NICK RIEMER

REDUCTIVE PARAPHRASE AND MEANING: A CRITIQUE OF WIERZBICKIAN SEMANTICS*

ABSTRACT. This article explores some fundamental issues of definition-based lexical semantic research through a critique of the Natural Semantic Metalanguage (NSM) theory of semantic and grammatical description (Wierzbicka 1996, *Semantics. Primes and Universals*, Oxford University Press, Oxford). NSM is criticized for attaching excessive importance to explanatory definition, for its adoption of the reductive requirement that a definiens be simpler than a definiendum, and for its use of 'canonical contexts' to disambiguate meaning. The principle of substitutability, according to which a definition of a term is accepted if it can be substituted for the term itself, is also critically examined, and the theory's use of syntactic phenomena as evidence for polysemy is shown to be inconsistent. Finally, suggestions that NSM may be a valid analytical method for only a subpart of the lexicon are rejected.

1. INTRODUCTION

One of the results of the empiricist temper prevailing in linguistics has been that the theoretical foundations of linguistic models are rarely examined as deeply as their specific proposals about the details of the domain being modelled. Nowhere, perhaps, is this bias more serious than in lexical semantics (cf. Nuyts 1993, p. 281). This article explores some fundamental issues of definition-based (as opposed to truth-function-based) lexical semantic research through a critique of the Natural Semantic Metalanguage (NSM) theory of semantic and grammatical description developed by Wierzbicka and her colleagues (the theory will henceforth be referred to as NSM; see Wierzbicka 1972, 1980, 1985, 1987, 1991, 1992, 1996, 1999; Goddard 1991; Goddard and Wierzbicka 1994; Goddard 1998, 2002 and the critiques collected in *Theoretical Linguistics* 29 give some idea of the debate which NSM has so far stimulated). NSM has been chosen for special

attention because, contrary to what might be thought, it embodies in a particularly striking form many prevalent ideas about the nature of word meaning and the desiderata of a lexical semantic model. Close attention to its theory and practice therefore promises to be instructive about many aspects of decompositional or definition-based lexical semantics.

The critique to be outlined here will be made on two main fronts. First, certain of the detailed aspects of NSM's linguistic argumentation and use of data will be criticized. Second, the chain of reasoning by which NSM's theoretical position is justified will be closely scrutinized. This second line of argument, which occupies the main part of the article, is of a more theoretical and abstract nature than is common in the largely data-dominated landscape of lexical semantic discussion to which NSM belongs; I hope it will become clear, however, that it is no less pertinent to 'empirical' lexical semantics for that. The result of these twin critiques will be to suggest that virtually none of the hypotheses, techniques and assumptions designed to confer superiority on NSM actually does so. As a result, while the depth, insight, explicitness and subtlety of the analyses developed within the framework can, and should, be admired, its broader methodological and theoretical claims should be rejected.

The article is organized as follows. Section 2 introduces the NSM program of semantic description, offering some general observations on its theoretical rationale and its conception of meaning. It concentrates especially on the role of definition and on the relation asserted to exist between universality and simplicity, and argues that NSM exaggerates the importance of definitions to linguistic theory. The next two sections advance some specific criticisms of NSM. In Section 3, I consider criticisms which are specific to NSM's mode of semantic analysis. I argue (Section 3.1) that the adoption of the reductive requirement that a definiens be simpler than a definiendum is misguided: what guarantees explanatory success is not that the definiens be simpler, but that it be already known. In Section 3.2, it is argued that NSM's use of 'canonical contexts' does not fix the meaning of primes in the required way, but leaves open the possibility of multiple ambiguity. Section 4 concentrates on three criticisms that apply particularly obviously in NSM, but which could be easily generalized to many definition-based (as distinct from formal) theories of semantics. First, the principle of substitutability is criticized, according to which a definition of a term is accepted if it can be substituted for the term itself (Section 4.1). Second, the theory's

analysis of polysemy, which is similar in its broad outline to that found in many semantic frameworks, is argued to be selectively applied (Section 4.2.1) and therefore untenable. Third, possible responses by NSM to disconfirming evidence are considered (Section 4.3). The concluding section draws out some consequences of these criticisms for a conception of linguistic semantics.

2. The Natural Semantic Metalanguage

2.1 *Introduction and General Issues*

NSM semantics represents a style of conceptual analysis characteristic of philosophical rationalism in the tradition of Leibniz. Semantic analysis in NSM involves the reductive paraphrase of definienda into a metalanguage constituted by a subset of ordinary language expressions claimed to represent universal primitive concepts. The following is a list of the English words whose meanings are considered to be primitive:

I, you, someone, people, something/thing, body; this, the same, other; one, two, some, all, much/many; good, bad; big, small; think, know, want, feel, see, hear; say, words, true; do, happen, move; there is, have; live, die; when/time, now, before, after a long time, a short time, for some time; where/place, here, above, below, far, near, side, inside; not, maybe, can, because, if; very, more; kind of, part of; like (Goddard 2002, p. 14).

NSM depends on the claim that each of these words can be translated without addition or loss of meaning into every language. Since the list could just as easily have been given in Malay or Mandarin, it is necessary to distinguish between each primitive meaning itself, represented by small capitals (e.g. GOOD), and the particular 'exponent' of the meaning in whatever language is in question (e.g. *good* in English, *bon* in French, etc.). The indefinable and universal status claimed for the primitives allows the theory to simultaneously avoid the circularity and terminological obscurity that 'dog most other semantic methods' (Goddard 2002, p. 5). 'Without a set of primitives', Wierzbicka comments (1996, p. 11), 'all descriptions of meaning are actually or potentially circular... Any set of primitives is better than none, because without some such set semantic description is inherently circular and, ultimately, untenable'. The set of NSM primitives, however, is preferable to a set of primitives established by stipulation because its membership is non-arbitrary: only those expressions which are

found to be both indefinable and universally intertranslatable (i.e., those which have equivalents in each language), are accepted as semantic primes. The meaning of any semantically complex (non-primitive) word in any language therefore reduces to a configuration of universal semantic/conceptual primitives.[1]

2.2 *NSM's Claims of Methodological Priority*

It is possible to distinguish two different claims of methodological priority consistent with NSM's methodological statements. The first, weaker claim would be the following:

(1)   The NSM set of primitives provides the best currently available lexico-grammar for descriptive and comparative semantics.

This statement does not claim anything about the accuracy of the current explications developed in NSM theory using this lexico-grammar. It only says that the best currently available semantic descriptions will use the primitives, not that the actual, existing NSM definitions are the best currently available.

The second, stronger claim is (2):

(2)   The actual definitions developed in NSM are the best currently available definitions in descriptive and comparative semantics.

Claim (1) is implicit in the entire NSM enterprise. The NSM scholars have not, however, been as clear as they might about whether they claim (2) as well. In reply to criticism from Murray and Button (1988), Wierzbicka (1988b, p. 687) states that current NSM definitions are open to revision. Many statements made by NSM practitioners, however, suggest that pending such revision, statement (2) holds. Thus, statements of NSM's definitional success (Goddard 2002, p. 7), scientificity and explanatory utility (Wierzbicka 1999, p. 10), its

---

[1] The claim that the components of the metalanguage are universal is often made in NSM writings (see Wierzbicka 1991, p. 7, which states that NSM is 'a hypothetical system of universal semantic primitives'). Elsewhere, however, this claim is scaled-down so that NSM is merely as universal as possible. Wierzbicka (1991, p. 7), for instance, says that the NSM mini-language is '*to a large extent* language-independent' (italics added).

objectivity, neutrality, and culture-independence (Wierzbicka 1999, p. 16) and ability to capture 'people's fundamental conceptual models' (Wierzbicka 1999, p. 10), all hang on the greater adequacy of its actual definitions compared to the definitions of any competing theory. The fact that NSM credits itself with these qualities suggests that (2) must be taken as being asserted: if NSM definitions were not more successful, scientific and explanatorily useful than their rivals, they would not be the 'best currently available definitions in descriptive and comparative semantics' (the best of the rival theories would be).

There is, however, an even more compelling reason for (2) to be attributed to NSM. As pointed out by Wierzbicka herself, the set of semantic primitives is only as good as its actual explanatory effectiveness: a set of universal semantic simples would be useless if it could not successfully elucidate semantically complex meanings. Since the whole NSM method is geared towards the provision of successful definitions, the existence of primitives must be taken as inseparable from their explanatory effectiveness:

The crucial point is that while most concepts ... are complex (decomposable) and culture-specific, others are simple (non-decomposable) and universal (e.g. FEEL, WANT, KNOW, THINK, SAY, DO, HAPPEN, IF); *and that the former can be explained in terms of the latter* (Wierzbicka 1999, p. 8; italics added).

The theory therefore stands or falls just as much on the issue of the adequacy of its definitions as it does on that of the universality of its elements. If NSM is to be open to genuine empirical testing, its explications of meaning cannot always be taken as provisional. The point must come where the paraphrases NSM offers are no longer promissory notes, but definitive analyses which can be submitted to decisive testing.

2.3 *Definition and Semantic Theory*

This section discusses the conceptual foundations of NSM and its construal of the task of lexical semantics.

One of the most original aspects of NSM is the fact that it is motivated less by any developed theoretical understanding of those domains often taken as relevant to the analysis of meaning (e.g. conceptions of the nature of cognition, categorization, reference, or truth: cf. the importance of these questions in, e.g., Jackendoff 1983, 1990; Lakoff 1987; Allan 2001) than from some rather practical considerations about

the nature of a particular metalinguistic practice, explanatory defini-
tion or 'explication', and the requirements that any actual definition or
explication should supposedly meet if it is to successfully convey a
word's meaning. The locus of NSM explanation is not therefore the
question 'what is happening when I understand the meaning of a
word?', but 'how can I explain the meanings of words (to others)?'.

This is an unexpected emphasis for a modern theory of lexical
semantics. For it is not obvious that the task of understanding
meaning – presumably the central task of semantic theory – should be
identified so completely with that of providing explanatory defini-
tions of individual words, in the sense of descriptions of separable
semantic components whose composition results in a representation
of lexical meaning. There are, indeed, many other metalinguistic
practices, such specification of truth-conditions or lexical relations,
description of typical contexts, text interpretation, or etymology, in
which meaning is just as crucially implicated and which, as a result,
have equal *prima facie* claim as candidates for the paradigms of
semantic theory. Of course, it should not be denied that the definition
of a word bears *some* relation to what we will want to think of as its
meaning. But the belief that any theory of semantics which, like
NSM, aspires to empirical and methodological rigour, should adopt
explanatory definition as its main task must be questioned: we should
not take it for granted either that the ultimate results of semantic
theory will necessarily resemble dictionary definitions (cf. Fodor et al.
1980; Miller & Leacock 2000; for a response see Wierzbicka 1996,
pp. 253–256), or that the best way to understand meaning is as a
determinate object open to representation (whether definitional or
not) in some metalinguistic medium (cf. Geeraerts 1993).

While a method built on definitional paraphrase may appear to
demystify meaning,[2] it remains open to the charge that it is merely a
theory of a particular metasemantic practice which has no necessary
connection to meaning as such.

2.4 *Grounding Meaning*

Non-behaviourist and non-truth-conditional semantic analyses are
subject to an apparently insuperable methodological boundary

---

[2] cf. Goddard (2002, p. 6): 'For many linguists and logicians working in other
frameworks, nothing is more mysterious and intangible than meaning. But adopting
reductive paraphrase as a way of grasping and stating meanings makes meanings
concrete, tangible'.

condition, expressed by Goddard (1994, p. 7) as the 'Semiotic Principle': 'A sign cannot be reduced or analyzed into any combination of things which are not themselves signs'. Further, NSM assumes that 'the meanings expressible in any language can be adequately described within the resources of that language' (Goddard 2002, p. 5). The signs into which meanings are analysed will therefore be words of natural language rather than the technical formalisms of other semantic theories.[3] In equating semantics with the formulation of definitions, and in stipulating that, as such, it is irreducible to anything non-linguistic, the validity of accounts of meaning based on reference, denotation, or neurophysiology is denied. Given these assumptions, and adding, uncontroversially, the methodological criterion of the undesirability of circular definitions, definitional paraphrase must be grounded in undefined elements which are not themselves susceptible of definition. It is only if the process of definition is halted at a level of undefined elements that definitions can be truly explanatory and circularity averted. Because the vocabulary of a natural language is, at least for practical purposes, finite, any attempt to define *all* its words will inevitably lead to implicitly or explicitly circular definitions, definitions, that is, in which the same expression appears as both definiendum and definiens.

The desirability of restricting the semantic metalanguage to a fixed number of elements is a powerful impetus at the centre of many semantic theories and in many explanatory frameworks in general (see Fillmore 1971; Jackendoff 1983; Allan 2001, p. 281; for some criticisms of primitives see Aitchison 1994). It is a common feature of perspicuous explanation that it characterize the explananda using a more constrained set of analytical terms than those in which the data are described pretheoretically. For NSM the question is which elements of the language are to be taken as indefinable. As pointed out by Goddard (2002, p. 13), 'one can never prove absolutely that any element is indefinable. One can only establish that all apparent avenues for reducing it to combinations of other elements have proved to be dead-ends'. The elements identified as primitive in NSM theory are those which are (a) semantically simplest, in the sense of not amenable to further definition, and (b) universal. Notice that in order for a definition to succeed (i.e. for it to be explanatorily effective) it need only possess the first of these properties. While it is obviously

---

[3] Matthewson (2003) casts considerable doubt on NSM's claim to use only natural language expressions, free of technical devices.

ineffective to explain the meaning of a word in terms of something *more* complex, it is not obvious that the most simple (least definable) meanings will *also* be those found universally. The NSM identification between the simplest and the most universal terms therefore deserves some discussion. As noted by Goddard (2002, p. 9), 'the ideal position from which to bear on the issue (of which words are definitionally most basic, i.e. simplest) would be to begin with a body of deep semantic analyses carried out on a purely language internal basis in a range of diverse languages'. This would establish which terms needed to be considered as indefinable. The analyst would then go on and look at whether the set of indefinable terms matched up cross-linguistically. Understandably, however, this has not been the course that NSM investigations have taken. As discussed by Wierzbicka (1996, p. 13), it was hypothesized from the very beginning of the theory that the sets of semantic primitives identified in each natural language corresponded to innate human concepts, and would therefore match.

Accordingly, the identification of the simplest meanings of a language with the universal ones is a significant aid in the isolation of the indefinable terms. Universality and simplicity cooperate in each other's discovery: if an element seems to be truly universal, it is likely to be indefinable, and if an element is indefinable, it may well be universal.

3. NSM-Specific Criticisms

We will now turn to some arguments against two specific aspects of the NSM program: its insistence that explications must be simpler than explicanda, and its commitment to a residue of indefinable terms.

3.1. *Greater Simplicity as the Criterion of Explanatory Success*

In order to be successful, a definition must, according to NSM, be couched in terms of something simpler: 'if a human being can understand any utterances at all (someone else's or their own)', states Wierzbicka (1996, pp. 11–12), 'it is only because these utterances are built, so to speak, out of simple elements which can be understood by themselves'. The nature of understanding

presupposed here must be questioned, however. One may agree that a successful definition must explicate a definiendum through definientia which are simpler, without accepting the existence of a canon of universal terms which represent the absolutely simplest possible elements of explanation. In other words, simplicity should not be assumed to be an invariant property of an expression that can be measured on an absolute scale. Goddard's (2002, p. 5) identification of 'simpler' with 'more intelligible' is therefore salutary. To label a sense as 'more intelligible' ('more able to be understood') brings out the fact that intelligibility is something manifested in events of understanding. Something that may be more intelligible to one person may be less intelligible to another. 'Intelligibility', that is, is not an absolute property: it can only be measured by how successfully something *is actually understood* by someone on some occasion. This relational character is obscured by the term 'simplicity', which suggests an unchanging property of an expression that is not dependent on the individuals engaged in understanding it.

What makes an explanation 'more intelligible'? Common sense suggests that the answer varies from case to case and depends on many variables. Appeal to experience, however, shows that in order to be effective, an explanation has to be couched not in terms of simpler elements (on a putative universal scale), as claimed in NSM, but in terms of something the addressee of the definition *already knows*. Prior knowledge, rather than anything else, is the criterion on which explication successfully depends (cf. Koptjevskaja-Tamm and Ahlgren 2003, p. 260). The following thought experiment is a stark illustration of this point. Imagine that a Georgian speaker is trying to explain to me (a native English speaker) the meaning of the word *c'q'al-i*. The Georgian speaker knows hardly any English, and I speak no Georgian. In particular, the Georgian does not know the English translation of *c'q'al-i*. She is, however, a chemist, and offers as her explication the formula 'H_2O', which allows me to identify *c'q'al-i* as meaning 'water'. As a theoretical and scientific definition, the explanation 'H_2O' is certainly less simple in the 'absolute order of understanding' (cf. Wierzbicka 1996, p. 10) than the word of which it is offered as the explanation. As a technical explanation within scientific chemistry, it is certainly also not universal. Yet this

definition would be successful, because the technical chemical terms of which it consists are already known to me.[4,5]

An NSM proponent might perhaps reply that this situation showed that, for me, the defining chemical term should in fact be considered as simpler. This claim, however, is untestable. Whether something is part of someone's prior knowledge can often be established empirically (for example, through ordinary questioning); whether it is simpler is a much less concrete and checkable matter. We have no way of ascertaining whether something is objectively simpler than something else (this ability would necessitate definitive analyses of the meanings involved, precisely what NSM claims to supply), but we are often able to establish what is already or not yet known. Prior knowledge therefore provides the only falsifiable hypothesis about the criterion of explanatory success. It, therefore, rather than simplicity, must be taken as the criterial condition for definitional success, *contra* NSM.

In claiming universality for its simplest semantic elements, NSM escapes this objection by, in effect, asserting an identity between the simplest meanings and the already-known ones. Since the semantic primes are assumed to be part of humans' innate conceptual structure, they are always available to the understanding as the building-blocks for more complex meanings: they are, in other words, always already known. The supposed innateness of the primes therefore constitutes a counter-argument to the criticism that NSM adopts in simplicity a mistaken criterion of explanatory success. But

---

[4] An NSM theorist might claim in rebuttal that a technical chemical classification is not a definition of the English word *water*, but a definition of the corresponding scientific concept. While this definition might identify the *referent* of the word, it does not specify the *sense*. But given that what one takes to be the sense of the word *water* is, as it were, a matter of definition, in that it is not open to any external and objective checking, but depends on the details of one's semantic theory, 'H$_2$O' has as great a claim as anything else as the definition of *water*. In the present example, it would certainly allow us to use *c′q′al-i* properly. The aspects of meaning left out by the chemical definition, and claimed by NSM to be part of the sense of the word, could be treated as part of the encyclopaedic knowledge we have about water, not as knowledge of the word's meaning as such – a distinction on which NSM often insists (e.g. Wierzbicka 1996, p. 262), even if it often considers a word's 'meaning as such' to be highly detailed and rich.

[5] As most recently recalled by Barker (2003), recapitulating Kripke (1980), the meaning of proper names and of natural kind terms like *water* are inherently resistant to explication through paraphrase; as a result, any NSM explanation of these terms cannot be considered as an explication of their meaning.

since semantic universals are hypothesized to exist precisely in order to render explanatory definition through simpler terms non-circular, a method of semantic analysis which takes prior knowledge as its criterion of explanation has no need of them. Only if greater simplicity is substituted for prior knowledge as the universal characteristic of semantic explanation does a level of ultimate simples become necessary: the process of definitional simplification cannot, clearly, go on for ever. But if semantic explanation is assumed to operate by relating definienda to meanings which are already known, no universal array of absolutely simple ideas need be supposed. It only makes sense to believe in the existence of semantic primitives if we believe that explanation proceeds via reduction to simpler elements. As the example of Georgian *c′q′al-i* shows, however, this is not necessarily the case.

If this argument is accepted, the NSM method of semantic analysis will begin to look increasingly unlike an adequate approach even to the definitional explanation of meaning: to define a meaning correctly we do not have to build it up out of a level of supposedly elementary particles, but only relate it to meanings with which the learner is already familiar. As noted initially, the sets of meanings related in this way will differ rather significantly from one learner to another. This is not a trivial point. I have mainly, in this discussion, been granting to NSM that it is possible to specify a list of criteria which can predetermine the possible success of a semantic explanation. I will end this section by calling that assumption into question. The contrary claim, in fact, seems closer to the truth: *whether a word is successfully explained or not by a given metalinguistic formula is not a question that can be answered in the abstract*. This is because successful explanation is subject to significant interpersonal variation: as is, I think, widely recognized among parents, language teachers, field workers and general stakeholders in the ordinary day-to-day explanation of meaning, what works well for one person may not work well for another. Whether or not a word's meaning has been successfully explicated, and its understanding thereby achieved, cannot be determined by the extent to which a proposed explication conforms to a pre-established scheme: an explication's effectiveness cannot be measured with an invariant algorithm, but is sensitive to the particularities of each situation in which the definition is needed – not just superficial particularities, but deep ones having to do with the cognitive, cultural, and historical contingencies of each individual in the learning experience. This is a truism which I take to be so obvious as

not to require any argument. For the sake of completeness, however, I invite the reader simply to reflect on their experience in explaining meanings to others, and to recall, in particular, those occasions, which inevitably will have arisen, on which the 'correct' definition of a term has not been grasped by a learner, necessitating the discovery of an alternative stratagem.

The success of a definition, then, is not guaranteed if its elements are part of a deductive system that captures the essential meaning of words by reflecting the 'absolute order of understanding'. Definitions are not abstract algorithms, but practical tools used by real speakers to solve real problems of understanding. They are thus not dependent on principles of logical coherence, but on whatever means work to communicate the meaning – whatever it takes for the learner to 'get it', including ostension, analogy, translation and, if necessary, circularity. The alleged impossibility of an algorithm to determine an expression's degree of simplicity and its consequent explanatory utility would not affect NSM if it did not claim for itself a high degree of *actual explanatory effectiveness*; if it did not, in other words, claim that the validity of its method is to be measured by the success of its definitions in actually explicating the meanings of definienda. But this is the very claim often made by NSM theorists. It is, for instance, the justification for the repudiation of circularity as a definitional tool (see, e.g. Wierzbicka 1996, pp. 274–278). Yet as anyone knows who has tried to explain the meaning of terms to language learners (whether it is a first or second language in question) explanation in even the simplest possible metasemantic terms may not succeed. Not only is a maximally simple paraphrase not a *sufficient* condition for successful explanation (in that as well as hearing or reading the definition, the learner must also understand, or 'get' it), it is not even a *necessary* one: successful explanation is often achieved, for many concrete words, ostensively rather than through paraphrase. To explain to a Chinese speaking botanist the meaning of the English word *conifer* we will adopt a very different procedure from the one we would use with a Chinese speaking 4-year old, but in each case our definition will have real explanatory value, by the only criterion that should surely count in an empirical theory, and the criterion which NSM in fact adopts: that of whether it succeeds in conveying the meaning of the word to the learner. The best definition will thus depend on a variety of contingent variables in the person to whom the definition is addressed.

This is not, as might be objected, a trivial point about the necessity of idealization: it will not do to say that maximally simple definitions are those which inevitably lead to understanding *under ideal conditions*. The claim here is that meaning, perhaps unlike other components of linguistic description, is so deeply embedded in the particularities of individual and social variation that it is impossible to abstract a single, invariant paraphrase which can serve as the successful definition of a term. If we accept actual explanatory adequacy as the criterion of measurement for definitional adequacy, we must acknowledge that the means for creating a successful definition of a word will vary radically from one situation to the next and that as a result there is no such thing as a necessary condition of definitional success.

3.2. *Canonical Contexts*

We turn now to the means by which the exact membership of the set of universal primes is determined, and the methods used to discover whether a certain language contains an exponent of a putatively primitive meaning. As noted by, among others, a number of the contributors to Goddard and Wierzbicka (1994), many – we might add, perhaps all – of the English exponents of the primes are polysemous, with only one of the many meanings expressed by each being identified as universal (for some discussion of this point, see Cattelain 1995). For example, in testing for the presence of an exponent of a primitive meaning in a particular language, it is not enough to simply ask whether the language in question has words for 'I, YOU, SOMEONE, PEOPLE, BIG, GOOD, TRUE' and the other presumed primes; instead, it is necessary to distinguish the sense claimed as universal from the others: is the primitive TRUE, for instance, better represented by the meaning present in (3) or (4)?

(3)   If you read it in a book it must be true.

(4)   You must be true to yourself.

In answering questions like this the theory encounters a problem of its own making. Because the direction of semantic explanation must always proceed from complex to simple, the allegedly universal sense cannot be distinguished in the most obvious way, i.e. simply by *defining* it through other words: since the semantic primitives are

indefinable, any such attempted definition would inevitably use more complex terms and hence be invalid. The solution to this problem is to 'indicate for each proposed prime a set of 'canonical contexts' in which it can occur; that is, a set of sentences or sentence fragments exemplifying grammatical (combinatorial) contexts for each prime' (Goddard 2002, p. 14) which allow the primitive meaning to be identified. For example, only the (a) sentences below are considered to involve primitive senses of the underlined verbs:

(5) a. This person can't move (Wierzbicka 1996, p. 30).
    b. Her words moved me.

(6) a. (When this happened), I felt something good/bad (Goddard 2002, p. 15).
    b. I am feeling your pulse.

Sentences like (5a) and (6a) define the canonical contexts (also called 'canonical sentences': Wierzbicka 1996, p. 30) which can be used to test the validity of NSM primes. 'Merely listing the English word *feel*', for example, 'does not indicate which of these contexts is intended' (Goddard 2002, p. 15). The canonical contexts are supposed to make it clear which of the many possible meanings are intended as semantic primes.

  Sentences (5a) and (6a) are, however, multiply ambiguous. Thus, (5a) could have at least the following three interpretations, of which presumably only (one of the many possible readings of) (7a) is the one intended:

(7) a. This person can't move (part of) their body.[6]
    b. This person can't change dwelling.
    c. This person can't change their ideas (about a particular issue).

Likewise, (6a) could refer to either of the following situations, of which presumably only (8a) corresponds to the canonical context:

---

[6] Durst (2003, p. 298) agrees that (7a) is the correct interpretation of (5a) and claims that it is 'quite unambiguous'. Consider, however, that it could equally be applied to an undertaker unable to shift a particularly heavy corpse, to someone unable to raise a single eyebrow and to someone unable to lift their own severed leg, as well as to someone completely paralyzed by a sporting injury.

(8)a.   (When this happened), I had a good/bad feeling.
   b.   (When this happened), I perceived something good/bad by touching it.

The existence of ambiguity in these canonical sentences is not accidental. Specification of a canonical context will never be enough to exclude all unwanted senses, since no sentence can uniquely determine a single meaning: the possibility of multiple interpretations can never be excluded, even in a rigorously formalized metalanguage. The canonical contexts thus do not provide an unambiguous delineation of a single meaning, but require significant contextualization in order to impose the required reading. To elicit from an informant an equivalent for 'move' in (5a), for example, an NSM theorist would have to engage in a considerable amount of stage setting – for instance, by asking the informant what one would say in certain characteristic situations in which the intended sense of 'this person cannot move' would be appropriate (someone confined to a wheelchair, say). In order to render these specifications explicit, replicable, and open to scrutiny – qualities which they must have if they are to be admitted as parts of a responsible procedure – it would be necessary to use semantically more complex terms, thus reversing the only direction of explanation which NSM endorses.

    The inherent ambiguity of canonical contexts means that they require disambiguation through definition in language. Adequate disambiguation cannot be provided, however, without violating the main principle of the analysis, namely that some elements must be left undefined.

4. MORE GENERAL PROBLEMS

In this section, we turn to several criticisms of NSM which are of wider interest to semantic theory in general. They concern the argument from substitution by which a definiens is validated as the correct analysis of a definiendum, the question of polysemy, and the role of disconfirmation in semantic theory.

4.1. *Substitution as an Index of Identity*

This section explores the status of substitutability in NSM. Section 4.1.1 contains the main discussion of the issue and Section 4.1.2 considers implications of the argument for NSM's claimed non-objectivism.

4.1.1. *Main discussion*

In NSM as in most other definitional semantic theories, a minimum requirement on a term's definition is that it be substitutable for the term itself. The *locus classicus* of this requirement is its articulation by Leibniz: two things are the same if they can be substituted one for the other with truth intact. In NSM, the principle in question can be reconstructed as having the following form:

1. Substitutability
Linguistic elements (x and y) can be substituted for element z
('unmarried' + 'male' can be substituted for 'bachelor')

therefore

2. Identity
The meaning of z is identical to the composition of the elements (x and y).
(the meaning of *bachelor* is identical to the composition of the two elements 'unmarried' and 'male').

The substitutability principle is regularly appealed to in order to test proposed NSM analyses: if the semantic paraphrase can be substituted for the definiendum, then it is accepted as accurate. Note that in NSM it is not identity of truth, but identity of *meaning* that is required between definiens and definiendum: only if the definiens can be substituted for the definiendum without loss or addition of meaning (*salvo sensu*) in the original context (*in locum*) is it accepted as its correct analysis (Wierzbicka 1988, p. 12).[7] The apparent circularity of this aspect of the argumentation will be considered shortly. First, however, it is necessary to observe that the conclusion from (1) to (2) is not *prima facie* warranted by the intuitive force of *identity* and *substitutability*. This is because substitutability and identity are quite different relations: identity is about the inner essence of something, whereas substitutability is about equivalence *with respect to a given function* – it concerns, in other words, *the role something has in a particular context*. Whereas the semantic identity of a linguistic unit is assumed to be fixed – it has an invariant 'essential nature' which is precisely what semantic analysis aims to uncover – substitutability varies from one situation to another, depending on what is at stake in

---

[7] We will ignore the fact that since NSM paraphrases target only the invariant part of an expression's meaning, *all* semantic explications involve meaning loss.

each substitution. The fact that one linguistic expression can be substituted for another in the context of a definitional practice therefore does not tell us anything more than that the two elements are functionally equivalent for this purpose. As a result, an attempt to argue from substitutability within a definition to semantic identity will necessitate a theory of the relationship between definition and meaning – a relationship which, if the present arguments are correct, is substantially different from the one usually assumed. In the absence of such a theory, a semantic method which simply analyses expressions into a definitional metalanguage should not, strictly, be thought of as a theory of meaning, but as a theory of definition.

As noted, preservation of truth is not the criterion adopted in NSM to regulate definitional substitutions: NSM scholars have repeatedly, and correctly, denied the accusation that their method is 'objectivist' in this sense (Goddard 2002, p. 8). Instead, the criterion of preservation of meaning is used: an NSM definition is accepted if it can be substituted *salvo sensu* for the definiendum (Wierzbicka 1988a, p. 12; Goddard 2002, p. 6): that is, if it involves neither addition nor loss of meaning with respect to the meaning of the definiendum. At this point an important problem arises. For what is the metalanguage in which the meanings of a definiendum and an NSM paraphrase can be represented, in order to determine whether or not they are, in fact, identical? Without such an independent determination the argument for the correctness of the NSM paraphrase is both stipulative and circular: we are asked to accept an NSM definition as a true representation of the meaning of a definiendum because it does not involve any addition to or loss from this meaning – because, in other words, it is a true representation of its meaning.

This is a duplication on a different level of the very problem for which NSM is suggested as the answer in the first place. As Wierzbicka puts it:

To compare meanings expressed in different languages and different cultures, one needs a semantic metalanguage independent, in essence, of any particular language or culture – and yet accessible and open to interpretation through any language. (1991, p. 6).

But this point applies just as much to the comparison of meaning necessary to verify the accuracy of an NSM paraphrase as it does to the comparison of meaning which NSM claims to facilitate for ordinary linguistic semantics. If it is to be demonstrated, rather than merely asserted, that a definiendum and its proposed NSM definiens have the same meaning, some additional and accurate

semantic representation is needed in which the meaning of both definiens and definiendum can be objectively examined. Paradoxically, however, such a metalanguage is precisely the tool that NSM claims to be uniquely supplying, and which we must therefore presume not to be available before the final realization of the NSM system. NSM frequently claims, indeed, that any other semantic metalanguage – including ordinary language, with its commonly decried inadequacies – is subject to the usual faults of ethnocentrism, circularity and terminological obscurity, which the developed NSM lexicon seeks to transcend. By its own admission, therefore, the semantic metalanguage necessary to assess the matching of definiendum and definiens does not exist.

This problem would matter less if NSM did not claim to provide a theoretically principled basis for semantic research which removes the distorting ethnocentrism bedevilling other semantic theories (see Wierzbicka 1991, p. 148, 1996, p. 239, 1999, pp. 23–24). If NSM saw itself as one among a number of equally subjective, culture-specific modes of semantic representation, it would be no more or less affected than its fellows by its ultimate reliance on intuitive semantic judgements. It is argued here that these judgements are absolutely unavoidable, and that they install an irreducible degree of subjectivity into semantic analysis. If semantic analysis is irreducibly subjective, there is little point in trying to render it culture-neutral, since this will not remove the even deeper level of subjective bias. As it is, however, NSM claims to be categorically different from comparable semantic theories in the scientificity, rigour and culture-neutrality of its method, and to '(submit) itself to a higher standard of verifiability than any rival method' (Goddard 2002, p. 11). But without a metalanguage in which the meaning of definiendum and definiens can be accurately and explicitly represented and contrasted, investigators' semantic judgements, as well as the intuitions and methodological proclivities on which they are based, are effectively placed beyond scrutiny, a fact which robs NSM of its claimed methodological superiority.

NSM theory thus presupposes a pretheoretical interim vocabulary in which initial observations about semantic facts can be couched, and judgements of semantic identity made explicit and legitimated. This vocabulary would be analogous, perhaps, to the ordinary vocabulary in which astronomical observations are phrased, and of which astronomical theories are the refinements: observations like 'there is a stationary light x degrees above the horizon'. The failure of

NSM to sustain its own claim to provide a maximally neutral medium for semantic description derives, it is argued here, from the fact that no such vocabulary exists: any semantic metalanguage depends on a high degree of subjective, intuitive semantic judgement.

So far I have principally been arguing that the absence of an objective metalanguage from the development stage of any NSM paraphrase simply compromises the theory's ability to *justify* its particular final paraphrases. I will now extend the argument in order, ultimately, to show that without such a metalanguage, a semanticist cannot even *refer to* the semantic features of a word which need to be reflected in its definition without continually running the very risks (terminological obscurity, circularity) which further threaten the theory's justificatory basis and which only the finished NSM lexico-grammar will escape. Unlike the relatively neutral observational vocabulary of astronomy, which involves uncontroversial notions on which observers can agree (degrees above the horizon, cardinal directions, brightness, etc.) and which do not strongly determine any one theoretical treatment, the initial observational language of semantics strongly influences the nature of the subsequent theoretical representation by constituting the very (culture-specific) terms in which the meaning of a definiendum is first represented, and which the NSM definition seeks to purify. Since these initial descriptions inevitably contain many semantically complex, multiply ambiguous words, they do not provide the firm and unambiguous basis for semantic description that NSM requires. If ordinary language semantic descriptions are thoroughly infected with obscurity, circularity and latent culture-specificity, they should not be relied upon at *any* stage of the process of semantic description: any preliminary characterization of an aspect of a term's meaning, on which the NSM paraphrase is based (e.g. 'wetness, freshness, succulence' as relevant to the Hanunóo word *latuy*: Wierzbicka 1996, p. 307, following Conklin 1964, p. 191), can be claimed as an inaccurate because potentially ethnocentric, unclear, or overly complex.

Let us examine a particular instance of this dilemma, Wierzbicka's treatment of the Japanese noun *amae* (1996, pp. 238–239). In developing an NSM paraphrase for this noun, Wierzbicka refers to many non-NSM descriptions and definitions of its meaning and that of related words, as found in existing lexicographical and other sources. These definitions are the pretheoretical descriptions that motivate –

and justify – the eventual NSM paraphrase, and they include the following:

‘helplessness and the desire to be loved’, ‘lean on a person’s goodwill’, ‘depend on another’s affection’, ‘act lovingly towards (as a much fondled child towards its parents)’, ‘to presume upon’, ‘to take advantage of’, ‘to behave like a spoilt child’, ‘be coquettish’, ‘trespass on’, ‘behave in a caressing manner towards a man’, ‘to speak in a coquettish tone’, ‘encroach on [one’s kindness, good nature, etc.]’, ‘presume on another’s love’, ‘coax’; ‘take advantage of’, ‘play baby’, ‘make up to [someone] and get their sympathy’, ‘coax’, ‘act spoilt’ (for *amae*, n); ‘depend and presume upon another’s benevolence’, ‘wish to be loved’, ‘dependency needs’ (for *amaeru*, vb).

These descriptions are, collectively and individually, highly ambiguous: how many different situations, for instance, can be conveyed by ‘coax’ or ‘trespass on’? They are also deeply culture-specific (consider ‘be coquettish’ or ‘act spoilt’). If the NSM set of primitives is to supply ‘constant points of reference, which slippery labels with shifting meanings cannot possibly provide’ (Wierzbicka 1996, p. 456), it must not simply inherit the weaknesses of the pretheoretical descriptions on which it is based. There is no point in an NSM paraphrase’s being couched in universal vocabulary if the initial descriptions which it has been designed to reflect are themselves highly culture-specific.

As examples of ethnocentric, semantically complex, and ambiguous talk about meaning, the initial descriptions license a wide range of possible NSM paraphrases, and can only be used as input to an NSM definition after undergoing a particular interpretation. Yet, given the ‘slipperiness’ of the descriptions, there is no way to justify any one of the possible interpretations over another. In the case of *amae*, for example, it is clear that the NSM paraphrase developed ‘(o)n the basis of these and other similar clues’ (Wierzbicka 1996, p. 238), represents just one of many possible preliminary meaning descriptions:

amae
(*a*)X thinks something like this:
(*b*)   when *Y* thinks about me, *Y* feels something good
(*c*)   *Y* wants to do good things for me
(*d*)   *Y* can do good things for me
(*e*)   when I am near *Y* nothing bad can happen to me
(*f*)   I don’t have to do anything because of this
(*g*)   I want to be near *Y*
(*h*)X feels something good because of this (Wierzbicka 1996, p. 239)

The NSM paraphrase is thus a refinement of (selected) pre-existing descriptions which, insofar as they are framed in ordinary language, are subject to its failings of ethnocentrism, culture-specificity and so on. Yet it is these descriptions to which the eventual paraphrase is explicitly tied. Wierzbicka justifies its various components in terms of their correspondence to aspects of the earlier descriptions, especially those in Doi (1981):

Doi emphasizes that *amae* presupposes conscious awareness. The subcomponent (*a*) '*X* thinks something like this . . .' reflects this. The presumption of a special relationship is reflected in the component (*b*) 'when *Y* thinks about me, *Y* feels something good'. The implication of self-indulgence is rooted in the emotional security of someone who knows that he or she is loved: 'it is an emotion that takes the other person's love for granted' (Doi 1981, p. 168). This is accounted for by the combination of components (*b*) 'when *Y* thinks about me, *Y* feels something good', (*c*) '*Y* wants to do good things for me', (*d*) '*Y* can do good things for me', and (*e*) 'when I am near *Y* nothing bad can happen to me' (Wierzbicka 1996, p. 239).

The line from any one of these statements to the component of the definition is far from unambiguous: the statements do not uniquely determine the particular NSM phrasing adopted, and the NSM phrasing does not uniquely connote the statements. It is therefore just one particular construal of these statements that is adopted, and others are concomitantly excluded. Even if the existing paraphrase is a good representation of the meaning of *amae*, the point still remains that we can have no other justification of the paraphrase's appropriateness than an intuitive one: since the preliminary semantic descriptions could motivate a number of different NSM realizations, according to the particular construals made of them, it is always up to the individual investigator to decide which paraphrase fits best. Given the divergence of possible opinions, this is hardly an open standard of verifiability at all, and NSM's claim to supply a maximally culture-neutral, non-arbitrary representation is therefore vitiated.

This is a problem from which no semantic theory may claim to escape. Any attempt to discuss meaning presupposes an initial metasemantic vocabulary in which the first, rough impressions of meaning are couched, and relies on the investigator's own intuitive judgements of identity and difference between definientia and definienda – quite in conflict with the foundational and purificatory instincts at the core of NSM analysis. Proposed refinements of this vocabulary will inevitably depend on the initial gross delineation of the semantic facts which it imposes. And in the absence of an

independently justified metalanguage in which claims of identity be-
tween definiens and definiendum can be justified, the theory remains
circular. As just observed, this is only a problem if unrealistic claims
are made of the theory. A theory which claims an absolute contrast
between its fully developed, 'purified' method of semantic description
and its observational predecessors inevitably deprives itself of a
means of justifying its choice of elements. In contrast, a method of
semantic analysis prepared to acknowledge its own inevitably
adventitious nature does not have to defend a claim of methodo-
logical priority over rival analyses.

A Wierzbickian might respond that the initial terms used to talk
about aspects of a word's meaning during the evolution of a full
NSM representation are no more than labels serving to name certain
intuitively grasped semantic properties of the word in question.[8] The
finished NSM paraphrase, on this view, would not be *shown* to be
semantically identical to the definiendum, it would simply be en-
dorsed as such after a process of introspection in which the investi-
gator scrutinized their intuitions and determined that the
definiendum and the NSM paraphrase matched in meaning. Intuited
properties, however, while inescapable in semantic analysis, are,
paradoxically, an unsatisfactory basis for the sort of analysis to which
NSM aspires, given the vagueness and variability of intuitions within
and between individuals, and the consequent unlikelihood that they
could ever be disciplined stringently enough to yield semantic judg-
ements of the requisite certainty, delicacy, or depth. Even if such
discipline was possible, the match between paraphrase and definien-
dum could only ever be asserted, never demonstrated – hardly a
satisfactory situation for a methodology that claims to provide 'clear
standards of precision' (Wierzbicka 1991, p. 283).

This last point should be stressed. Intuitions themselves cannot
figure directly in the explicit argumentation of semantic analysis, but
must first be named in language. As intuitions, indeed, they are
theoretically inert, since the nature of the semantic property identified
by a named intuition can only be specified through an elaboration of
those conventionally accepted terms which can be definitionally
related to, or accepted as satisfactory analyses of, the label in

---

[8] These properties may be intuitively grasped either by the framer of the NSM
definition, or by those responsible for the reports on which the NSM definition is
based. In the second case, the intuitions only enter the process indirectly, since the
definition is only the product of second hand knowledge.

question. The conventional properties of the label must, in other words, coincide with the properties of the intuition. Thus, one may choose the label 'positive evaluation' for an intuited semantic feature of the words *nice, kind, tasty, happy, pretty*, etc. (Goddard 2002, p. 16), but this will only be accurate in the process of framing definitions of these terms as long as the meaning of 'positive evaluation' is itself compatible with the meaning of the words being defined. For example, it is possible to associate the noun 'evaluation' with calculation and deliberation of a rather cold, detached and unspontaneous kind – quite frequent connotations of the noun, I suggest. If these connotations are mistakenly taken to be part of the intuited semantic content of the definienda, and enter into the subsequent definitions, the meanings of *nice, kind, tasty, happy, pretty*, which do not include these connotations, will be misrepresented in the finished paraphrase.

The point that a label like 'positive evaluation', when used to mark an intuited property, needs to be appropriately chosen is, no doubt, entirely obvious. Less obvious, perhaps, is the point that while the intuited semantic property may fall within the semantic range of the metalinguistic description chosen to label it (in the case of *nice*, 'positive evaluation'), many other semantic properties which have, in fact, *not* been intuited will also fall within this range: as has just been shown in the case of 'positive evaluation', the range of the application of the metasemantic label will usually be *greater* than that of the intuited semantic feature (this, simply because of the very imprecision of ordinary language which NSM acknowledges and tries to escape). As a result, it will be necessary to specify some way of narrowing down the range of connoted semantic properties expressed by the label so that it applies to the intuited feature of the definiendum alone, excluding unwanted semantic properties. The claim made here is that language will never be able to be matched precisely enough onto intuitions for this (this is why there are so many possible ways of describing the meaning of a word, all of which conform to our intuitions), and that, as a result, there is an irreducible core of intuition in semantic analysis which prohibits the type of regimented and unique description of meaning which NSM claims to provide.

The chain of reasoning that issues in the finished paraphrase is not, therefore, of the kind characterized by the rigorous and deductive working out of argumentative steps, but one in which intuition, subjectivity and hence indeterminacy, enter at crucial points, especially as concerns the relation between a proposed gloss and the

intuited semantic feature to which it refers. The justification for one particular semantic description over another cannot therefore be made objective and rigorous, but always rests on necessarily subjective, intuitive judgements of semantic appropriateness. In order to escape ethnocentrism, it is not enough for a definition to be *framed* in supposedly universal terms: it must also *be based on culture-neutral evidence*. A definition does not stop being ethnocentric simply because its formulation uses universal elements, since it may embody an entirely culture-dependent perspective at a deeper level. This, I suggest, is always the case. NSM claims to do more than provide a lexicon of universal elements which can be used to couch definitions which would have the same meaning in any language. It also claims that the particular definitions it offers provide a reliable basis for comparative research into meaning. The first claim has been widely questioned; here, I have tried to show that even if the primitives are accepted as universal meanings, the definitions in which they figure continue to embody highly culture-specific, subjective descriptions of meaning. To adapt a frequent Wierzbickian metaphor, we always see meaning through the prism of our own *selves*: even granting that the NSM primitives are universal, the theory cannot eliminate the subjectivity of the semantic judgements necessary to the development of its paraphrases. The view from nowhere (or from almost nowhere) promised by NSM should be seen as illusory.

4.1.2. *NSM and Objectivism*
We now must note a sense in which, despite its disavowals, NSM remains thoroughly objectivist. Given that NSM aims to identify the meaning of each definiendum, and that the definiens must in each case therefore be unknown until after the NSM analysis has been achieved, there is (once a maximally simple and universal set of primes has been evolved) no other criterion to regulate the definitional substitution than preservation of truth. If it is acknowledged that intuitions are not reliable or deep enough to serve, and that the method of substitutability *salvo sensu* is circular, the only remaining criterion of whether an NSM paraphrase adequately represents the meaning of a definiendum is whether it is true under the same conditions. NSM, therefore, faces the paradox that the only possible justification that would furnish its procedure with the methodological certainty it claims is the one it explicitly rejects.

4.2. *Diagnosing Polysemy*

Any attempt to describe meaning must acknowledge the existence of different senses attached to a single word. Without such recognition, description through ordinary language paraphrase or through intuitively manipulable formal concepts becomes impossible. In order to describe the meaning of the English word *crown*, for instance, a theory must allow at least the following three different (though related) senses to be distinguished, as exemplified in (9)–(11):

(9)   Kings and queens wear crowns at official ceremonies. (COBUILD: *crown*)

(10)   The crown of his head is completely bald. (COBUILD: *crown*)

(11)   This would have cost twenty Swedish crowns.

Without recognition of three different meanings in these contexts, it would be impossible to develop a unitary definition that accounted for all and only the three types of use we see here. Since 'monarch's head covering', 'top part' and 'unit of currency' have virtually nothing in common – 'thing' could not even be advanced as a genuine common factor, since it is not clear that a 'top part' of something is itself a 'thing' – any definition that did not acknowledge a difference between them could hardly avoid including as examples of *crown* indefinitely many things (and non-things) which are not known as *crowns* in English.

NSM has an even greater need than other semantic theories to acknowledge different senses within the one word. This is because languages often appear to violate a key hypothesis of the NSM program, the *Strong Lexicalisation Hypothesis* (henceforth SLH), according to which '(e)very semantically primitive meaning can be expressed through a distinct word, morpheme or fixed phrase in every language' (Goddard 1994, pp. 13; see Bohnemeyer 2003 for arguments against the SLH). Apparent disconfirmations of this principle abound. The meanings IF and WHEN, for example, both hypothesized to be primitive and therefore under the scope of SLH, are realized by the same morpheme in Japanese, conjunctive *-ba* (Goddard 1998, p. 138). Similarly, Pitjantjatjara *kulini* does service for both primitive meanings THINK and HEAR. *Prima facie,* these examples would appear to be strong disconfirmations of SLH. They are claimed, however,

not to be disconfirmations at all. This is because *-ba* and *kulini* are analyzed as polysemous: they do not, that is, merge the two supposedly primitive meanings into a single general sense, but are ambiguous (polysemous) between them. In this way, supposition of polysemy allows the SLH to be maintained because the primitive meanings remain distinct.

Polysemy is a reasonable supposition only if subject to controls. It would clearly be unsatisfactory if any word which appeared to merge putative primitives could be dismissed as polysemous. NSM research must therefore motivate a diagnosis of polysemy in order to show that it is not simply hypothesized as an ad hoc fix of disconfirmations of the theory. As in a number of other approaches to semantics, this is achieved through an appeal to syntax (see, e.g. Weinreich 1966, pp. 177–183; Croft 1998 contains relevant discussion): 'lexicographers agree on at least one mechanical diagnostic of polysemy, namely the possession of mutually exclusive syntactic frames or combinatorial possibilities' (Goddard and Wierzbicka 1994, p. 32). In other words, an expression is taken to be polysemous between two senses if each is associated with differing syntactic possibilities: a different syntactic frame/combinatorial possibility shows a different (polysemous) meaning. We will refer to this principle as the 'Syntactic Evidence for Polysemy Principle' (SEP). Exactly what does and does not count as a distinct syntactic possibility or combinatorial frame has never, to my knowledge, been explicitly stated by NSM researchers. The case of *kulini*, however, may be taken as representative, Goddard noting (1991, pp. 33–34) that only the 'think' sense takes a quasi-quotational complement introduced by *alatji* 'like this', and only the 'hear' sense takes a nonfinite 'circumstantial' complement.

I have elsewhere presented counterevidence to SEP, showing cases in which clearly identical meanings are associated with differing syntactic options (Riemer 2003, forthcoming). In Section 4.2.1, however, I will show that SEP is only selectively applied in current NSM analyses. On a criterion of theory-internal consistency, then, the current use of SEP is methodologically untenable.

4.2.1.   *SEP is applied selectively*
A point not often appreciated in discussions of this subject is that if SEP is to be a useful and rigorous diagnostic for polysemy, it has to be absolute: the claim has to be that *whenever* a lexeme is associated with more than one syntactic frame or combinatorial possibility, then it has different (polysemous) meanings. If the principle is not

absolute, some other criterion – most obviously a semantic one – will have to be invoked in order to adjudicate between unclear cases. Since SEP has been advanced as a way of regulating possible semantic paraphrases, such an appeal to a semantic criterion would be circular. It is clear, however, that SEP is not advanced as an absolute indicator of polysemy in NSM theory. In a discussion of *advise*, for example, Wierzbicka (1996, p. 243, based on the discussion in Wierzbicka 1987, pp. 181–183) notes the existence of two different syntactic frames in which the verb can appear:

(12)   The doctor advised Bill to have complete rest.

(13)   The doctor advised complete rest.

If SEP were applied consistently, it would be necessary to claim that *advise* was polysemous, with (12) and (13) instantiating different senses.[9] Yet this is precisely what Wierzbicka denies: 'the verb itself does have an invariant meaning, evident in both these frames (associated with the alleged meanings 1 and 2)' (1996, p. 243). The SEP is thus apparently only applied selectively.

  The existence of SEP is also a problem for existing NSM primes which show differing syntactic possibilities. The prime *know* can appear without any lexical or clausal object as in (14), with an NP direct object as in (15), with a *that* complement (16), a *what* complement (17) or an *if* complement (18), as its uses in the following NSM paraphrases show:

(14)   I know now: this good thing will not happen (Wierzbicka 1996, p. 179 *disappointment*)

(15)   I want to know more about it (Wierzbicka 1996, p. 179 *surprise*)

(16)   I didn't know that it happened (Wierzbicka 1996, p. 180 *sad*)

(17)   I don't know what I can do (Wierzbicka 1996, p. 181 *distressed*)

(18)   I don't know if I can do anything (Wierzbicka 1996, p. 182 *upset*)

---

[9] Alternatively, it would be necessary to define 'syntactic frame' in such a way that the difference between (12) and (13) did not count. The NSM theory has not, to my knowledge, ever provided such a definition.

Yet *know* is a prime and therefore has the same meaning in all five (syntactically different) contexts: at the very least, the contrast between (14), where *know* governs a direct object, and (15)–(18), where its object is a noun or a clause, looks syntactically deep enough to fall under SEP. The fact that these gross differences in syntactic frames are not even commented on should raise some questions about the use of SEP in NSM. If it is used to distinguish allegedly polysemous senses in Japanese or Pitjantjatjara, why is it not also used in English?

4.3. *Disconfirmation and 'Partial Coverage'*

Empirical disconfirmation of NSM analyses does not necessarily constitute good grounds for rejecting the theory: as with any other framework, further facts may always be brought to light which will explain the apparent disconfirmations in some other way. Goddard (2002, p. 6), however, suggests the following avenue of NSM response to empirical challenge:

> Perhaps the venture will work out well in some respects and not so well in others; there is no reason to assume *a priori* that it is an all or nothing affair.

This amounts to the suggestion that NSM primitives might underlie some but not all of universal semantic structure. Given the theory's strongly universalist claims, however, it cannot afford to entertain the possibility of 'partial coverage' of the semantics of language: the whole attraction of the NSM program, as of any theory of semantic primitives, lies in its claim to provide a key that unlocks *all* meaning. Exhaustivity is, indeed, integral to the notion of a set of semantic primitives: the semantic primitives of a language are, precisely, those words which are required for the definition of the language's *entire* vocabulary. As a result, there is something paradoxical in the idea that a set of semantic primitives might apply to some but not all words. The 'alphabet of human thought' is not a real alphabet if it cannot be used to spell everything: if the primitives cannot be used everywhere, a critic might ask, why should they be used anywhere? We can grant to NSM the right to pursue its research in the face of disconfirming evidence, on the supposition that further facts will be uncovered which will bring failures of existing analyses under the explanatory control of the theory by showing why they fail and, ideally, allowing predictions to be made about whether a particular, as yet unexplored, area of the lexicon would be likely to yield to NSM analysis. We should not, however, accept the possibility of a

restricted NSM that is used simply wherever it can be made to work, in the face of acknowledged failures elsewhere. Accepting this would be an annulment of the theory's claim of methodological rigour, and a dissolution of its broader metaphysical postulates about the nature of meaning. If some vocabulary proves to be resistant to definition using the set of primes, claims that the primes are the building blocks of meaning *tout court* become unsustainable, and the theory is left unable to answer the charge that those of its definitions which are apparently successful are not in fact the correct semantic analyses of their definienda. Therefore, the only attitude to disconfirming evidence which NSM can afford to adopt is that later research will allow apparent disconfirmations of the theory to be brought under its scope and that, as a result, the theory can maintain its claim that the existing primitives underlie all meaning.

NSM scholars often seem to appreciate that the value of NSM theory lies in its universality: NSM's main attraction is that it provides a way of analysing *all* meaning. Goddard himself, for example, in the same article from which the previously quoted comment is drawn, explicitly discounts the possibility of a partial NSM:

...taken as a whole, the metalanguage of semantic primes is intended to enable reductive paraphrase of the entire vocabulary and grammar of the language at large, i.e. it is intended to be comprehensive. (Goddard 2002, p. 16).

In my view, such comprehensiveness must indeed be seen as integral to the NSM project, so that *any* degree of final acknowledged empirical failure should be enough to stimulate a revision of its theoretical claims (though not necessarily of its practice). This is a respect in which NSM is quite different from a semantic theory with less universal leanings. It is only because NSM aspires towards universality and comprehensiveness that its proposal to only use the primitives where they work becomes untenable. If the value of the primes is that they underlie all meaning, the theory cannot afford to restrict them to only that subset of meaning for which they actually work. A more exuberant theory of semantic description which did not claim a single metalanguage as the only possible analytical scheme for meaning would be much better able to respond to disonfirming evidence through the adaptation of its paraphrases to linguistic facts. Thus, while other semantic theories are in the same position as NSM, in that disconfirming evidence is not per se a reason for abandoning them, the fact that they are less constrained allows them more agility in responding to new facts: different words

can always be chosen to escape problems. NSM, by contrast, inherently opts for an all-or-nothing degree of confirmability. Restricting its applicability to only parts of semantics should not therefore be an option.

5. Conclusion

The preceding sections have called into question some basic aspects of NSM. It has been argued that NSM's requirement that a definiens be simpler than a definiendum represents a misunderstanding of the nature of semantic explanation (Section 3.1) and that its use of canonical contexts does not fix the meaning of primitives in the required way (Section 3.2). Several features common to both NSM and other varieties of semantic theory based on reductive paraphrase have also been criticized. These included the principle of substitutability (Section 4.1), the treatment of polysemy (Section 4.2), and the modes of response to disconfirming evidence (Section 4.3). In all cases it was suggested that NSM lacks precisely the methodological rigour and objectivity to which it lays claim as part of a 'scientific' linguistics, and that any program sharing its assumptions and goals is equally problematic.

   The fact that, on the arguments of this article, the enterprise of semantic description in general is thoroughly subjective should not cause the value of descriptive semantic studies like those advanced in NSM to be questioned, unless, that is, they make claims of methodological superiority they are unable to sustain. Descriptive semantic studies and theories are as useful as the representations they allow of their subject matter. The explanatory attractiveness of NSM-style primitive-based definition comes from the same source as that of other types of natural language definition: it reduces the amount of arbitrariness in the lexicon by setting up relations of dependency between different words, and responds to the intuition of compositionality – the intuition that the meanings of words correspond to the combinations of the meanings of other words. These intuitions can be served just as well, however, by a different metalanguage, even a metalanguage characterized by technical terms, circularity, and other qualities condemned in NSM. To the extent that the definitions of such a metalanguage can be incorporated into an explanation of the object language, they should not be dismissed.

The criticisms levelled here against NSM, as noted at the outset, leave its definitional practice intact: the present arguments should not dissuade anyone from framing definitions using the NSM primitives. The only changes which, if accepted, they impose upon the theory affect its conception of the significance of the paraphrase method and the acceptance of its claimed universals. In this respect, the effect of the present considerations would be simply to justify an attitude that many of those who have contributed to NSM investigations might well have already adopted: that NSM, like many other linguistic models, in defining a constrained metasemantic vocabulary, fulfils a valuable instrumental function as a framework for the description of meaning, and that the use of this metalanguage is entirely independent of whether or not one accepts the associated claims for the primitives themselves. Representation of meaning, in other words, is an entirely different activity from its explanation. Any semantic theory which claims, like NSM, that meaning is uniquely constituted by the elements which serve to represent it is, therefore, ignoring the fact that a metalanguage is a tool designed to serve specific ends, and that as the ends are different, so different tools will be appropriate.

REFERENCES

Aitchison, J.: 1994, *Words in the Mind: An Introduction to the Mental Lexicon*, Blackwell, Oxford, 2nd edition.

Allan, K.: 2001, *Natural Language Semantics*, Blackwell, Oxford.

Barker, C.: 2003, 'Paraphrase Is not Enough', *Theoretical Linguistics* **29**, 201–210.

Bohnemeyer, J.: 2003, 'NSM Without the Strong Lexicalization Hypothesis', *Theoretical Linguistics* **29**, 211–222.

Cattelain, E. E. J.: 1995, 'Must a Universal Semantic Metalanguage Be Composed of Primitives?', *Pragmatics and Cognition* **3**, 159–179.

Conklin, H.: 1964, 'Hanunóo Color Categories', in D. H. Hymes (ed.), *Language in Culture and Society*, Harper and Row, New York, pp. 189–192.

Croft, W.: 1998, 'Linguistic Evidence and Mental Representations', *Cognitive Linguistics* **9,** 151–173.

Doi, T.: 1981, *The Anatomy of Dependence*, Kodansha, Tokyo.

Durst, U.: 2003, 'The Natural Semantic Metalanguage Approach to Linguistic Meaning', *Theoretical Linguistics* **29**, 157–200.

Fillmore, C.: 1971, 'Types of Lexical Information', in D. D. Steinberg and L. A. Jakobovits (eds.), *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*, CUP, Cambridge, pp. 370–392.

Fodor, J. A., M. F. Garrett, E. C. T. Walker and C. H. Parkes: 1980, 'Against definitions', *Cognition* **8**, 263–367.

Geeraerts, D.: 1993, 'Vagueness's Puzzles, Polysemy's Vagaries', *Cognitive Linguistics* **4**, 223–272.

Goddard, C.: 1991, 'Testing the Translatability of Semantic Primitives into an Australian Aboriginal Language', *Anthropological Linguistics* **33**, 31–56.

Goddard, C.: 1994, 'Semantic Theory and Semantic Universals' in C. Goddard and A. Wierzbicka (eds.), *Semantic and Lexical Universals: Theory and Empirical Findings*, Benjamins, Amsterdam, pp. 7–29.

Goddard, C.: 1998, 'Bad Arguments Against Semantic Primitives', *Theoretical Linguistics* **24**, 129–156.

Goddard, C.: 2002, 'The Search for the Shared Semantic Core of All Languages', in C. Goddard and A. Wierzbicka (eds.), *Meaning and Universal Grammar – Theory and Empirical Findings*, Vol. 1, Benjamins, Amsterdam, pp. 5–41.

Goddard, C. and A. Wierzbicka: 1994, 'Introducing Lexical Primitives' in C. Goddard and A. Wierzbicka (eds.), *Semantic and Lexical Universals: Theory and Empirical Findings*, Benjamins, Amsterdam, pp. 31–54.

Jackendoff, R.: 1983, *Semantics and Cognition,* MIT, Cambridge, MA.

Jackendoff, R.: 1990, *Semantic Structures*, MIT, Cambridge, MA.

Koptjevskaja-Tamm, M. and I. Ahlgren: 2003, 'NSM: Theoretical, Methodological and Applicational Problems', *Theoretical Linguistics* **29**, 247–262.

Kripke, S.: 1980, *Naming and Necessity*, Blackwell, Oxford.

Lakoff, G.: 1987, *Women, Fire and Dangerous Things*, University of Chicago Press, Chicago.

Matthewson, L.: 2003, 'Is the Meta-Language Really Natural?', *Theoretical Linguistics* **29**, 263–274.

Miller, G. A. and C. Leacock: 2000, 'Lexical Representations for Sentence Processing', in Y. Ravin and C. Leacock (eds.), *Polysemy*, CUP, Cambridge, pp. 152–160.

Murray, D. W. and G. Button: 1988, 'Human Emotions: Some Problems of Wierzbicka's 'Simples'', *American Anthropologist* **90**, 684–686.

Nuyts, J.: 1993, 'Cognitive Linguistics', Review Article of Lakoff 1987 and Langacker 1987, *Journal of Pragmatics* **20,** 269–290.

Riemer, N.: 2003, 'Semantic Evidence and the Mental Representation of Polysemy', in P. Slezak (ed.), *Proceedings of the Joint International Conference on Cognitive Science,* University of New South Wales, Sydney.

Riemer, N.: 2005, *The Semantics of Polysemy: Reading Meaning in English and Warlpiri*, Mouton de Gruyter, Berlin and New York.

Riemer, N.: forthcoming, 'Syntactic Optionality and Sense Individuation', in M. Amberber and N. Riemer (eds.), *Selected Linguistics Papers from the 2003 Joint International Conference on Cognitive Science*, Elsevier, Amsterdam.

Weinreich, U.: 1966, 'On the Semantic Structure of Language', in J. Greenberg (ed.), *Universals of Language*, MIT Press, Cambridge, MA, pp. 142–216.

Wierzbicka, A.: 1972, *Semantic Primitives*, Athenäum, Frankfurt.

Wierzbicka, A.: 1980, *Lingua Mentalis: The Semantics of Natural Language*, Academic Press, Sydney.

Wierzbicka, A.: 1985, *Lexicography and Conceptual Analysis*, Karoma, Ann Arbor.

Wierzbicka, A.: 1987, *English Speech Act Verbs: A Semantic Dictionary*, Academic Press, Sydney.

Wierzbicka, A.: 1988a, *The Semantics of Grammar*, Benjamins, Amsterdam.

Wierzbicka, A.: 1988b, 'Semantic Primitives: A Rejoinder to Murray and Button'. *American Anthropologist* **90**, 686–689.

Wierzbicka, A.: 1991, *Cross-Cultural Pragmatics. The Semantics of Human Interaction*, Mouton de Gruyter, Berlin, New York.

Wierzbicka, A.: 1992, *Semantics, Culture and Cognition; Universal Human Concepts in Culture-Specific Configurations*, OUP, New York.

Wierzbicka, A.: 1996, *Semantics. Primes and Universals*, OUP, Oxford, New York.

Wierzbicka, A.: 1999, *Emotions Across Languages and Cultures*, Cambridge/Paris, CUP/Editions de la Maison des Sciences de l'Homme.

Department of English
University of Sydney
Woolley Building A20
Sydney NSW 2006
Australia
nick.riemer@arts.usyd.edu.au