t \rangle \rangle}. P(w)$ because, given the fact that the intensions of 'cold' and 'is cold' are identical, this is the function that yields the latter when applied to the former via intensional function application. This way we keep the compositional bottom-up explanation of the intension of 'Snow is white' intact – we just tucked an extra layer of explanation to account for the intension of 'is'.

Now, this explanation rests on the implausible claim that the intension of 'is' has something to do with the intensions of 'cold' and 'is cold'. But 'cold' and 'is cold' were picked arbitrarily; any other adjective and corresponding copular verb phrase would have done the job. What I suggest is that besides ordinary adjectives we also posit a *schematic adjective* and besides ordinary verb phrases consisting of a copula and an adjective complement we posit a *schematic phrase*. I think
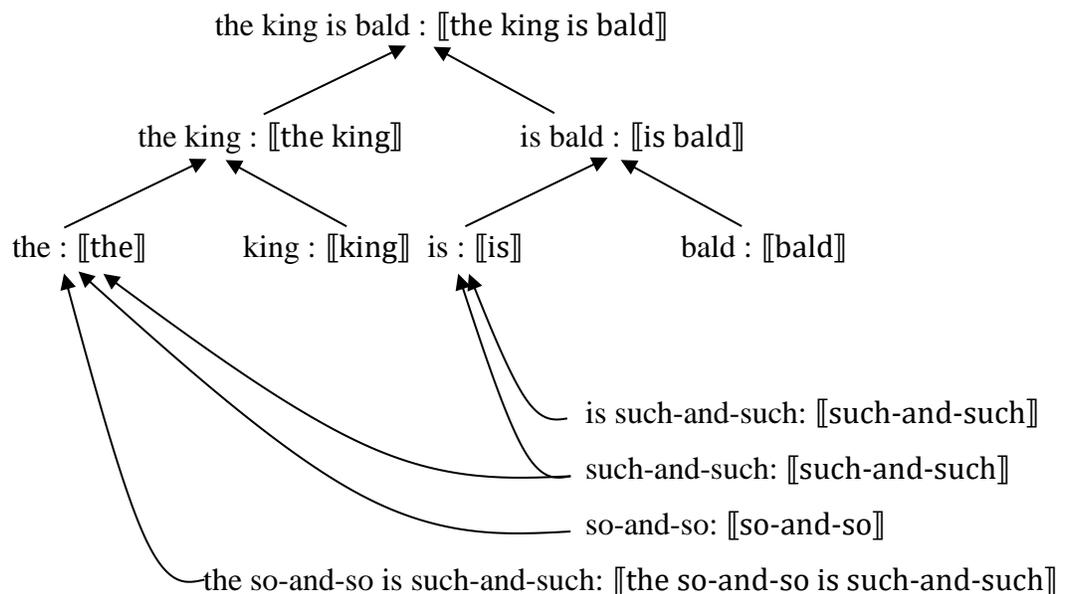
we should regard these as genuine linguistic expressions. In English, we even have a way of pronouncing them – 'such-and-such' and 'is such-and-such' do the job. 'Such-and-such' is a *bona fide* lexical item which shows up in any decent dictionary and which has been around since the 15th century. Not surprisingly, dictionaries are not too good in explaining its meaning, other than pointing out that it is used to talk about something not specified. We can do better – we can say that the intension of 'such-and-such' and well as the intension of 'is such-and-such' is a *function type* specifiable via the schema $\lambda w_s \lambda x_e. \Psi(x, w)$. Then we can explain why 'is' has the intension is does in terms of the intensions of the two schematic expressions. I suggest that both are semantic primitives.

Here is a schematic representation of the semantic explanation suggested. The nodes represent semantic facts, i.e. that a certain expression has a certain meaning; arrows point from explanans to explanandum. The leaves of the tree are semantic facts to be explained non-semantically. He straight arrows trace syntactic structure; the curved ones mark semantic explanations for the meaning of 'is', the sole closed category expression within the sentence.

$$\text{snow is white} : \lambda w_s. \ white(snow(w), w)$$

$$\text{snow} : \lambda w_s. \ snow(w) \qquad \text{is white} : \lambda w_s \lambda x_e. white(x, w)$$

$$\text{is} : \lambda w_s \lambda P_{\langle s, \langle e, t \rangle \rangle}. P(w) \qquad \text{white} : \lambda w_s \lambda x_e. white(x, w)$$

$$\text{is such-and-such} : \lambda w_s \lambda x_e. \Psi(x, w)$$

$$\text{such-and-such} : \lambda w_s \lambda x_e. \Psi(x, w)$$

The explanation for why the semantic value of 'The king is bald' is similar, albeit slightly more complex. Just like with 'Snow is white' we keep the bottom-up explanation intact but add additional layers to account for the intensions of the closed category expressions 'is' and 'the'. The explanation of why 'is' has the intension $\lambda w_s \lambda P_{\langle s, \langle e, t \rangle \rangle}. P(w)$ is the same as before, based on the intensions of 'such-and-such' and 'is such-and-such'. The explanation of why the intension of 'the' is $\lambda w_s \lambda P_{\langle s, \langle e, t \rangle \rangle} \lambda Q_{\langle s, \langle e, t \rangle \rangle} \exists x_e (\forall y_e (P(y, w) \leftrightarrow x = y) \wedge Q(x, w))$ relies on the schematic

clause 'the so-and-so is such-and-such', the schematic noun 'so-and-so', and the schematic phrase 'is such-and-such'. The intensions of the last two are function types from possible worlds to functions from entities to truth-values. The intension of the schematic clause is a function type from possible worlds to truth-values that assigns the value True to a possible world w just in case there is a unique object o that is within the extension of 'so-and-so' at w and o is also within the extension of 'is such-and-such' at w. In other words, we follow Russell, who said that "'the so and so is such and such' will mean that 'x is so and so and nothing else is, and x is such and such' is capable of truth."[37] 'The' has the intension it does because that is just the function that yields the intension of 'the so-and-so is such-and-such' when first combined with the intension of 'so-and-so' and then with the intension of 'is such-and-such'.

the king is bald : ⟦the king is bald⟧

the king : ⟦the king⟧   is bald : ⟦is bald⟧

the : ⟦the⟧   king : ⟦king⟧   is : ⟦is⟧   bald : ⟦bald⟧

is such-and-such: ⟦such-and-such⟧

such-and-such: ⟦such-and-such⟧

so-and-so: ⟦so-and-so⟧

the so-and-so is such-and-such: ⟦the so-and-so is such-and-such⟧

These explanations are, of course, at best preliminary. They are only as good as the syntax and semantics they are built upon, and those are subject to revisions as we learn more about natural languages. They serve here to illustrate a philosophical point: that there could be a principled way to supplement the usual compositional explanations with further explanations for the meanings of closed category words within them. These explanations rely on a distinguished class of expressions – some lexical ('such-and-such', 'so-and-so'), some not ('the such-and-such is so-

---

[37] Russell (1911): 128.

and-so') – whose meanings are more abstract then the meanings of other expressions. If ordinary expressions of a certain syntactic category have a certain type of meaning, the meaning of the schematic expression will be just that type. Thus we can think of these meanings as *abstractions* – the meaning of the adjective 'such-and-such' is what all adjective meanings have in common, the meaning of 'very such-and-such' is what the meanings of all phrases of the form [$_{AP}$ very [$_A$ X]] have in common, the meaning of the noun 'so-and-so' is exactly what all noun meanings have in common, the meaning of 'above the so-and-so' is what the meanings of all phrases of the form [$_{PP}$ above [$_{DP}$ [$_D$ the][$_N$ so-and-so]]] have in common, and so on.

On the model proposed, the semantic primitives in natural languages are (i) open category lexical expressions, (ii) schematic lexical expressions, and (iii) schematic complex expressions containing exclusively closed category lexical expressions and a single open category lexical expression as their ultimate syntactic constituents. The meaning of every other expression is explained in two phases. First, we give semantic explanations for the meanings of the closed category expressions that occur in them as constituents, and then providing a bottom-up compositional explanation.

7. Conclusion

The proposal of the previous section vindicates the idea that semantic theories as developed by linguists do what one might want them to do: explain why certain expressions mean what they do in terms on the meanings of other expressions. The explanatory model offered is conservative: it accepts all the bottom-up explanations strongly suggested by the structure of compositional theories as genuine while accommodating the intuition that function words are not semantic primitives. That function words are within the purview of semantic theory in a way content words are not is widely appreciated among working semanticists: a cursory look at the content of any semantic journal shows that the meanings of 'only', 'each', 'not' is a matter of active research while the meanings of 'elm', 'walk', or 'happy' are not.[38]

---

[38] For defense of the claim that grammar is all about the "little words", see von Fintel (1995).

The consequence of adopting this view for philosophy of language are significant. It entails that when a philosopher proposes a theory about what the meanings of the so-called logical constants in natural languages are, she is stepping on the toes of linguists. Words like 'and' and 'or' are function words, and a proper characterization of their meaning is impossible without paying close attention to compositional detail. More importantly, it also entails that when we speculate about what might explain the meanings of content words we should look for plausible non-semantic accounts, rather than appealing to the meanings of larger expressions in which these words occur. The meaning of 'snow' has a significant role to play in explaining why 'snow is white' means what it does, and so, on pain of circularity we cannot appeal to the latter in seeking an account of the former.39

References

Alston, W. 1974. 'Semantic Rules.' In M. K. Munitz and P. Unger (eds.), *Semantics and Philosophy*. New York: New York University Press, 17 – 48.

Armstrong, D. 1971. 'Meaning and Communication.' *Philosophical Review* **80**: 427 – 47.

Carnap, R., 1947. *Meaning and Necessity*. Chicago: University of Chicago Press.

Cresswell, M., 1985. *Structured Meanings: The Semantics of Propositional Attitudes*. Cambridge, MA: MIT Press.

Davidson, D., 1965. 'Theories of Meaning and Learnable Languages.' Reprinted in *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 2001, 3 – 16.

Davidson, D., 1967. 'Truth and Meaning.' Reprinted in *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 2001, 17 – 42.

Fine, K., 2012. 'A Guide to Ground.' In F. Correira and B. Schneider eds., *Metaphysical Grounding*. Cambridge: Cambridge University Press, 37 – 80.

von Fintel, K., 1995. 'The Formal Semantics of Grammaticalization.' *NELS 25 Proceedings* (GLSA, UMass Amherst), Vol. 2: *Papers from the Workshops on Language Acquisition & Language Change*, 175 – 189.

Frege, G. 1884. *The Foundations of Arithmetic*. Oxford: Blackwell, 1950.

Frege, G., 1892. 'On Concept and Object.' Reprinted in B. McGuinness (ed.), *Collected Papers on Mathematics, Logic, and Philosophy*. Oxford: Basil Blackwell, 1984, 182 – 94.

Frege, G., 1919. 'Notes for Ludwig Darmstaedter.' In H. Hermes et al. (eds.), *Posthumous Writings*. Chicago: University of Chicago Press, 1979, 253 – 7.

Hornstein, N., 1999. 'Movement and Control.' *Linguistic Inquiry* **30**: 69–96.

Horwich, P., 1998. *Meaning*. Oxford: Oxford University Press.

Horwich, P., 2005. *Reflections on Meaning*. Oxford: Oxford University Press.

Jacobson, P., 2002. 'The (dis)organization of grammar.' *Linguistics and Philosophy* **25**: 601 – 26.

Kaplan, D., 1989. 'Demonstratives.' In *Themes from Kaplan*, Almog, J., Perry, J., and Wettstein, H. (eds.), Oxford: Oxford University Press, 481–563.

Kim, J., 1989. 'Mechanism, Purpose, and Explanatory Exclusion'. *Philosophical Perspectives* **3**, 77 – 108.

Lewis, D., 1970. 'General Semantics.' *Synthese* **22**: 18–67.

Lewis, D. 1975. 'Languages and Language.' Reprinted in *Philosophical Papers, volume 1*, 1983, 163 – 89. Oxford: Oxford University Press.

Lewis, D., 1980. 'Index, Context, and Content.' Reprinted in his *Papers in Philosophical Logic*, 1998, 21 – 44. Cambridge: Cambridge University Press.

Ludlow, P., 2005. 'A Note on Alleged Cases of Non-Sentential Speech.' In R. Eluguardo and R. Stainton (eds.), *Non-Sentential Speech*. Dordrecht: Kluwer, 95 – 108.

Merchant, J., 2004. 'Fragments and ellipsis.' *Linguistics and Philosophy* **27**: 661–738.

Russell, B., 1905. 'On denoting.' Reprinted in A. P. Martinich (ed.), *The Philosophy of Language.* Oxford: Oxford University Press, 2001, pp. 212 – 220.

Russell, B. 1911. 'Knowledge by acquaintance and knowledge by description.' *Proceedings of the Aristotelian Society (New Series)*, vol. 9, (1910-1911), 108 – 28.

Sainsbury, M., 1991. *Logical Forms: An Introduction to Philosophical Logic.* Oxford: Blackwell.

Schaffer, J., 2012. 'Grounding, transitivity, and contrastivity.' In F. Correira and B. Schneider eds., *Metaphysical Grounding.* Cambridge: Cambridge University Press, 122 – 38.

Searle, J., 1969. *Speech Acts.* Cambridge: Cambridge University Press.

Speaks, J., 2011. 'Theories of Meaning.' *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2011/entries/meaning/>.

Stainton, R., 2006. *Words and Thoughts: Subsentences, Ellipsis, and the Philosophy of Language*, Oxford: Oxford University Press.

Stalnaker, R. 1997. 'Reference and Necessity.' In *A Companion to the Philosophy of Language* C. Wright and B. Hale (eds.), Oxford: Blackwell, 534–54.

Stanley, J., 2000. 'Context and logical form.' *Linguistics and Philosophy* **23**: 391 – 434.

Stanley, J. and Z. Szabó 2000. 'On Quantifier Domain Restriction.' *Mind and Language* **15**: 219 – 61.

Szabó, Z., 2000a. 'Compositionality as Supervenience.' *Linguistics and Philosophy* **23**: 475 – 505.

Szabó, Z., 2000b. 'Adjectives in context.' In I. Kenesei and R., Harnish (eds.), *Perspectives on Semantics, Pragmatics, and Discourse*. Amsterdam: John Benjamins, pp. 119 – 146.

Szabó, Z., 2010. 'The Determination of Content.' *Philosophical Studies* **148**: 253 – 272.

Travis, C. 1985. 'On what is strictly speaking true.' *Canadian Journal of Philosophy* **15**: 187 – 229.

Wittgenstein, L. 1921. *Tractatus Logico-Philosophicus.* London: Routledge, 1981.