Luca Gasparri

Lexical Meaning in Truth-Conditional Semantics

Diametros nr 39, 182-202

2014

Diametros 39 (2014): 182–202
doi: 10.13153/diam.39.2014.571

LEXICAL MEANING IN TRUTH-CONDITIONAL SEMANTICS

– Luca Gasparri –

Abstract. The paper offers a critical review of the role played by lexical meaning in the earlier stages of philosophical semantics and truth-conditional semantics. I shall address, both historically and theoretically, the relative neglect of lexical semantics within these fields, and argue that the approach to word meaning fostered in extensional frameworks is overall inconsistent with the customary assumption that truth-theoretic semantics can be considered a semantic theory *proprio sensu*.

Keywords: word meaning; lexicon; reference; theories of meaning; semantics.

1. A Strange Oblivion

Philosophical semantics has customarily been poorly interested in the study of lexical meaning, and has preferred to focus on the structural and compositional features of sentences. If we take a look at how the debate on semantic issues has evolved from its foundations to the present day, this is somewhat surprising. First, because the idea of providing a semantic theory for natural languages was largely inspired by the work of lexicographers.[1] Second, because many of the issues raised by Frege in his seminal 1892 papers dealt precisely with the analysis of the meaning of individual words. Granted, even the commitment to composition later endorsed by philosophical semantics and model-theoretic semantics is due to an epistemological lesson that disciplines interested in the study of meaning have learnt via Frege's account of functional application.[2] Still, it is curious that despite Frege's persisting attention for what he called the *Gedankenbausteine* (roughly, "the building blocks of thought"),[3] his emphasis on the fact that the analysis of the contribution of lexical constituents to the truth conditions of sentences should be conceived as a tool to provide an explicit representation of the "essential meaning" of lexemes,[4] philosophical semantics absorbed just the compositional agenda of his approach to meaning. How did this happen?

[1] Geeraerts [2010].

[2] E.g., Heim, Kratzer [1998].

[3] Frege [1979].

[4] Frege [1980].

The main reason could be that, before Tarski, it was generally held that the study of lexical meaning was free of genuine problems. In Wittgenstein's *Tractatus*, for example, the semantics of names is reduced to their naming whatever object or worldly feature they denote in circumstances of use, and phrasal constituents are viewed as symbols that can be processed and interpreted only in meaningful linguistic acts. Since, the argument goes, we have reasons to maintain that all we can say about the meaning of lexemes is their use *qua* phrasal constituents or *qua* elements of a speech act, questions about the sense of free--standing lexical units are doomed to be pointless. Words taken in isolation do not have any sense; and even if they had something similar to a sense, such quasi-sense would be irrelevant for the enterprise of a semantic theory.[5]

Some time after the first Wittgenstein, Tarski's identification of semantics with the theory of reference fostered a partially renewed interest in the analysis of lexical expressions, but the situation did not change much. Tarski viewed the lexical units of formal languages as descriptive constants mapping on a domain of interpretation and classified them according to the set-theoretic type of their reference. Proper names were paired with entities in the domain, unary predicates were paired with subsets of the domain, *n*-place predicates were paired with sets of ordered *n*-tuples combining entities in the domain, and so forth.[6] Meaning analysis was accordingly conceived as standing in a one-to-one correspondence with grammatical categorization,[7] and the semantics of descriptive constants assumed the shape of a function taking their syntactic type as input and giving a proper portion-type of the domain as output. Yet, a problem soon became apparent: lexical analysis could not be plausibly extended to natural languages and reduced to the identification of syntactic roles at the same time. Take the unary predicate *cat* (roughly, $[\![cat]\!] = [\lambda x\colon x \in D\ .$ *true* if x is a cat and *false* otherwise]). Even if the correspondence between the grammatical category of *cat* and the type of domain mapping regularly exhibited by unary predicates predicts that the reference of *cat* will be a set of non-ordered entities in the domain, there is no way we can know in advance, just by relying on this procedure, that the set designated by *cat* will be the set containing all and only the cats available in the domain. Spelling out the semantics of the predicate via an appropriate function gives us no clue as to how the domain should be carved in order to obtain the subset matching the extension of *cat*. Yet, it seems that the knowledge about how

[5] Wittgenstein [1922].

[6] Tarski [1944, 1983].

[7] Partee [1976], Marconi [1997].

this operation should be performed is an essential part of our grasp of the meaning of natural language lexemes.

In point of fact, Tarski did not pursue anything more than a taxonomy of formal predicates based on the set-theoretic type of their denotation, and gave a number of reasons for not extending his theory to natural languages (e.g., the problem of there being no systematic way of deciding whether a given natural language sentence is well-formed).[8] Nonetheless, many have insisted on the application of a Tarskian agenda to natural language semantics, following the immense influence exercised on the philosophy of language by Davidson's adoption of Tarski's "material adequacy condition".[9] Tarski's "Convention T" suggested that a formal definition of the predicate *true* should depend on the provision, for every sentence S in an object language, of a matching sentence P in a metalanguage, functioning as a translation of S. Accordingly, "T-sentences" were conceived as having the form "S is true in L iff P", and the notion of "adequate theory of meaning" was reduced to the notion of "theory capable of generating a T-sentence for every sentence in the object-language". Davidson reversed Tarski's approach, which was intended to arrive at a theory of truth via a theory of meaning, proposed that meeting the requirements of Tarski's Convention T could be seen as the basic requirement for a complete theory of meaning, and assumed that a complete theory of meaning for natural languages could be built on the basis of an apparatus requiring little more than first-order quantificational logic as supplemented by set theory, and a purely extensional definition of truth.[10] However, there are at least *three* major problems with this approach. First, natural languages are far richer than the well-behaved formal languages Tarski was dealing with, and contain features (e.g., reported speech, adverbial expression, imperatives) whose semantic breakdown seems to require resources beyond those of first-order logic or extensional analysis. Second, while Tarski relied on the notion of "translation" as a means to provide a definition of truth (one of the requirements of Convention T was that the sentence on the right side of a T-sentence had to be a translation of the sentence on the left side), Davidson's use of the notion of "truth" to provide an account of meaning deprived him of a way to constrain T-sentences so as to ensure that they generated correct specifications of sentential meanings (e.g., how do we rule out "*Schnee ist weiss* is true iff grass is green"?). Third, Davidson's writings on the subject had all

[8] McDowell [1998].

[9] Davidson [1984].

[10] Malpas [2009].

a programmatic character. Davidson never formulated along the above lines a theory of truth for a natural language, nor for a portion of a natural language. He merely insisted on the philosophical motivations behind the project, but he never pursued it, he did not indicate how to pursue it, nor it is entirely clear how it could be pursued, especially if one thinks about how the semantics of sentences ascribing propositional attitudes could be spelled out while endorsing Davidson's antipathy for the conceptual apparatus of possible worlds and his rejection of the notions of "sense" and "intension".[11] In light of all this, a question arises naturally: is a Tarski-style semantics for natural language lexemes at least an element of their complete analysis, or is it something we should expect a mature semantic theory to be void of?

2. Gradual Recollection

Quine argued fervidly for the second option.[12] He maintained that since meaning could not be accounted for exclusively in terms of reference, Tarskian semantics could not be interpreted as a theory of meaning. In his *Notes*, Quine dealt mainly with the behavior of modal and intensional contexts, but felt that the problems he was addressing could be tackled at best by recognizing that they stemmed from the assumption that "meaning" and "designation" could be considered equivalent notions. Consider the case of non-designating words such as fictitious proper names. One can say that the phonographic unit *Pegasus* has no reference on condition that she understands its meaning, and it is only in result of the knowledge of its meaning that one can establish that the extension of *Pegasus* is empty. Not only *Pegasus* does not commit us to the existence of a winged horse because we can use the machinery of first-order logic and Russell's theory to convert *Pegasus* into a definite description having the form "the thing that is Pegasus" or "the thing that pegasizes".[13] More importantly, Quine continued, assume that meaning is just a matter of reference and take two coreferential terms such as *Morning Star* and *Evening Star*. If the semantics of *Morning Star* and *Evening Star* were to be provided in purely referential terms, anyone who knew the meaning of *Morning Star* and *Evening Star* should have a notion of their synonymy. Since in a Tarskian framework terms that have exactly the same denotation are *ipso facto* synonymous, the relation of synonymy existing between *Morning Star* and *Evening Star* should be an essential part of the ability to use

[11] Casalegno [1997].

[12] Quine [1952].

[13] Juhl, Loomis [2009].

such expressions in a competent fashion. Contrary to evidence, awareness of their coreferentiality (beyond the level of modes of presentation) should be an immediate correlate of the knowledge of their linguistic meaning, not something speakers should determine synthetically.

Quine later thought that the quasi-psychological notion of meaning he had in mind was unsuited to be used for any scientific purpose, and that serious semantics should simply avoid the attempt to provide a theory of the non--extensional aspects of non-logical words. Reference was clearly insufficient to capture the contribution of lexical items to the truth conditions of sentences, and he thought that the missing part of the picture was elusive enough to support the intuition that its investigation would have inexorably fallen outside the domain of respectable science. Quite ironically, however, it was precisely the combination of Quine's remarks with Church's Fregean revival at the beginning of the 1950s what contributed to the *renaissance* of lexical meaning that led to Carnap's mature semantics.[14]

It is well-known than the pivotal notion of Carnap's treatment of lexical meaning is that of "intension".[15] In the contemporary formulation of the notion, intensions are functions from indices, usually possible worlds, to extensions.[16] Exactly as we saw with the one-to-one correspondence between set-theoretic types and syntactic categories delineated by Tarskian semantics, in Carnap intensional types are determined by the grammatical category of the expression they are coupled with. But there is one important novelty. Intensions now serve the purpose of distinguishing expressions that speakers characterize as having a different meaning even if they have the same reference. Semantic dissimilarity among coreferential terms is accordingly regarded as an intensional divergence: two expressions are said to be intensionally non-equivalent, and hence to differ in meaning, if there is at least one possible world in which their extensions differ. Still, as it had happened with Tarski, Carnap's semantics is able to predict the intensional type of lexical items on the basis of their syntactic properties while failing to differentiate the intensions of expressions belonging to the same syntactic class. The theory succeeds in envisaging that common nouns can be treated as predicates and that their intensions fall under the same logical type (i.e., they all are functions from possible worlds to sets of individuals), but it is unable to appreciate why and how they single out different sets of objects. This has two

[14] Marconi [1997].

[15] Carnap [1956].

[16] E.g., Lycan [2008].

consequences. The first is that Carnap's system is insufficient to analyze the meaning of descriptive constants the way we would want a complete referential semantics to be able to do, because the information about the meaning of lexemes returned by extensional analysis does not come in a format or in a level of specification sufficient to determine their reference in circumstances of use. The second is that if sentential meaning is identified with truth conditions and it is assumed that the composition of the truth conditions of a sentence cannot be exclusively sensitive to reference, then the analysis of sentential meaning offered by Carnap is itself constitutively incomplete.[17]

This observation can be easily corroborated by considering the way Carnap's intensional semantics copes with the distinction between sentences whose necessary truth or falsity depends on semantic factors and sentences whose necessary truth or falsity depends on logic. Semantic necessity and logical necessity simply collapse on one another. We can separate them in specific cases by appealing to intuitions, but the theory does not provide us with any well--behaved instruments to single out semantic necessity and isolate it from logical necessity. If we only assume that lexical intensional types are homomorphic with grammatical types and that phrasal intensions are to be distinguished by looking at the truth value they designate in different worlds, we are given no general method to discriminate between sentences that denote a given truth value in all worlds for semantic reasons and sentences that denote a given truth value in all worlds for logical reasons. How do we proceed then?

3. Meaning Postulates

Carnap provided his own solution to the *impasse* with the theory of meaning postulates. In its basic form, a meaning postulate is a stipulation on the set-theoretic relation existing between the extensions of two or more non--functional words.[18] For example, consider (1).

(1) $(\forall x)$ (bachelor(x) $\supset$ $\neg$married(x))

In (1) it is stipulated that, for every x, if x is a member of the subset of the domain singled out by *bachelor*, then x falls outside the extension of *married*. Although meaning postulates were expressedly designed as constraints on the relations among the extensions of lexemes, it is natural to ask whether they can be viewed as restrictions on properties of the intensions, not the extensions, of content words.

[17] Marconi [1989].

[18] For a more detailed overview, see Zimmerman [1999].

To answer this question, take (1) and consider it a valid postulate of our semantic theory. Then take the conditional (2).

(2) $(\forall x)$ (bachelor(x) $\rightarrow \neg$married(x))

At this point, we have two options: (i) maintaining that (2) is true only in the set of possible worlds where (1) holds; (ii) saying that (2) must be true in all possible worlds. In the former case, (1) would be discriminant just at the extensional level and would not relate the intensions of the descriptive constants occurring in its formula; in the latter case, (1) would encode both an extensional and an intensional constraint.

To determine which of the two alternatives is to be preferred, let us proceed *ad absurdum* and assume that the intersection of the extensions of *bachelor* and *married* is not necessarily empty. Accordingly, assume that (1) is genuinely contingent and that the individuals clustered under the extension of *bachelor* fall *de facto*, but not *de jure*, outside the extension of *married*. In this case, we would need to admit that there is at least one possible world, call it *w**, which is extensionally equivalent to our world and differs from our world in this single particular: people speaking English in our world are substituted, in *w**, by people speaking a strange language English*, whose phonology, morphology, syntax and semantics are identical to those of English but have the curious property of allowing for an overlap among the extensions of *bachelor** and *married**. Should we take the *prima facie* conceivability of *w** as a proof of the contingency of meaning postulates, as an argument in favor of (i), and hence as a reason to believe that the English words *bachelor* and *married* can "have the same meaning" that *bachelor** and *married** have in English*? It seems not, for if the extensions of *bachelor* and *married* do not overlap while those of *bachelor** and *married** do, and if the domains of English and English* are duplicates, then there has to be some difference between the extensions of *married* and *married** and between the extensions of *bachelor* and *bachelor**. In the meantime, our semantic theory has remained denotational, with the obvious consequence that we are still moving in a system where the intensional value of lexical items is covariant with their referential interpretation. In light of these premises, the only possible conclusion is that the meaning of *married* is different from the meaning of *married**, and that the meaning of *bachelor* is different from the meaning of *bachelor**. Which means, in turn, that meaning postulates, although possibly arbitrary, cannot be contingent, and that the relativization of language-to-world restrictions pictured by (i) is not a viable option. In a world where *bachelor* and *married* were not extensionally exclusive, they simply would have a meaning different from the one they have in the lexicon of English. To stipulate (1) is thus

to assert that (2) is true in all possible worlds, and the proposition of a meaning postulate is *ipso facto* the proposition of an intensional clause constraining the meaning that words can have in a language.

Once meaning postulates have been introduced, we can finally formulate a more disciplined definition of the notion of "lexical analyticity" and outline a working distinction between semantic truths and logical truths. In a system where descriptive constants were all independent from one another both extensionally and intensionally, it would be impossible to consider a conditional like (2) true in all possible worlds. To prove this, consider (3).

(3) $(\forall x)$ (bachelor(x) $\vee$ $\neg$bachelor(x))

Assume an intuitive notion of "analyticity" and say that the class of analytically true sentences can be defined as comprising the sentences whose validity is a result of logic (e.g., (3)) and the sentences whose validity is a result of semantics (e.g., (2)).[19] In Carnap's own terms, while (3) can be characterized as analytic in compliance with the evidence of its L-truth, if we want to corroborate our intuitions regarding the analyticity of (2), which is not L-true, we need (1). While the analyticity of (3) can be grounded on its L-truth, lacking (1) it is impossible for us to maintain that our intuitions regarding the analyticity of (2) are sound. Once we have (1), we can switch from an intuitive to a well-behaved notion of analyticity in the treatment of sentences whose necessary truth or falsity does not depend exclusively on logical calculus but depends also on semantic interpretation, and we can posit that a sentence S is analytically true either if S is L-true or if the necessary truth of S is entailed by an appropriate meaning postulate.

4. False Translations

One could have the more than understandable impression that Carnap's introduction of meaning postulates was intended to allow meaning analysis to frame in a rigorous and systematic fashion speakers' intuitions about the relations of synonymy, antonymy, hypernymy and hyponymy connecting non-functional words in natural lexica, perhaps in the hope to invest the recovery of this network into a structure-oriented account of the meaning of individual lexemes. After all, the idea that as the web of relations existing among nodes of a network becomes more dense and specified we gain more information about the content of single nodes is intuitively attractive. Plus, meaning postulates seem to offer a consistent

[19] Norris Lance, O'Leary Hawthorne [1997].

advantage over standard versions of componential analysis, since they can be introduced to represent the meaning of lexemes without making any assumption about the conceptual components of the senses that they may aim to analyze.[20] Nonetheless, Carnap insisted that his appeal to meaning postulates was not intended to mirror language use, and that they should be conceived as genuine stipulations rather than as competence-related devices by which to formalize features of the actual semantic relations connecting words in natural lexica.[21] So let us ask: how do we turn meaning postulates into instruments by which to analyze lexical *competence*? One natural way to proceed is to describe meaning postulates as having to do with semantically-based inference[22] and use them to account for speakers' spontaneous ascription of validity to inferences like (∀*x*) ("*x* is a bachelor" → "*x* is not married") or (∀*x*) ("*x* knew that *y* was mistaken" → "*y* was mistaken"). After all, when a speaker S does not recognize the validity of (∀*x*) ("*x* knew that *y* was mistaken" → "*y* was mistaken"), we tend to conclude that S has an inadequate knowledge of the semantic properties of the verb *know* (in this case, S probably ignores that *know* is factive), rather than that S's refutation of the conditional is not problematic just because we are dealing with a semantic machinery whose meaning postulates are not expected to be sensitive to use.

Bearing this agenda in mind, let us take a short detour and consider again the topic of semantic interpretation.[23] Consider a random well-formed declarative sentence in English like (4).

(4) "There is a cat floating over a chair".

Suppose we want to translate (4) into a sentence of a formal language L whose system of symbols and syntax have been specified as including, among others, the following instructions.

(5) (a) K(*x*) = "*x* is a cat";
(b) C(*x*) = "*x* is a chair";
(c) F(*x*, *y*) = "*x* is floating (/ floats) over *y*".

Based on (5), the "translation" of (4) in L should be something close to (6).

(6) (∃*x*)(∃*y*) (K(*x*) & C(*y*) & F(*x*, *y*)).

[20] More on this in Lyons [1995].

[21] Carnap [1956].

[22] The terminology is that of Brandom [1994].

[23] We'll come back to the evaluation of semantic inferences at the beginning of § 5.

Now, it seems there is something substantially misplaced in the idea that (6) can be taken as a proper "translation" of (4), since the symbols of L that (5) pairs with the descriptive constants of (4) are left uninterpreted in (6). On the one hand, we want our translation to be a procedure associating sentences in an source language to sentences in a target language whose rules of interpretation are independent from the rules of interpretation for sentences in the source language; on the other, only the syntax of (6) is autonomously interpretable. It might seem, at least *prima facie*, that (6) could be converted into an autonomously interpretable sentence by processing its descriptive constants on the basis of the instructions provided in (5). But the idea of providing a set of rules of interpretation for (6) by processing its lexical constituents in light of (5) is pointless as well, since our starting goal was to analyze (4) via (6), and by doing so we would simply make the possibility to interpret (6) parasitic on the semantic transparency of (4), whose lack is precisely what we wanted to fix via the introduction of (6). As an overall result, (6) cannot be counted as a translation of (4), given that (6) cannot be interpreted without (5); plus, the only procedure apparently available to turn (6) into an interpreted sentence would be to construe its descriptive constants via (5), a choice that would make the comprehension of (6) depend on the comprehension of (4), which, unhappily, is exactly the sentence we wanted to see analyzed by (6).

The nut is hard to crack, but here is one possible way to go. If we want to transform (6) into a translation of (4), we need to supply some self-standing set of instructions to define the truth conditions for formulas in L, and then specify how the syntactic constructions and the symbols appearing in (6) contribute to its truth conditions. Once these instructions will have been provided, we should be able to give the truth conditions of (6) with respect to a set of functions assigning values to the variables in its formula.[24] Naturally, since the truth conditions of (6) thus obtained will be relative to the array of rules chosen to represent how the syntax and the descriptive constants of (6) contribute to the propositional content of formulas in L, the truth conditions of (6) will be relative to such array of rules and, in turn, to the model that grounds their selection. So ours will be just *an* interpretation of (6). Anyway, *an* interpretation of (6) will be sufficient for our purposes, since what we want is simply to determine what is needed to translate a proposition expressed in English into a formal language like L and take the sentence thus obtained as a "translation" of the source sentence. Let us call our interpretive procedure I and define its domain as D_I. We can now define I by

[24] Chierchia, McConnell-Ginet [2000].

saying that (6) is true with respect to I iff there exist an object u_1 and an object u_2 satisfying the requirements expressed in (7).

(7) (a) $I(K) \subseteq D_I$;
(b) $I(C) \subseteq D_I$;
(c) $u_1 \in D_I$;
(d) $u_2 \in D_I$;
(e) $I(F) \subseteq D_I \times D_I$;[25]
(f) $u_1 \in I(K)$;
(g) $u_2 \in I(C)$;
(h) $(\forall u \in D_I)\ (u \notin I(K) \vee u = u_2)$;
(i) $\langle u_1, u_2 \rangle \in I(F)$.

So far, so good. Now let us ask: is the provision of (7) sufficient to consider (6) an interpreted translation of (4)? More intuitively, is what we understand by processing (6) in light of (7) exactly what we understand by processing (4) as competent speakers of English? The answer seems bound to be negative: again, there seems to be some aspect of the meaning of descriptive constants that eludes truth-theoretic analysis, and as long as we are unable to load (7) with some additional instructions or rules to take care of this shortage of lexical transparency (an incorporation whose feasibility in a referential framework is to be proven), it is impossible for us to take (6) as an interpreted translation of (4). Of course, it is true that in standard semantics we are not required to specify more than the logical type of the symbols appearing in the formula we have provided to model the source sentence. For example, when (7) (a) writes "$I(K) \subseteq D_I$", it is signaled that the extension of the predicate K under I will be a set of individuals in D_I. But since none of the instructions in (7) suffices to determine the referents picked out in D_I by the descriptive constants appearing in (6), and since all that is expressed in (7) about the intensions of such descriptive constants is their membership to a set--theoretic type, there is something conceptually wrong in referring to the translation provided by (6)–(7) for (4) as something justifiably definable as a "translation".

Two supplementary reasons can be given in favor of this remark. The first has to do with the fact that the "translation" of (4) provided via (6) does not select a specific model, in Montague's sense.[26] Rather than being an interpretation in a model, (6) looks like a rule to define interpretations in models and, hence,

[25] I.e., I(F) is a subset of the set of all the ordered pairs we can build in D_I.

[26] Montague [1974].

something very close to a function from models to interpretations of (4). The selection of the model mirroring the interpretation of (4) preferred by speakers of English has to depend on the introduction of further constraints (for example, eligibility of content and compliance with use),[27] that the instructions listed in (7) are unable to deliver. The second, more intuitive, is based on the observation that spelling out the content of (6) via (7) is insufficient to reproduce the informational grasp entertained by competent speakers of English in the comprehension of (4). In other words, it is impossible to determine on the only basis of (6) and (7) what competent speakers of English believe when they evaluate (4) as a true sentence, which implies (given the syntactic transparency of (6)) that the knowledge of the meaning of *cat*, *float over* and *chair* cannot be reduced to information about the logical type of predicates in a formal language.

One could object that this line of argument breaks into an open door and overstates a trivial methodological divide, since model-theoretic analysis is designed to frame only the phrasal effects of composition: no wonder that it is unable to shed light on the individuation of lexical meaning, which functions as a raw input to compositional processes. But it is fair to say there is something more to this. First, some have observed that the Tarskian apparatus needs to be understood in a particular way to make it deliver a theory of meaning. For example, Tarski's definition of the truth predicate assumes a quasi-mathematical notion of sentential truth (where whether or not a sentence is true is viewed, in essence, as a matter of logic), but since the truth of natural language sentences is mostly a contingent matter, his approach to the truth predicate has been deemed unsuitable to give a theory of sentential meaning.[28] Granted, the recursive characterization of truth offered in a Tarskian apparatus *can* be limited to the description of truth conditions and thereby used to express some semantic properties of sentences and of their lexical constituents. The problem is that what we obtain in such cases is a procedure that generates descriptions of the truth conditions of sentences while being unable to return a valuation of their truth,[29] and that something very similar seems to obtain when it comes to analyzing lexemes: what we have is a theory illuminating what it means to be a (syntactic) type of descriptive constant, but not a theory analyzing the semantic content thanks to which tokens of descriptive constants single out referents in language use. So, we have a theory that seems to respond to the same explanatory

[27] Lewis [1983], Sider [2001].

[28] E.g., Soames [1984]; Etchemendy [1988].

[29] Heck [1997].

requirements we find in theories of truth *lato sensu*, which we want to tell us something about what it means for a proposition *p* to be true, but not to provide us with the extension of the set comprising all the *p*-like items that actually are true.[30] Overall, something we can hardly characterize as a semantic theory.

Second, if we take the equivocal and all-encompassing label "theory of meaning" and try to transform it into something more accurate, there are at least *three* specific kinds of theories of meaning.[31] (i) The first, that we may dub a *metaphysics of meaning*, is a theory interested in accounting for "what it means to mean", for what properties must be satisfied in order for a physical token to count as a meaningful symbol or a meaningful expression, and for what meanings are in the most general sense. (ii) The second, that we may call a *semantic theory*, is a theory interested in pairing (classes of) meanings with (classes of) expressions in a language and, hence, in building a formal system capable of predicting the interpretation of expressions in that language. (iii) The third, that we may label a *foundational theory of meaning*, is a theory interested in determining the extralinguistic facts in virtue of which the expressions of a language have the meaning they have (e.g., Kripke's causal theory of reference). What is the ecological niche in this trinomial of the type of meaning analysis we have been discussing so far? The expected answer is that it should be a (ii)-type theory. But is it? It seems not, since when the meaning of non-functional words comes into play, the only thing a denotational system is able to do is to relate syntactic types with types of extensions, but not to relate syntactic tokens with tokens of extensions. (7)-like instructions may well be an element of the account of the interpretation of (4) that we would expect to see provided by a semantic theory defined along the lines fixed in (ii), but it just is not that account. This is not to reach the absurd conclusion that speakers of a language, in order to count as competent users of a given content word, must always be able to represent its extension (and determine its reference in contexts of use). This is clearly not the case when we deal with the valuation of vague predicates such as *bald* or of highly polysemous verbs such as *take*.[32] But it is to say that once we allow that the meaning of lexemes must be analyzed in truth-theoretic terms, it is impossible for us to do justice to the minimal facts a theory of word meaning should be expected to explain. The reason why I point this out is not that truth-theoretic analysis cannot capture features such as emotive meaning, metaphoricity, or semantic change

[30] E.g., Kirkham [1995], Schmitt [2003].

[31] Compare with Lewis [1970] and Speaks [2010].

[32] E.g., Williamson [1994], Sorensen [2001], Rescher [2009].

(which one could doubt should be taken into account by a semantic theory in the first place),[33] but that once the externalist commitment underlying our formalist machinery is forced to the point of claiming that extensions are intrinsic to meaning, word meaning analysis should be systematically sufficient to predict word reference. Yet, this result seems to fiercely resist our efforts. Interestingly, even in frameworks couched in the apparatus of formal semantics and with explicit computational ambitions such as Pustejovsky's generative lexicon[34] or Asher's two-level approach to lexical content,[35] referential grounding is reinterpreted as an interface phenomenon arising from the interaction between the lexicon and cognition in the broader sense: in Chomskyan terms,[36] as something speakers do with words via their meaning rather than as something word meanings do by themselves.

This being the problem, we have two essential ways out. (i) We may propose that our system's provisional inability to formulate explicit predictions about the referential grounding of non-functional words can be fixed by loading our analytic apparatus with formulas of inter-word connection of the type provided by meaning postulates. Which is to say, once our semantic machinery will be able to represent in a complete fashion the inferential competence of a proficient speaker of English, it will *ipso facto* acquire referential competence. (ii) Alternatively, we may maintain that referential grounding cannot be in any sense or degree parasitic on inferential information, for even if we agreed that the provision of information constraining the validity of semantically-based inferences would allow us to build a working model of the relational side of our lexical competence, still such integration would not allow us to obtain the association of interpreted predicates to subsets of the domain containing objects whose lexical labeling is not given *a priori*. We need to evaluate these two possibilities in detail.

5. A Referential Conundrum

Suppose we take the interpretation of (4) we have provided via (6) and we load it with a set of postulate-like clauses covering all the inferential relations existing among non-functional entries of the lexicon of English. This integration should allow our machinery to predict the inferences validated by a competent

[33] See Seuren [2009].

[34] Pustejovsky [1995].

[35] Asher [2011].

[36] Chomsky [2000]. More in Pietroski [2005].

speaker of English in result of her command of the semantic potential of the descriptive constants occurring in (4), such as those provided in (8).

(8) (a) "There is a cat floating over a chair";
(b) "(a) → There is an *animal* floating over a chair";
(c) "(a) → There is a cat floating over *an object*";
(d) "(a) → There is a cat *moving* over a chair".

What we need to ask is whether this integration to our theory will be sufficient to make it represent all that is involved in the interpretation of (4) by a competent speaker of English. Now, (8)-like additions do not tell us anything complete about the meaning of individual words, but they unquestionably increase our level of awareness about what speakers know at the lexical level about cats, chairs, and so forth. Suppose therefore we have a semantic machinery where all the set-theoretic relations among the extensions of non-logical words are specified according to (8)-like clauses. Will the provision of an exhaustive set of (8)-like clauses be sufficient to ground the lexical items occurring in (4)?

This does not seem to be the case. Our now-inferentially-competent machinery (name it ICM) will predict that the linguistic behavior of competent speakers is such that their treatment of the lexical items occurring in (4) regularly singles out extensions among which a known web of set-theoretic relations must obtain. But suppose we interface ICM with a domain and we ask it to distribute the objects of the domain among the extensions of the lexical types it is acquainted with. For instance, suppose that ICM is a computer program that we have inserted in a machine having the ability to access objects in an artificial domain exactly the same way our cognitive and perceptual apparatus allows us to build modes of presentation for worldly objects.[37] Will ICM be able to single out the extensions of the lexical items of a sentence on the basis of an (8)-like knowledge? Again, the answer to this question must be negative.[38] To obtain this result, we would need at least to make ICM competent in recognition tasks and integrate it with an apparatus capable of performing feature extraction and object identification, then interface its recognition abilities with some non-linguistic component contrasting the instruction for meaning assignation stored in its LTM-words with the modes of presentation assembled by perceptual outputs.[39] Even intuitively, an uninterpreted symbol cannot become interpreted by simply connecting it to more

[37] See Violi [2001].

[38] See Partee [1980] and Devitt [1981] on the "language-to-world grounding" problem.

[39] This point has been largely corroborated by recent research in artificial intelligence, such as Steels [2012] and Steels, Hild [2012], which assume a prototype-theoretic model of mental words.

uninterpreted symbols. If you have no notion about the lexicon of German, accessing a complete set of postulate-like clauses expressed in German for German lexical items is not going to make you able to use of those words in a referentially competent fashion. It would be like trying to learn the lexicon of German by studying a German-German dictionary without having any prior knowledge of German.[40]

To further articulate this point, take the lexicon of a natural language N and imagine that ICM has a punctuated competence of type $[\![cat]\!] = [\lambda x\text{: } x \in D \text{ . } true$ if x is a cat and *false* otherwise] on all the content words of N. ICM can compute only the languages of first-order logic and set theory and is only able to associate the entries of N's lexicon with recursively defined extension-types. Suppose we grant ICM the access to a complete set of postulate-like clauses singling out all the set-theoretic relations expected among the extensions of non-functional words of N. Will ICM be a competent user of N's lexicon? No, because it will still be unable to pair specific subsets of the domain with N's lexical entries. Now suppose that the lexicon of N is maximally integrated from an inferential point of view and, hence, that the postulate-like clauses we have inserted in the system are sufficient to relate any non-functional word w_0 of N to any other non-functional word w_n of N by predicting the set-theoretic relation existing between their extensions. Will ICM be a competent user of N's lexicon now that the inferential integration of the lexicon of N matches that of an holistic system? Well, it is difficult to see how the addition of maximum inferential integration to the object language could make a difference in allowing ICM to replicate the referential proficiency displayed by competent speakers of N, if *all* the expressions of its lexicon have not been already interpreted.

To prove this, say that N has an inferentially continuous lexicon with just *one* interpreted entry w_0 and that the entire set of postulate-like clauses constraining the relations between the extensions of words of N are known to ICM. Say also that w_0 is a unary predicate of N and that we want to assess, given these premises, whether ICM will be able to predict the interpretation of all the unary predicates in that lexicon. Unfortunately, this won't do. The output we may obtain at the end of the process is a specific extension for w_0, plus a range of expectations about the set-theoretic relations between the reference we associated to w_0 and the reference of other unary predicates of N, which means a representation of the possible interpretations of the unary predicates of N provided the interpretation that we have fixed for w_0, but not the recognition

[40] Harnad [1989].

of a single model of N satisfying these requirements. In other words, once the extension of w_0 and the array of set-theoretic relations that must obtain between the interpretation of w_0 and that of all other descriptive constants of N have been fixed, there may be more than one global function assigning extensions to descriptive constants that satisfy the extensional constraints between the interpretation of w_0 and that of all other non-functional words of N.

On a broader note, it is worth noting that the same conclusion can be derived by pointing out that inference-based accounts cannot yield a complete analysis of lexical meaning because inferential interdependence is not strictly meaning constitutive, in the following sense. Suppose that Z is the set of valid sentences of a language N on which the meaning of a word w_0 depends, or from which it can be inferred. Suppose also that K is the set of non-functional words occurring in members of Z and that $w_0 \notin$ K. Add a logico-syntactic filter to exclude from K the set of non-functional words occurring in parts of the members of Z which w_0 is clearly not dependent on (e.g., if w_0 is *bachelor*, Z will contain "Bachelors are unmarried men and lovers are lovers" but *lover* will remain outside K, though there may be other sentences of Z requiring the inclusion of *lover* in K). Now take a random element k_i of K. The semantic value of k_i will in turn depend on the set K* of descriptive constants occurring in some other set Z* of valid sentences of N where k_i occurs, some of which will be not included in Z. Since the meaning of w depends on k_i and k_i depends in turn on the meaning on K*, it seems natural to conclude that w_0 depends on K* as well, and that if we appropriately reiterate this line of reasoning, what we obtain is that the semantics of w_0 depends on (or can be properly inferred from) all the true sentences of N. The argument seems plausible, but rests on a subtle fallacy. Suppose again that ICM is competent on w_0 iff it has access to Z and has a notion of their truth. The argument states that it makes sense to assume that the command of Z is required for competence on w_0, *and* that if competence on w_0 requires competence on k_i, competence on any sentence where both w_0 and k_i occur is a proper constituent of competence on w_0 iff ICM is inferentially competent on k_i and, hence, on K*. Now suppose that w_0 is *cat* and k_i is *animal*, and take $(\forall x)$ (cat(x) → animal(x)). What does it mean to say that competence on *cat* is constituted by ICM's access to $(\forall x)$ (cat(x) → animal(x))? Here lies the problem, for as long as we can only say that competence on *cat* can be inferred from access to $(\forall x)$ (cat(x) → animal(x)) and not from the access to the meaning of *animal*, we are implicitly excluding the command of the meaning of *animal* from the variables that constrain the meaning of *cat*. Upon closer examination, the inferential competence of ICM would be nothing but a database representing relations among phonographic units, *not*

among meanings. In the case of natural languages, by contrast, it seems that the knowledge of the validity of $(\forall x)$ (cat(x) → animal(x)) is part of our competence on *cat* and *animal* only because we have an independent and primitive notion of the meaning of *cat* and *animal*.

Granted, representing lexical knowledge in terms of intralinguistic, inference-based connections among words provides us with a working instrument to frame important aspects of the semantics of lexemes, but if the observations presented in this paper are correct, inference-based meaning analysis is bound to remain silent on equally fundamental questions about word meaning. A rough list can include: (i) Are word meanings atomic primitives, Fodorian concepts, lists of prototypical features, instructions to fetch representations, rules to token bits of our encyclopedic knowledge, or what else? (ii) Do word meanings contain features resisting their representation in terms of bases for semantic inferencing? (iii) If the inferential properties of lexemes are insufficient to account for their referential grounding, should we go internalist and exclude reference from the core preoccupations of semantic theorizing? (iv) Should the analytic apparatus of semantic theorizing incorporate insights on the nature of lexical meaning coming from approaches to language different from the Tarski-Montague lineage that originated formal semantics (e.g., Wittgenstein-Austinian claims about meaning as use, Putnam's division of linguistic labor, Bartsch's work on linguistic norms)? (v) How far can standard truth-theoretic semantics be pushed in dealing with the phenomena of fuzziness, underspecification, polysemy and vagueness so abundantly emphasized by cognitively-oriented research on lexical meaning in the last couple of decades?

Many of these questions are currently under discussion. Overall, the analysis of lexical items seems to require an appeal to facts (e.g., concepts, mental structures, cognitive interfaces) extending far beyond what an externalist, broadly referential approach to semantic content is capable of accounting for. To my best guess, the way this need will be reconciled or deemed incompatible with standard truth-conditional semantics will have a significant impact on both its format and its credit as a productive framework for the study of natural languages.[41]

[41] This paper originated from the handout of a presentation I gave at Institut Jean Nicod (ENS, Paris) in April 2013. My gratitude goes to François Recanati for inviting me to Paris and to the audience that attended the talk: their questions and observations greatly helped me strengthen my arguments and expand the notes I had prepared for the presentation into a proper article. I also thank an anonymous referee for pressing me to address more explicitly a couple of important points I had unjustly taken for granted.

References

Asher [2011] – N. Asher, *Lexical Meaning in Context. A Web of Words*, Cambridge University Press, Cambridge 2011.

Brandom [1994] – R.B. Brandom, *Making It Explicit*, Harvard University Press, Cambridge, MA 1994.

Carnap [1956] – R. Carnap, *Meaning and Necessity*, enlarged ed., Chicago University Press, Chicago, IL 1956.

Casalegno [1997] – P. Casalegno, *Filosofia del Linguaggio: Un'Introduzione*, Carocci, Rome 1997.

Chierchia, McConnell-Ginet [2000] – G. Chierchia, S. McConnell-Ginet, *Meaning and Grammar. An Introduction to Semantics*, 2nd edn., MIT Press, Cambridge, MA 2000.

Chomsky [2000] – N. Chomsky, *New Horizons in the Study of Language and Mind*, Cambridge University Press, Cambridge 2000.

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

Devitt [1981] – M. Devitt, *Designation*, Columbia University Press, New York, NY 1981.

Etchemendy [1988] – J. Etchemendy, „Tarski on Truth and Logical Consequence," *Journal of Philosophical Logic* (43) 1988, pp. 51–79.

Frege [1979] – G. Frege, *Posthumous Writings*, Blackwell, Oxford 1979.

Frege [1980] – G. Frege, *Translations from the Philosophical Writings of Gottlob Frege*, ed. P. Geach, M. Black, 3rd edn., Blackwell, Oxford 1980.

Geeraerts [2010] – D. Geeraerts, *Theories of Lexical Semantics*, Oxford University Press, Oxford 2010.

Harnad [1989] – S. Harnad, „Minds, Machines, and Searle," *Journal of Theoretical and Experimental Artificial Intelligence* (1) 1989, pp. 5–25.

Heck [1997] – R. Heck, „Tarski, Truth, and Semantics," *Philosophical Review* (106) 1997, pp. 533–554.

Heim, Kratzer [1998] – I. Heim, A. Kratzer, *Semantics in Generative Grammar*, Blackwell, Oxford 1998.

Juhl, Loomis [2009] – C. Juhl, E. Loomis, *Analyticity*, Routledge, London 2009.

Kirkham [1995] – R.L. Kirkham, *Theories of Truth. A Critical Introduction*, MIT Press, Cambridge, MA 1995.

Lewis [1970] – D. Lewis, „General Semantics," *Synthese* (22) 1970, pp. 18–67.

Lewis [1983] – D. Lewis, „New Work for a Theory of Universals," *Australasian Journal of Philosophy* (61) 1983, pp. 343–377.

Lycan [2008] – W.G. Lycan, *Philosophy of Language. A Contemporary Introduction*, 2nd edn., Routledge, London 2008.

Lyons [1995] – J. Lyons, *Linguistic Semantics*, Cambridge University Press, Cambridge 1995.

Malpas [2009] – J. Malpas, *Donald Davidson*, [in:] *The Stanford Encyclopedia of Philosophy (Fall 2009 Edition)*, ed. Edward N. Zalta, URL = http://plato.stanford.edu/archives/fall2009/entries/davidson/.

Marconi [1989] – D. Marconi, *Rappresentare il Significato Lessicale*, [in:] *Mente Umana, Mente Artificiale*, ed. R. Viale, Feltrinelli, Milan 1989.

Marconi [1997] – D. Marconi, *Lexical Competence*, MIT Press, Cambridge, MA 1997.

McDowell [1998] – J.H. McDowell, *Meaning, Knowledge, and Reality*, Harvard University Press, Cambridge, MA 1998.

Montague [1974] – R. Montague, *Formal Philosophy: Selected Papers of Richard Montague*, ed. R.H. Thomason, Yale University Press, New Haven, CT 1974.

Norris Lance, O'Leary Hawthorne [1997] – M. Norris Lance, J. O'Leary-Hawthorne, *The Grammar of Meaning. Normativity and Semantic Discourse*, Cambridge University Press, Cambridge 1997.

Partee [1976] – B. Partee, *Montague Grammar*, Academic Press, New York, NY 1976.

Partee [1980] – B. Partee, *Montague Grammar, Mental Representations and Reality*, [in:] S. Oehman, S. Kanger (eds.), *Philosophy and Grammar*, Reidel, Dordrecht 1980.

Pietroski [2005] – P. Pietroski, *Meaning Before Truth*, [in:] *Contextualism in Philosophy*, G. Preyer, G. Peters (eds), Oxford University Press, Oxford 2005.

Pustejovsky [1995] – J. Pustejovsky, *The Generative Lexicon*, MIT Press, Cambridge, MA 1995.

Quine [1952] – W.V.O. Quine, *Notes of Existence and Necessity*, [in:] *Semantics and the Philosophy of Language*, ed. L. Linsky, University of Illinois Press, Urbana, IL 1952.

Rescher [2009] – N. Rescher, *Unknowability. An Inquiry into the Limits of Knowledge*, Lexington Books, Lanham, MD 2009.

Schmitt [2003] – F.F. Schmitt, *Theories of Truth*, Blackwell, London 2003.

Seuren [2009] – P.A.M. Seuren, *Language in Cognition*, Oxford University Press, Oxford 2009.

Sider [2001] – T. Sider, *Four-Dimensionalism: An Ontology of Persistence and Time*, Clarendon Press, Oxford 2001.

Soames [1984] – S. Soames, „What Is a Theory of Truth?," *Journal of Philosophy* (81) 1984, pp. 411–429.

Soames [1999] – S. Soames, *Understanding Truth*, Oxford University Press, Oxford 1999.

Sorensen [2001] – R.A. Sorensen, *Vagueness and Contradiction*, Oxford University Press, Oxford 2001.

Speaks [2010] – J. Speaks, *Theories of Meaning*, [in:] *The Stanford Encyclopedia of Philosophy (Spring 2010 Edition)*, ed. Edward N. Zalta, URL = http://plato.stanford.edu/archives/spr2010/entries/meaning/.

Steels [2012] – L. Steels (ed.), *Experiments in Cultural Language Evolution*, John Benjamins, Amsterdam 2012.

Steels, Hild [2012] – L. Steels, M. Hild (eds.), *Language Grounding in Robots*, Springer, New York, NY 2012.

Tarski [1944] – A. Tarski, „The Semantic Conception of Truth," *Philosophy and Phenomenological Research* (4) 1944, pp. 341–376.

Tarski [1983] – A. Tarski, *Logic, Semantics, Metamathematics, Papers from 1923 to 1938*, ed. J. Corcoran, Hackett Publishing Company, Indianapolis, IN 1983.

Violi [2001] – P. Violi, *Meaning and Experience*, Indiana University Press, Bloomington, IN 2001.

Williamson [1994] – T. Williamson, *Vagueness*, Routledge, London 1994.

Wittgenstein [1922] – L. Wittgenstein, *Tractatus Logico-Philosophicus*, Routledge and Kegan Paul, London 1922.

Zimmerman [1999] – T.E. Zimmerman, „Meaning Postulates and the Model-Theoretic Approach to Natural Language Semantics," *Linguistics and Philosophy* (22) 1999, pp. 529–561.