Paraphrase is not enough

CHRIS BARKER

In 1972, the publication of Wierzbicka's *Semantic Primitives* initiated the Natural Semantic Metalanguage approach to linguistic semantics. NSM sets out to analyze the meaning of natural language utterances using only paraphrases built from a set of concepts called semantic primes. The main hypothesis is that a definite small set of unanalyzable primes that occur in every natural language suffices to characterize all natural meaning.

In that same year (1972), the first publication of Kripke's *Naming and Necessity* initiated an extraordinarily influential discussion in the philosophy of language. Kripke, Putnam, and others argued that the meaning of proper names (e.g., *Socrates* or *Hesperus*, etc.) cannot be adequately paraphrased by any description or set of descriptions. They extended these arguments to natural kind terms (*water* or *tiger*, etc.), and tentatively to other types of predicates. Since proper names and natural kind terms carry a significant portion of the burden of meaning in every natural language, these arguments strongly suggest that the NSM project cannot provide a complete account of natural language meaning even in principle.

In addition to the non-paraphrasability of names and natural kinds, I identify below a separate class of challenges for NSM. Research over the last thirty years has provided abundant evidence that the meanings of words and expressions are exquisitely sensitive to the structure of the sentences in which they are embedded. Because NSM strictly limits the way that paraphrases are able to talk about the linguistic structure undergoing paraphrase, it cannot provide adequate accounts of such fundamental semantic phenomena as indexicality, performatives, or presupposition, unless the concepts used for paraphrases are non-basic, i.e., systematically decomposable into more basic semantic elements.

Theoretical Linguistics 29 (2003), 201–209

0301–4428/03/0029–0201

1. Names are not paraphrasable

Russell argued that proper names could be paraphrased by a description, so that the name *Socrates*, for instance, might mean 'the Greek philosopher who drank hemlock'. But the description theory of names is vulnerable to error: what if historical research reveals that there was no hemlock available in Greece at that time? Clearly, we were mistaken: Socrates must have drunk some other poison. But if the name Socrates is defined in part by the property of drinking hemlock, we are forced to conclude that the name refers to some other individual. [My presentation in sections 1 and 2 is indebted to Burgess (to appear).]

So Searle proposed that names take for their meaning a cluster of properties, so that Socrates might be whichever unique person fulfilled most of the following requirements: was Greek, was famous, drank poison, drank hemlock. Proving that the person indicated could not have drunk hemlock does not prevent him from being Socrates, since he still satisfied most of the other descriptive properties.

But now we confront the problem of ignorance: imagine a student is aware of virtually none of the cluster of properties except perhaps that Socrates was a philosopher, and yet he can still use the name successfully and with its usual meaning. Apparently speakers can borrow the credentials of more knowledgeable speakers: because the student has observed his professor discussing Socrates, that is enough to allow him to use the name too.

Kripke sharpened and extended the arguments from error and from ignorance, and added an argument of his own, claiming that names are rigid designators: they refer to the same individual in every possible world (if they refer at all). If the meaning of a name were to depend crucially on some description or cluster of descriptions, they would not be rigid. Imagine that we paraphrase *Hesperus* as 'the brightest object in the morning sky'. Given the way things are, Hesperus and Venus name the same object. But what if Mars comes so close to Earth that it outshines even Venus? Under the description theory, *Hesperus* would come to refer to Mars; yet we have a strong intuition that even in the counterfactual situation, *Hesperus* would continue to refer to Venus.

In general, no matter how numerous or complex the set of properties, it would always be possible to imagine a situation in which the person named

Socrates or the planet named Hesperus would fail to have every single one of those properties, and yet we still intuitively feel that the names 'Socrates' and 'Hesperus' would continue to refer to the same objects they did before. This is Kripke's modal argument.

Instead, suggests Kripke, names receive their meaning via historical transmission. There is a christening event, perhaps involving a description or an act of ostention, in which a name is associated with an object; and other people come to associate that name with the object in question through a causal chain leading back to the original event.

Yet another argument against a description theory of names can be constructed on the basis of Putnam (1973). Imagine a world virtually identical to Earth called Twin-Earth. Everything that happens on Earth happens on Twin-Earth, right down to the exact fluttering of each leaf. On Twin-Earth, there is a person just like me typing the same words I am now typing. This person has all of my qualitative properties: age, height, weight – he even has a double-ganger (me!) on the planet that he calls Twin-Earth (and that you and I call Earth). If the name 'Rex' has for its meaning (the properties denoted by) a linguistic description of my dog, then there is a creature on Twin-Earth that possesses each one of those properties. So when I say "I walked Rex this morning", there are two creatures that satisfy the description embodied by the name 'Rex'. Either my use of the name fails (surely not the right result), or else I am in danger of accidentally referring to twin-Rex. But that surely is not correct: when I say 'Rex', I mean my dog here on the same Earth with me, regardless of the existence of another creature with identical properties. (Salmon (1986:66–7) develops this kind of example considerably more carefully than I have done here.)

The arguments from error, ignorance, rigid designation, and Twin-Earth strongly suggest that at least some words have meanings that cannot be adequately expressed by any linguistic paraphrase.

2. Natural kind terms are not paraphrasable either

It would be bad enough for NSM if proper names were the only type of expression whose meaning could not be paraphrased. But Kripke and Putnam show that natural kind terms such as *tiger* or *water* behave

similarly to names with respect to error, ignorance, rigid designation, and Twin-Earth. For instance, imagine some paraphrase that attempts to capture the essence of tigerness, perhaps along the lines proposed by Wierzbicka (Durst's example (16)): the fact that it is an animal, its habitat (the veldt), its size, its appearance (has stripes), behavior (hunts, eats meat), etc. Now imagine some marsupial species comes to fill the ecological niche normally occupied by tigers, and through convergent evolution comes to have all of the properties that we used to attempt to characterize tigers. Would these creatures, however similar, be tigers? Clearly not.

No matter how exhaustive the paraphrase, there will always be a counterfactual thought experiment that reveals that natural kind terms acquire their meaning via baptism and historical transmission. One way of putting it is that some common nouns are names, that is, names of kinds, and, like proper names, immune to paraphrase.

Furthermore, these arguments are independent of the tension between scientific definitions of natural kinds (i.e., having a certain genetic code or having a certain chemical composition) and folk conceptions. Kripke's reasoning is entirely consistent with giving priority to the naive conception of tigers or water. When scientists investigated the chemical properties of water, they merely discovered that water was H2O, they did not discover what water actually was—they already knew what water was, just as any two-year old or a chemically-ignorant adult is fully competent in their knowledge of the meaning of *water*.

Wierzbicka (1972:21) herself suggests that names and natural kind terms cannot be explicated directly or indirectly in terms of the semantic primes. She 'tentatively' offers the following partial characterizations:

> The man called John = the man thinking of whom we say 'John'
> a cat = an animal thinking of which one would say 'cat'

Note that the words *John* and *cat* are part of the paraphrases here, even though they are not semantic primes. Clearly such definitions violate two central precepts of the modern NSM project as presented by Durst: that the semantic primes alone are sufficient to explicate all utterances (part of indispensability, his section 2.3); and that paraphrases should not be circular (no circular definitions, section 3.2).

Later work in NSM seems to be somewhat more optimistic, at least about the status of natural kinds (see Durst's discussion and citations in section 4.1).[1] For instance, Wierzbicka (1985:163) explicitly rejects her earlier position and asserts that 'folk genera' are in fact reducible. She then proceeds to offer qualitative property-based definitions of *tiger* and a number of other natural kind terms, though she does not address directly any of the Kripke/Putnam criticisms of the descriptivist approach.

Wierzbicka's definition of tiger is fairly lengthy. Kripke (p. 327) speculates that his anti-descriptivist arguments apply to other sorts of predicates, including color terms; therefore let us consider as an example Wierzbicka's paraphrase for *green* as given by Durst:

X is green = in some places many things grow out of the ground
when one sees things like X one can think of this

Even assuming we are provided with suitable paraphrases for *in, grow, out of,* and *ground* (none of which are primes), this paraphrase is deeply unsatisfying as a characterization of the meaning of the word *green*. Like a turn at charades, the best we can hope for is to guide the attention of the listener towards the intended concept, and trust they will guess the right one. Note that in Kripke's terms, the way in which the paraphrase attempts to do this is not by characterizing the property itself, but by evoking a situation in which the concept would be salient. In this case, there are a number of equally valid solutions to the riddle, including "brown", "trees", and "grass". In any case, the paraphrase is certainly not criterial, much less an explanation of the nature of "green".

It seems to me that Kripke, Putnam, and Wierzbicka (1972) got it right: names and natural kinds cannot be adequately reduced to paraphrases. If so, reductive paraphrase using only a fixed set of semantic primes cannot be a complete theory of meaning, even in principle.

---

[1] To be sure, there are also a few dissenting voices in the philosophical literature. For instance, Jackson (1998) sketches some possible strategies for defending a description theory of names and natural kinds. Yet even on Jackson's view, the descriptions are not required to be the meanings of any linguistic paraphrase, in violation of the fundamental assumptions of NSM.

3. Meanings interact with linguistic structure

I am sympathetic to the claim that a theory of meaning (or of anything) cannot truly be understood unless it can be paraphrased in simple language. Thus it makes sense to avoid technical terms in favor of their paraphrases. Yet when NSM rejects analyzing I and YOU as the first and second person singular pronouns, it cannot be because these technical terms are too difficult to be paraphrased. After all, NSM has no difficulty paraphrasing far more subtle concepts.

Rather, it can only be because these terms explicitly refer to linguistic properties of the expression to be defined. But a wide variety of fundamental semantic phenomena clearly show that the meanings of words interact intimately with the syntactic and semantic structure of the utterances in which they occur. The inability of NSM paraphrases to talk about the linguistic structure of the expression under explanation is a severe shortcoming, given the stated goals of NSM. I will briefly discuss three examples, though I have no doubt that more could be found.

3.1. *Indexicals*

Perhaps the most obvious example of this pernicious lack of expressive power is the treatment of indexicals. As Durst explains, if concepts such as first person and second person are unavailable, there is no choice but to accept I and YOU as distinct primes. But this is glaringly inadequate, since, as Durst himself admits, they share a large semantic overlap, including the concept of PERSON, not to mention the fact that both are entailed to participate in the speech event.

As a result, indexicals represent a significant number of the proposed primes, including I, YOU, NOW, THIS, and HERE. If it were possible to talk about the containing utterance in the paraphrase, these could be assimilated to PERSON, TIME, THING, and PLACE. For instance, the kind of paraphrase I have in mind for I (following Kaplan) might be

I = the speaker of the utterance in which this word occurs

Similar paraphrases are available for the other indexicals. Thus in its current state, NSM contains several pairs of primes that systematically

share a significant semantic overlap; hence the primes clearly can and should be further reduced to more basic elements in search of deeper semantic insight.

3.2. *Presupposition*

As a second, more subtle example of how the meaning of a lexical item can interact with its linguistic environment, consider the presuppositions of verbs such as *realize* or *managed. Realize* differs from *believe* in large part by virtue of presupposing the truth of its complement. Thus it is inappropriate to utter *John realized that he was sick* unless the speaker believes that John is sick. In contrast, *John believed that he was sick* can be true even when the speaker knows that John actually isn't sick.

In order to distinguish *realize* from *believe*, it is not enough to add a clause such as 'I think X is true' (where X is the content of the complement) to the paraphrase of *realize* but not *believe*. That would incorrectly predict that *John didn't realize that he was sick* ought to be acceptable even if the speaker doesn't think that John is sick. After all, the condition 'I think that X is true' would be under the scope of the negation. But in fact it is only acceptable to say *John didn't realize that he was sick* if the speaker believes that John *is* sick. Indeed, this is the hallmark of presuppositions: their entailments survive even when embedded under negation.

Clearly, presuppositional meaning behaves entirely differently compared with non-presuppositional meaning. There is no way to express presuppositional meaning in a theory that allows paraphrase as the only means of representing meaning, unless perhaps the semantic primes include a suitable range of presupposition triggers. On analogy with the treatment of indexicals, then, presumably NSM will need to add multiple lexical items that differ from existing primes only in that they trigger presuppositions: REALIZE in addition to THINK, POSSESS in addition to HAVE (since the possessive in *My uncle is bald* presupposes that I have an uncle), MANAGE in addition to DO – enough primes to characterize all the myriad lexical items and syntactic constructions that trigger presuppositions.

Just as in the case of indexicality, adding multiple primes would obscure an important underlying semantic regularity, resulting in a set of

supposedly unanalyzable concepts that in fact share a significant common semantic component.

3.3. *Performatives*

Performatives provide a third kind of lexical meaning that cannot be captured in NSM without explicitly recognizing the structure of the utterance in which a word is embedded. Roughly, performatives are utterances that do things in the world. Christenings, marriages, promotions, and promises are all actions that can be accomplished by uttering expressions of a certain form under the appropriate conditions. In her study of speech acts, Wierzbicka recognizes this problem and offers (1987:23) this approximation: 'I assume that by saying this in this way I can (or will) cause it to happen'. But this is clearly insufficient. If I say *I promise to come*, not only have I assumed something, a contract has been created. No matter what the speaker may or may not assume, no contract is created merely by uttering *I promised to come* or *Bill promises to come*. The difference in potential perlocutionary force among these expressions is in part a function of the lexical meaning of the verb in question and in part a function of the linguistic forms of the expression in which that verb is embedded (in particular, first versus second person, and present versus past tense). I see no way to provide an adequate account of the differences in meaning without adding a series of performatives to the set of primes, once again creating a systematic overlap in meaning among the primes.

4. Summary

The arguments of Kripke and others present a grave challenge for any attempt to render the meaning of proper names and natural kind terms by means of paraphrase. If those arguments are correct, then NSM is necessarily incomplete as a theory of meaning.

Furthermore, I have argued that in any theory that limits itself exclusively to paraphrase, indexicality, presupposition, and performatives each induce a systematic partial duplication of certain concepts. If so, then NSM fails to provide a set of irreducible primes, since multiple pairs of primes share manifestly analyzable semantic sub-components.

Therefore, despite the undeniable intuitive appeal of NSM, and despite its admirable and lastingly valuable cross-linguistic investigations of lexical meaning, I am forced to conclude that the basic hypothesis of NSM simply isn't viable.

Thanks to David Braun, Svetlana Godjevac, Chris Potts, and Brian Weatherson.

University of California at San Diego

References

Burgess, John P. (to appear) Saul Kripke, Naming and Necessity. To appear in Acumen Publishing in Volume 5 of the series Central Works of Philosophy, edited by J. Shand.

Jackson, Frank. 1998. Reference and Description Revisited. *Philosophical Perspectives* 12:201–218.

Kripke, Saul. 1972. Naming and Necessity. In Davidson, Donald and Harman, Gilbert (eds). *Semantics of Natural Language*, Reidel, Dordrecht, 253–355, 763–769. Revised version published in 1980 by Blackwell in Oxford and by Harvard University Press in Cambridge, MA.

Putnam, Hillary. 1973. Meaning and Reference. *Journal of Philosophy* 70:699–711.

Salmon, Nathan. 1986. *Frege's Puzzle*. Bradford, Cambridge, MA.