

BUILDING A UNIVERSAL SEMANTIC  
METALANGUAGE:  
THE SEMANTIC THEORY OF ANNA WIERZBICKA

by  
Cliff Goddard

For some thirty years now, Anna Wierzbicka has been one of the most prolific, insightful, and lively scholars in the field of linguistic semantics. Her books and articles have ranged over diverse areas of lexical semantics, grammatical semantics, and pragmatics. At the level of theory, she is widely known for her insistence that universal semantic primitives exist as meanings of words in ordinary language. In recent years, her theory – now known as the 'natural semantic metalanguage' (NSM) approach – has undergone considerable expansion and modification. This article presents an overview of current NSM theory, covering the expanded inventory of primitives, the novel concepts of allolexy and non-compositional polysemy, and new proposals about the syntax of the semantic metalanguage.

*1. Introduction*

The main principles of Anna Wierzbicka's approach to semantic representation have remained constant since the publication of her 1972 book *Semantic Primitives*. The foundational postulate is that semantic analysis can be both discrete and exhaustive; i.e., that any complex meaning can be decomposed without circularity and without residue into a combination of discrete other meanings. It follows that there must be a finite set of indefinable meanings (semantic primitives) which are the terminal elements of semantic analysis. In the tradition of Descartes, Pascal, Arnauld, and, above all, Leibniz, Wierzbicka views semantic primitives as the simple concepts in which articulate thought is composed; in Leibniz's phrase, 'the alphabet of human thought'.

From these assumptions it follows that semantic analysis must be conducted through paraphrase in natural language. Technical neologisms, logical formalisms and other artificial symbols are no good as terms of semantic analysis because their meanings are not clear until and unless they are explained in simple, ordinary language. Semantic primitives, then, can be seen as the 'indefinable words' of ordinary natural language, the words one would be left with after an exhaustive reductive paraphrase analysis of the entire lexicon. Aside from its intrinsic interest, Wierzbicka's huge output of empirical semantic studies (Wierzbicka 1980, 1985, 1987, 1988, 1991a, 1992a, among

others) can be seen as an attempt to establish, test, and refine a viable set of semantic primitives.

The first set of primitives numbered a mere 14 elements (Wierzbicka 1972). In 1980 it was expanded slightly to 15, and several other elements were mentioned as possible candidates for future inclusion. The lexicon of semantic primitives remained in this relatively austere state until the 'expanding set' phase began in the late 1980s (see §2 below). The current inventory numbers 55-60 elements. The claim is that using this tiny lexicon one can frame reductive paraphrase explications for all the lexical items in any language, as well as for all language-specific grammatical constructions.

This way of phrasing Wierzbicka's semantic ideals draws attention to a second goal of her program, namely, the pursuit of universality. Though always present in her thinking, this goal came increasingly to the fore in the second half of 1980s as her colleagues and graduate students at the Australian National University, many of whom were engaged in field-work on non-European languages, wrestled with the translatability of some of the proposed semantic primitives of that time. For Wierzbicka, any serious challenge to the translatability of her system is a matter of the utmost gravity. For although it is derived mainly from deep semantic investigations into European languages, hers is a hypothesis about the fundamental nature of human language and cognition. Recent times have seen concerted (and largely successful) efforts to identify exponents of semantic primitives in other languages, especially non-European languages. §3 and §4 give an account of new analytical concepts (allolexy and non-compositional polysemy) which have emerged from this exercise.

In the latter part of the 1980s, Wierzbicka's attention also began to shift from the lexical to the syntactic aspects of her system. She began to emphasise that what she was after was not merely a mini-lexicon but a complete 'mini-language' for semantic analysis. The semantic primitives which comprise the vocabulary of the system must have a universal combinatorial syntax which could and should be specified. To reflect this new emphasis, the term 'natural semantic metalanguage' (NSM) was adopted. §5 reports on developments in the syntax of the natural semantic metalanguage. This includes the concept that primitives may have alternative 'valency options'; for instance, that the element DO may occur either with a subject ('actor') alone or with a second ('patient') argument as well (in English, as DO TO). This notion represents, to some extent, a weakening of assumptions of earlier work,

but it allows the resolution of some persistent difficulties and at the same time opens the way for a new semantically-inspired approach to grammar construction.

Notwithstanding these innovations, the fundamental claim of Wierzbicka's program has remained constant over her thirty years of semantic work. It is the claim that the semantic cores of all human languages coincide – that from any language one can 'carve out' a minimal representational system of indefinable lexical elements with specific combinatorial properties and that all such systems are fundamentally isomorphic. One can therefore speak of a single 'natural semantic metalanguage' (NSM) with numerous versions, one in each of the world's languages.

## 2. *The expanding NSM lexicon*

It would be difficult to go over all the reasons for the initial expansion of the NSM lexicon, which began in 1989 with the publication of Wierzbicka's 'Expanding Set' paper as part of a *Quaderni di Semantica* 'Roundtable' on semantic primitives. To some extent Wierzbicka had been provoked by Goddard's (1989) contribution to the same roundtable, but it is apparent from the scope of her article that the ideas expressed in it were the fruit of years of thought, semantic seminars, and debates with absent colleagues, especially Andrzej Bogusławski.

In the 'Expanding Set' article she revised earlier analyses of time, place, knowledge, causality, evaluation, and action. In place of her well-known, but counter-intuitive, analysis of tense in terms of 'worlds becoming worlds' she installed a stative temporal primitive WHEN and a relational temporal primitive AFTER, to provide for temporal sequence. In place of her earlier analysis of location in terms of 'parts of the world' she adopted the spatial primitive WHERE (OR BE SOMEWHERE). Rejecting an earlier analysis of 'X knows this' as 'X can say this' and of 'A because B' as 'if not B, then not A', she admitted both KNOW and BECAUSE into the stable of semantic primitives. She also abandoned earlier attempts to explicate GOOD and BAD (in terms of wanting) and DO (in terms of happening and wanting). Other new additions were the quantificational elements TWO and ALL, the concepts of similarity and identity (LIKE and THE SAME), and the intensifier VERY.

One interesting aspect of this expansion phase was the realisation that, paradoxical as it seems, it is necessary to have pairs of opposites and converses as primitives. For example, GOOD and BAD are both needed because it is impossible to explicate either in terms of the other ('not good' ≠ 'bad'; 'not bad' ≠ 'good').

Over the next few years, still more primitives were proposed (cf. Wierzbicka 1992b) till by 1993, the list had stabilised at 39, no less than triple the size of the original inventory. It was this set of 39 primitives whose cross-linguistic viability was tested – and on the whole upheld – in the *Semantic and Lexical Universals* project (Goddard and Wierzbicka Eds, 1994). As can be seen from Table One, it was now possible to arrange the primitives into groups of similar elements and to see the relation between these groupings and conventional syntactic and semantic categories. The 'inventory' began to seem more and more like a language in miniature.

**Table One: Proposed semantic primitives**  
(Goddard and Wierzbicka Eds, 1994)

Substantives:	I, YOU, SOMEONE/PERSON, PEOPLE, SOMETHING/THING
Mental predicates:	THINK, SAY, KNOW, WANT, FEEL
Speech:	SAY
Actions, events:	DO, HAPPEN
Determiners/quantifiers:	THIS, THE SAME, OTHER, ONE, TWO, MANY/MUCH, ALL
Evaluators, descriptors:	GOOD, BAD, BIG, SMALL
Time/place:	WHEN/TIME, AFTER, BEFORE, WHERE/PLACE, UNDER, ABOVE
Metapredicates:	NO, IF, CAN, LIKE, BECAUSE, VERY
Taxonomy, paronymy:	KIND OF, PART OF
Similarity:	LIKE

The publication of the *Semantic and Lexical Universals* (SLU) volume was a watershed in the development of Wierzbicka's framework. For the first time a strictly comparable set of studies established that it is possible to find clear exponents for almost all the proposed set of semantic primitives in a sample of typologically and genetically diverse languages.

The languages and authors involved were as follows: Acehnese (Mark Durie, Bukhari Daud and Mawardi Hasan), Arrernte (Jean Harkins and David Wilkins), Ewe (Felix Ameka), French (Bert Peeters), Japanese (Masayuki Onishi), Longgu (Deborah Hill), Kalam (Andrew Pawley), Kayardild (Nicholas Evans), Mandarin Chinese (Hilary Chappell), Mangap-Mbula (Robert Bugenhagen), Misumalpan (Ken Hale), Samoan (Ulrike Mosel), Thai (Anthony Diller), Yankunytjatjara (Cliff Goddard). Unless otherwise indicated, all statements about these languages adduced in this paper come from the relevant chapters of the SLU volume.

As often happens during periods of rapid theoretical development, however, by the time *Semantic and Lexical Universals* appeared in print, work in press was entertaining the possibility of yet another two dozen primitives. In *Semantic, Primes and Universals* (1996) the number of proposed semantic primitives had reached 53 in number; see Table Two. A small set of further candidates – DIE, FOR SOME TIME, TRUE, and BODY – is being investigated in current work.

**Table Two: Proposed semantic primitives**  
(Wierzbicka 1996)

Substantives:	I, YOU, SOMETHING/THING, SOMEONE/PERSON, PEOPLE
Mental predicates:	THINK, KNOW, WANT, FEEL, SEE, HEAR
Speech:	SAY, WORDS
Actions, events, movement:	DO, HAPPEN, MOVE
Existence and Life	THERE IS, LIVE
Determiners:	THIS, THE SAME, OTHER
Quantifiers:	ONE, TWO, ALL, MANY/MUCH, SOME
Evaluators, descriptors:	GOOD, BAD, BIG, SMALL
Time:	WHEN/TIME, AFTER, BEFORE, A LONG TIME, A SHORT TIME
Space:	WHERE /PLACE, ABOVE, BELOW; FAR, NEAR, SIDE, INSIDE
Logical concepts:	IF, NOT, CAN, IF ... WOULD, BECAUSE, MAYBE
Intensifier, augmentor:	VERY, MORE
Taxonomy, paronymy:	KIND OF, PART OF
Similarity:	LIKE

Once again, the additions were not made willy-nilly, but as a result of careful consideration of recalcitrant semantic problems. Limitations of space permit us here to review the 'new primitives' in only the briefest of terms. Wierzbicka had long maintained that 'see' and 'hear' were definable in terms of KNOW, and 'eyes' and 'ears', respectively. Essentially, that 'see' = 'come to know using the eyes' and 'hear' = 'come to know using the ears' (Wierzbicka 1980). But at the same time it was recognised that there were difficulties with this. It meant that 'eyes' and 'ears' had to be explicated not as 'sense organs', but purely in physical and positional terms (e.g. 'eyes' = 'two small movable parts of the upper face'), which was counter-intuitive, and which also meant that the full semantic structures for 'see' and 'hear' would be very complex, once all the necessary detail was substituted in. Admitting SEE and HEAR as semantic primitives overcomes these difficulties.

Another (provisional) new primitive is WORD(S), which supplements SAY in the domain of language and speech. In its favour are, on the one hand, problems with explicating 'word' itself (as in 'she said a bad word', or 'he said two words'), and on the other, its apparent indispensability for adequate explication of swearing, magical and 'quasi-magical' speech, and performative verbs, as well as concepts such as 'language' and 'paraphrase'.

In the early days Wierzbicka had explicated existence as 'being part of the world', but this option closed off when 'world' was struck from the primitives list. Against the earlier idea that motion can be defined as 'change of place over time' is the fact that 'change' is no more straightforward than 'motion' itself, and also that not all movement can plausibly be regarded in this way (for instance, if I move my hands I am moving, but it seems counter-intuitive to see this in terms of my hands changing locations). In response to these problems, THERE IS and MOVE were added to the list.

The element LIVE is a very recent addition. The difficulty of defining the concept of 'life' has long been recognised in philosophy, science and medicine. It might seem that DIE can be explicated as 'cease to be alive', but this runs into problems because the concept of 'life after death' is not a contradiction, as this putative explication of DIE would suggest; space does not permit us to pursue such issues here.

The temporal and spatial domains have both been enriched in the latest versions of the NSM lexicon, firstly, by the addition of the deictics NOW and HERE. For a long time it was assumed that these could be explicated in terms of THIS and temporal/spatial reference to the

speech act situation ('now' = 'when I say this', 'here' = 'where I say this'), but these explications no longer seem satisfactory, if only because the range of application of NOW and HERE is typically wider than the proposed explications would predict. The temporal domain has also been expanded by the time-period expressions A LONG TIME and A SHORT TIME, and by the durational primitive FOR SOME TIME.

The spatial domain has been expanded by the distance expressions FAR and NEAR, and the relational notions of SIDE (laterality), INSIDE (containment), and, very tentatively, ON (contact). These additions give the metalanguage greatly improved capacity for dealing with concrete objects and actions. The quantificational realm has been elaborated by the addition of an 'indeterminate number' primitive SOME OF. The augmentor element MORE is also new.

In the domain of modality, it has been recognised that the primitive CAN was being used in more than one way in earlier work. The new primitive MAYBE installs something like 'possibility' into the system, leaving CAN to cover 'potentiality', cf. Bolinger (1989). Finally, a new bi-clausal 'counterfactual conditional' construction IF ... WOULD has been installed alongside the plain conditional IF, acknowledging that counterfactuals (e.g. 'if we hadn't found water, we would have died') resist explication in terms of the conditional alone.

All in all, we see a further move toward a mini-language of semantic representation which more closely resembles full natural languages. As Wierzbicka (1996) observes, the expanded lexicon reflects a much more 'embodied' and concrete perspective than the ancestral set of 14. One can now see clear correspondences between the proposed semantic fundamentals and the major lexical classes and grammatical categories found in the world's languages.

For example, though the universality of personal pronouns was reflected in the presence of I and YOU in the very earliest primitive inventories, the addition of PEOPLE and LIVE allows the categories 'human' and 'animate', so frequently invoked by typologists, to be readily constructed within the NSM metalanguage. Along with cognition (THINK, KNOW) and illocution (SAY), the new list includes predicates of perception (SEE, HEAR), all of which are often grammaticalised into evidential systems (Wierzbicka 1994b). The emotion/sensation primitive FEEL often finds its way into special experiencer constructions. The temporal primitives are often grammaticalised into tense and aspect systems, and the spatial primitives can be found grammaticalised in the elaborate systems of

locational deixis found in North and Central America (cf. Goddard In press a).

Has the 'expansion phase' of the NSM primitive inventory yet come to an end? Probably not, although we can be fairly sure that the rate of any future expansion will be slower than we have seen lately. Though, as mentioned, there are several expressions currently under consideration as additional primitives, the current inventory may also be trimmed back as a consequence of cross-linguistic testing of the newer primitives. Above all, it must be borne in mind that such modifications, though they may puzzle or frustrate some observers, are inevitable in view of the empirical character of the NSM program. It would be suspicious if the hypothesis about the shared lexical core of all languages did not change in response to continued semantic analysis and cross-linguistic investigation.

To conclude this review of the NSM lexicon, it should be pointed out that over the expansion phase Wierzbicka's conception of primitive meanings has changed in one important respect. Like Leibniz (Ishiguro 1972: 45-46), she used to believe that primitives should be intuitively independent, because – so it seemed – if two concepts have something in common they could not both be semantically indivisible. The changing shape of the primitive inventory has forced her to revise this stance. There are many intuitive connections among primitives which cannot be 'factored out' as discrete components. (Actually, this was apparent to some extent even in the very earliest lists: I and YOU are not independent of SOMEONE, though it was well-established (Sørensen 1963; Benveniste 1971; Wierzbicka 1976) that none of them could be defined away in terms of the others.)

Most obvious are the affiliations between pairs of opposites, such as GOOD and BAD, NEAR and FAR, and pairs of converses, such as ABOVE and BELOW, BEFORE and AFTER. Though they are related, it is impossible to extract any 'common feature' from such pairs. Abstract terms such as 'evaluation', 'distance', 'vertical orientation' and 'temporal sequence' are plainly more obscure and difficult to translate across languages than simple basic terms like GOOD, BAD, NEAR, FAR, and so on, and could not possibly replace the latter in a plausible universal semantic metalanguage. Opposites and converses, however, are only the beginning. There are all sorts of other, subtler non-compositional relationships as well (for instance, between SOMEONE and OTHER, WANT and SAY, IF and MAYBE), some of which we will consider in §4.

To recognise the existence of non-compositional relationships is to recognise that paraphrase semantics has its limits. In particular, it cannot account for all incompatibility. It is easy enough to show, for instance, that a contradictory sentence like 'I am unhappy and joyful' contains the components 'I feel something bad' and 'I feel something good', but compositional semantics does not spell out the fact that they are incompatible. That is, compositional semantics does not tell us that 'X is good' implies 'X is not bad', that 'Y is big' implies 'Y is not small', that 'A is above B' implies 'A is not under B', and so on. Putting this point another way, there is more to antonymy than the Law of the Excluded Middle. Of course, there are all sorts of other incompatibilities between primitives as well. For instance, ONE cannot be considered MANY; a PART OF something cannot be a KIND OF something.

If we wanted to train a computer to detect semantic incompatibility, compositional semantics (and the Excluded Middle) would not suffice. We would have to give it a whole series of additional instructions certifying relationships between primitives. (These could be seen as akin to Fodor's 'meaning postulates'). Depending on one's point of view, this may or may not seem a serious problem. It would be serious if the goal were to create a purely formal calculus which could be used for 'automatic reasoning', one of Leibniz's aspirations, and one shared by the contemporary Artificial Intelligence community. It is not serious, however, if the goal is to create a metalanguage for human beings to understand one another, i.e. for a humanistic semantics. No speaker of any language needs to be told that GOOD is not BAD, or that ONE is not MANY, and so on. Knowing these relationships comes with knowing the meanings themselves.

### 3. *Allolexy*

The *Semantic and Lexical Universals* project furnished a wealth of experience both about the formal expression of semantic primitives in typologically different languages and about the methodology required for rigorous cross-linguistic matching of meanings. One important tool developed for this work was a schedule of 'canonical sentences', i.e. contexts in which each lexical primitive was predicted or expected to be able to occur universally (see §5 below). Of course, it has to be stressed that the term 'lexical' is here used in a broad sense. A good exponent of a primitive meaning may be a word, a bound morpheme, or even a

phraseme, just so long as it expresses the requisite meaning and has the requisite combinatorial properties. Similarly, exponents of the same primitive may belong to different parts of speech in different languages; for example, KNOW is a verb in English but an adjective in Yankunytjatjara, TWO is an adjective in English but a verb in Samoan.

The concept of allolexy was foreshadowed as early as 1980, when Wierzbicka observed that though the expression 'this someone' was rather odd the same meaning seemed to be expressed by the phrase 'this person'. Accepting this, she noted: 'The language of explications cannot be kept entirely free of contextual variation. For example, to assure the readability of English-based explications both the forms 'me' and 'I' have to be used, standing for the same primitive' (1980: 36). Much later (Wierzbicka 1989a) the term 'allolexy' was coined to designate situations in which several different word-forms (allolexes) in complementary contexts express a single meaning. Initially, two types of allolexy were recognised.

Positional allolexy is illustrated by English *I* and *me*. It is impossible to state any semantic difference between them (*I* in *I did something* or *I don't know* has the same meaning as *me* in *you did something to me* or *people might say something about me*), and their distribution is determined solely by position, i.e. *I* pre-verbally and *me* elsewhere. Combinatorial allolexy is illustrated by the relationship between *something* and *thing*. In combination with determiners and quantifiers, *thing* functions as an allolex of SOMETHING: i.e. *this something* = *this thing*, *the same something* = *the same thing*, *two somethings* = *two things*, and so on. (The situation with *someone* and *person* is not clear. For some time, it was held that *person* is a combinatorial allolex of SOMEONE, i.e. *this someone* = *this person*, *the same someone* = *the same person*, etc., but against this is the fact that *person* also functions as the singular or 'individuated' form of PEOPLE, cf. *one person*, *many people*.)

Other types of combinatorial allolexy turned up in other languages. Both in French and Japanese, for instance, two forms expound the semantic primitive THIS. One is found where THIS functions as a 'quasi-substantive' making an anaphoric reference to a situation spelt out in previous clause (for instance, to say 'this is good' or 'this is bad') – Japanese *kore*, French *cela*. The other is found where THIS functions as a nominal modifier (for instance, to say 'this person' or 'this thing') – Japanese *kono*, French *ce*.

Yankunytjatjara illustrates case allolexy. The Yankunytjatjara verbs *kulini* THINK, *wangkanyi* SAY and *palyani* DO select ergative case subjects,

while *mukuringanyi* WANT (and other elements like *ninti* KNOW and *-tjara* HAVE AS PART) select nominative case subjects. Clearly, the choice of ergative vs. nominative case does not carry any independent meaning in these contexts. If it did, a word like SOMEONE would have different meanings in Yankunytjatjara expressions corresponding to 'someone did ...' and 'someone wants ...'. It would also mean that there would be a meaning difference between a sentence composed in Yankunytjatjara primitives and the corresponding sentence composed in English primitives. We can conclude that the ergative case-forms are allolexic variants of their unmarked counterparts. Similar arguments apply to accusative case-forms in NSMs based on many European languages.

It might appear that the conditioning factor is whether the argument is functioning as an A, an S, or an O, but this explanation is not available since notions like 'A', 'S', and 'O' are not applicable to the semantic metalanguage. Conventional definitions of these notions (e.g. Dixon 1979, Andrews 1985: 68-69) rely on making reference to prototypical or primary transitive verbs, but THINK and SAY (though complement-taking) are not prototypically transitive. It is therefore necessary to regard case allolexy as lexically conditioned – that is, as combinatorial allolexy determined by the identity of the verbal lexeme involved. Significantly, a survey of languages will show considerable variation in case marking so far as the NSM 'predicates' are concerned; for instance, in Arrernte, another Central Australian Aboriginal language, the exponents of THINK and of SAY select nominative, rather than ergative, case.

The concept of allolexy can resolve a difficulty for NSM semantics which has puzzled many observers, namely, how to reconcile the existence of language-specific inflectional categories with the claim that there can be perfect translatability between the NSMs based on different languages. Consider obligatory tense-marking in English. In an English sentence like *I did something*, as in (1a) below, the word *did* is semantically complex, conveying both DO and past tense. But notice what happens if we 'paraphrase out' the semantic content of past tense, which we may suppose to be 'at some time before now', as in (1b). In this context, choice of the form *did* as opposed to *do* becomes automatic and allolexic, since it would be ungrammatical to exchange the forms. This situation can be designated 'inflectional allolexy'.

- 1a. I did something.
- 1b. At some time before now, I did/\*do something.

This highlights the difference between a full natural language, in this case English, and the natural semantic metalanguage based upon it. In ordinary English, *do* and *did* are semantically distinct, but in an English-based NSM they are allomorphs because in the NSM time-reference will always be independently represented in lexical terms.

The full description of any NSM will include a tabulation of the semantic content of all the inflected word-forms of the language on which the NSM is based, somewhat in the style of the Word-and-Paradigm approach to inflection. Given this, full translatability between NSMs can be assured. This can be illustrated by comparing the (a) and (b) sentences below, which show how three sample sentences composed of identical primitives would be expressed in English and Malay (Bahasa Melayu). In these examples, the elements occur in the same order in the two languages. (The only difference is that Malay has the connective *yang* intervening between SOMETHING *sesuatu* and BAD *buruk*, in the expression 'something bad' *sesuatu yang buruk*.)

Malay has no obligatory inflectional categories, so Malay-based NSM has no inflectional allomorphy. Thus, English *did* in (2a) and *do* in (3a) both correspond to a single word-form *buat* DO in (2b) and (3b). Malay pronouns are invariable in form, so the Malay NSM lacks the English *I/me* positional allomorphy. Thus, English *I* in (2a) and *me* in (3a) both correspond to the single word-form *aku* in (2b) and (3b). Negation in Malay is expressed by a single word-form *tak* (or, in more formal speech *tidak*) whereas English *no*, *not*, and *don't* (among others) are all allomorphs of negation. Thus, English *don't* and *no* in (4a) both correspond to *tak* in (4b). Despite these differences in formal realisation, it should be apparent that there is no problem with inter-translatability.

- 2a. At a time before this, I did something bad.  
 2b. *Pada waktu sebelum ini, aku buat sesuatu yang buruk.*
- 3a. If I do this, people will think something bad about me.  
 3b. *Kalau aku buat ini, orang akan fikir sesuatu yang buruk pasal aku.*
- 4a. I don't want to say 'no'.  
 4b. *Aku tak mahu kata 'tak'.*

A final kind of allomorphy may be called portmanteau allomorphy, where a single word (morpheme, phraseme) expresses a combination of

primitive meanings. A simple example is English *can't*, combining CAN+NOT. It seems that many languages have portmanteau allomorphs involving negation, sometimes combining it with several other elements, as for instance in 'negative imperative' words such as Latin *noli*, which are amalgams of NOT WANT with YOU DO. In Yankunytjatjara, the word *alatji* is a portmanteau for LIKE THIS, equivalent to *nyanga purunypa*. Again, it seems many languages have such allomorphs, or even more specialised ones for combinations such as DO LIKE THIS, HAPPEN LIKE THIS, OR SAY LIKE THIS. Portmanteaus with DO are also common; for example, the Australian language Kayardild has portmanteaus for 'do this', 'do well', 'do badly', and 'do like someone else'.

It may be that in some languages certain combinations of primitive meanings are expressible only via a portmanteau. In Polish, for example, there is a common expression *dawno (temu)* 'a long time ago', which presumably means A LONG TIME BEFORE NOW. But although exponents exist for the components of this complex expression (*dlugo* A LONG TIME, *przed* BEFORE, *teraz* NOW), it is extremely odd to combine them; *\*dlugo przed teraz* sounds much worse than 'this someone' in English.

#### 4. Non-compositional polysemy

Unlike allomorphy, polysemy is hardly a novel concept, but it has found a new application in the quest to identify exponents of semantic primitives. At the simplest level, it is always necessary to be aware that exponents of the same primitive may be polysemous in different ways in different languages. For example, the Yankunytjatjara exponent of WANT *mukuringanyi* has secondary meanings approximating English 'like, be fond of' and 'need', so that its range of use does not correspond to that of English *want*. Only after the polysemy is sorted out and justified on language-internal grounds does the equivalence of the primary meanings of *want* and *mukuringanyi* become clear (Goddard 1991).

Many polysemies of this type recur in unrelated languages. To mention only a few: HAPPEN with 'appear' or 'arrive' (Mangap-Mbula, Ewe, French), DO with 'make' (Misumalpan, Arrernte, Samoan, Kalam), SAY with 'speak' or 'make sounds' (Thai, Mandarin, Yankunytjatjara, Kalam), BEFORE with 'first', 'go ahead', or 'front' (Samoan, Kayardild, Ewe); FEEL with 'taste', 'smell', or 'hold an

opinion' (Acehnese, Ewe, French, Mandarin, English). The basis for such polysemies is seldom difficult to discern (cf. Wierzbicka 1994a).

Rigorous semantic analysis sometimes paints a picture of polysemy which challenges preconceptions about the literal and the figurative (cf. Goddard 1996). For example, in some languages the exponent of FEEL or of WANT turns out to be a word which can also have a body-part meaning, whether specific, such as 'belly' or 'guts' (Yankunytjatjara, Kalam), or general, such as 'insides' (Mangap-Mbula). In such cases, contrary to the tradition which would have it that the concrete meaning is 'prior', we must recognise the primitive FEEL or WANT sense as the simpler meaning and the body-part meaning as the complex, extended one.

Curious as these cases may be, from a theoretical point of view there is nothing very exciting about the fact that a word may have two (or more) related meanings, one indefinable and the other definable. Much less expected was the finding, which emerged clearly from the SLU volume, that in some languages a single word-form may express two different indefinable meanings. In Yankunytjatjara and Samoan a single form (*kutjupa* and *isi*, respectively) expounds both SOMEONE and OTHER. In Mangap-Mbula and Kalam a single form (-*so* and -*ak*, respectively) expounds both WANT and SAY. In Yankunytjatjara and Arrernte a single form (*tjinguru* and *peke*, respectively) expounds both MAYBE and IF.

Even from these few examples, it is apparent that such overlaps recur across languages and that they are not semantically arbitrary. IF and MAYBE, for instance, have enough affiliation that it is not jarring to discover that both are expressed by the same element in some languages. (Or at least, not as jarring as it would be to find that IF was expressed by the same form as YOU, TWO, OR BAD!) What should we call such recurrent, semantically motivated match-ups? Since the meanings involved are primitive, we seem to be dealing with distinct words, but the term 'homonymy' is unsatisfactory here because it would suggest an arbitrary coincidence of form. Wierzbicka has therefore chosen to use the term 'polysemy' in an extended sense to register the fact that though there is no compositional relationship, the meanings in question (SOMEONE and OTHER, SAY and WANT, IF and MAYBE, and so on) are associated in non-compositional ways.

Let us look at some examples of non-compositional polysemy among semantic primitives from a methodological point of view. How can it be shown that a single form expresses distinct primitive meanings? Consider the Yankunytjatjara sentences below. In (5) we see

the form *kutjupa* functioning adnominally, with the meaning OTHER. In (6a) and (6b), on the other hand, it is the head of an NP in its own right, with the meaning SOMEONE. Though the same lexical form is used, it occurs in distinct syntactic (combinatorial) contexts, so there can be no confusion about which meaning is intended.

5. *Palu ngayulu munu kungka kutjupa.*  
But I and woman other  
'But (there was) me and another woman.'
- 6a. *Kutjupa-ngku iti katingu!*  
someone-ERG baby takePAST  
'Someone's taken the baby!'
- 6b. *Ngayulu nyangu kutjupa nyara waru-ngka nyina-nytja-la.*  
I seePAST someone there fire-LOC sit-NOML-LOC  
'I saw someone sitting over there by a fire.'

Against this, it might be claimed that *kutjupa* is not a true nominal head in (6a) and (6b) – that there has been ellipsis of an 'understood' categorical head such as *anangu* 'person/human being'. That is, it could be claimed that *kutjupa* in (6a) is short for *anangu kutjupa* and *kutjupangu* in (6b) is short for *anangu kutjupangu*, in both cases meaning 'another person/human being'. However, this analysis is not semantically viable because the referential range of *anangu* is confined to human beings whereas that of nominal *kutjupa* is not. For instance, traditional folktales often tell of babies being stolen away from their families by such creatures as *pangkalangu* 'ogre' and *kungkapanpa* 'bogey woman'. In such tales a distraught mother could say the likes of (6a), but since there is no presumption that a human being is responsible, *anangu kutjupa*(*ngku*) is not substitutable in place of *kutjupa*.

Similarly, example (7) shows *kutjupa* being used to make an indefinite reference which turns out to relate to God. (Many Yankunytjatjara people have adopted Christianity.) Once again, *anangu kutjupa* 'another person/human being' could not be substituted for *kutjupa* here.

7. *Kutjupa-ngku rawangu nyuntunya nyanganyi*  
someone-ERG always youACC seePRES



*munu kulini. Nganalu? Godalu.*  
 and thinkPRES whoERG GodERG  
 'Someone is always watching and thinking about you. Who? God.'

A second example of polysemy among the exponents of semantic primitives is provided by the Samoan verb *fai*. This may expound either SAY or DO, but the two meanings are associated with different morphosyntactic properties. *Fai* SAY is a non-ergative verb, selecting an absolutive S. It can introduce direct or indirect speech, in (8a) and (8b), and combines with noun phrases referring to speech; for example, *fai le upu* 'say the word', *fai le tatalo* 'say a prayer', *fai le tala* 'tell a story, make a report or statement'.

8a. *Ona toe fai atu lea 'o le fafine, "Se ...*  
 then again say DIR then ABS the woman friend  
 'Then the woman said again, "Friend, ..."' (Mosel 1987: 459)

8b. *Na e fai mai au oti?*  
 PAST you say hither PERF die  
 'You said he has died?'

*Fai* DO, on the other hand, is an ergative verb. When it occurs in a transitive sentence as in (9a) it selects an ergative A. As well, *fai* DO often occurs in the so-called long (suffixed) form *fai=a*, which is usual when an ergative verb is preceded by a pronoun. This applies even when *fai* DO is used in a non-transitive frame, i.e. without any explicit O, as in (9b). *Fai* SAY never appears in the long form; it would be ungrammatical to substitute *faia* for *fai* in (8b), for instance.

9a. ... *'ua fa'apênâ lava ona fai e le tama.*  
 PERF like this EMPH that do ERG the youth  
 '... the youth did it like this.' (Mosel 1987: 122)

9b. *'O ai na faia?*  
 PRES who PAST do?  
 'Who did it?'

A more complex example of non-compositional polysemy is found in the polysynthetic Papuan language Kalam, where a single stem *g-* appears to serve as exponent of three primitives, HAPPEN, DO and FEEL.

But once again the three meanings are kept separate by distinct grammatical frames. As an intransitive verb *g-* means HAPPEN, as in (10), whereas as a transitive verb it means DO as in (11). (Transitive *g-* can also mean 'make', but this is a separate matter.)

10. *Tap etp gp?*  
 thing what it.has.happened  
 'What's happened?'

11. *Tap tmey gpay.*  
 thing bad they.have.done  
 'They have done bad things.'

The primitive FEEL is expressed in an 'experiencer construction' which takes the form: Experiencer – Condition – Verb+tense+3sg. The experiencer appears as noun or free pronoun with objective case-marking. It is illustrated in (12a) and (12b) (Pawley 1994: 407-8; Wierzbicka 1994a: 462; data from Pawley).

12a. *Yp tep gp.*  
 me good it.feels  
 'I feel good.'

12b. *Yp yuwt gp.*  
 me pain it.feels  
 'I feel pain.'

Examples (10)–(12) show three distinct grammatical constructions (intransitive, transitive, experiencer) in which the same stem *g-* functions as an exponent of HAPPEN, DO and FEEL. It is therefore valid to say (as Pawley does) that these meanings do not 'belong' solely to the stem *g-*, but rather to *g-* in each of the specified constructions. But equally, and more importantly for the viability of a universal meta-language, it is possible to unequivocally express in Kalam meanings like 'something good happened', 'they did something good', and 'I feel good'.

One may be inclined to ask: How can it be shown that *g-* does not in fact have a single 'abstract' meaning, more abstract than any one of DO, HAPPEN, or FEEL? On the assumptions of the NSM approach, this is a non-question. One cannot show anything about hypothetical meanings

which are so abstract that they cannot be stated. This is not to say that the meanings have nothing to do with one another and that it is just coincidence that all three are expounded in Kalam (and in some other Papuan languages, cf. Foley 1986: 119) by the same 'lexical' element. But again, the affiliations are not compositional.

Though research on these matters is only just beginning, it is possible to draw some distinctions between different kinds of non-compositional relationships, and to investigate their role in polysemies between exponents of semantic primitives. It should be noted, however, that regardless of non-compositional affiliation, primitive meanings cannot be expounded by a single lexical form unless they possess sufficiently distinctive syntactic properties. This is presumably one reason why pairs of opposites and converses cannot be expounded by the same forms, but also (and more interestingly) it is one reason why I, YOU and SOMEONE are kept lexically distinct in all languages. As Wierzbicka (1994a: 450) remarks: 'patterns of polysemy of this kind could lead to unresolvable ambiguities on a large scale'. The same presumably applies to the quantifiers ONE, TWO, MANY, and ALL, and to the classifiers KIND OF and PART OF. The syntactic properties of THINK, KNOW and SAY may also be too close for them to be expressed by a single lexical item.

The strongest non-compositional relationships can be termed 'entailment-like'. The relationship between an antonym and the negation of its opposite is an entailment-like relationship (strictly speaking, the relationship is between propositions composed of primitives, rather than between the primitives themselves, but we can let that pass). This kind of relationship also applies between I and YOU, on the one hand, and SOMEONE, on the other. If I or YOU appears as the argument of any predicate, it follows that the same predicate may be applied to SOMEONE. Despite this, no polysemy is attested between I, YOU and SOMEONE, presumably because (as noted above) their syntactic properties are too similar.

A similar entailment-like relationship obtains between PEOPLE and SOMEONE, but there is at least one significant syntactic difference, namely, that PEOPLE cannot occur with the quantifier ONE.

There is also another relationship between PEOPLE and SOMEONE, of the kind which may be termed 'implicational'. It is that most of the 'someones' that one deals with in everyday life are, as a matter of fact, human beings (PEOPLE). To appreciate this, it has to be remembered that an important attribute of the primitive SOMEONE is that it is not

restricted in its range of application to human beings: probably most (perhaps all) cultures entertain the notion that non-human 'beings' (someones) exist. Cultures differ in their views as to how often the average person would have direct contact with such beings, but the average person in all cultures no doubt lives out life for the most part among human beings.

Together, these two facts account for the frequent overlaps between the exponents of PEOPLE and of SOMEONE (cf. Wierzbicka 1994a: 452-3). Universally, it seems, languages provide for a unique exponent of SOMEONE in interrogative contexts and very often in certain 'indefinite' contexts as well (as in 'I don't know who', 'I'm thinking of someone'). But quite often, there is overlap between the forms for PEOPLE and for SOMEONE in other contexts. For instance, in Samoan *tangata* when unmarked is PEOPLE, but the singular expression *le tangata* means 'a human being'. In Yankunytjatjara the word *anangu* without a modifier is usually interpreted to mean PEOPLE, but in combination with quantifiers it expresses the meaning 'human being'; for instance, *anangu kutju* 'one human being'. In Mangap-Mbula the expression *zin tomtom* means PEOPLE (generic, human) and *ta tomtom* means SOMEONE (not necessarily human). *Zin* is elsewhere a pluraliser and *ta* the numeral ONE.

An implicational relationship also explains how it is that the form for OTHER often expounds SOMEONE, as we saw above is the case in Yankunytjatjara. Although I and YOU are both 'someones', in face-to-face communication one would only rarely refer to oneself or one's interlocutor in the third person, as SOMEONE. On the other hand, it is very natural to refer to someone outside the speech situation not just as SOMEONE, but as 'someone else' (ANOTHER SOMEONE). Given the syntactic differences between (adnominal) OTHER and (substantive) SOMEONE, it is natural that the two meanings could be expressed by the same form (Wierzbicka 1994a: 451). Overlaps between SOMEONE and OTHER are also found in Arrernte and Samoan.

DO/HAPPEN polysemy has its rationale in an entailment-like relationship between DO TO and HAPPEN. Whenever someone does something to someone or something, it is possible to look at it from the point of view of the 'patient' as an instance of HAPPEN TO (that is, X DID SOMETHING TO Y implies SOMETHING HAPPENED TO Y). The fact that DO (TO) and HAPPEN TO both require time adjuncts is one reflection of this relationship. However, there are important syntactic differences between DO and HAPPEN – DO can take two referential arguments but

HAPPEN takes only one. As detailed above, complete HAPPEN/DO polysemy is found in Kalam; it also occurs partially in Ewe.

Polysemy between HAPPEN and FEEL, also found in Kalam, probably has its roots in an implicational relationship, namely, the fact that a person is likely to FEEL something when something HAPPENS TO them. The grammar of FEEL is different enough from that of HAPPEN to make polysemy viable.

Implicational relationships also provide the rationale for SAY/DO polysemy and SAY/WANT polysemy (Kalam, and partially in Mangap-Mbula). Sometimes saying something to someone affects that person. In this respect SAYING TO can be 'like' DOING TO. As well as this, the typical manner in which people express themselves (say things) is by speaking, which is a physical activity, i.e. a kind of doing.

As for the SAY/WANT overlap, it may reflect the fact that it is hard for people to know what someone else WANTS unless the person SAYS something about it. It is known that in some languages one cannot talk about the wants of other people without using locutions like 'He says: I want' or 'He says: I will'. Thus, it is not difficult to see how polysemies involving SAY and WANT could develop (Wierzbicka 1994a: 467-8). Another possible source for SAY/WANT polysemy is the tendency for the verb SAY to come to function as a complementiser; cf. the discussion of LIKE THIS below.

The partial MUCH/BIG polysemy found in Thai and Samoan is also understandable in terms of an implicational relationship: if MUCH of something is in one place, it tends to imply that something BIG is in that place. Similarly, IF/CAN/MAYBE polysemy makes sense in terms of the implication that 'if something can happen, maybe it will happen'. BECAUSE/AFTER polysemy, attested partially in many languages, makes sense in terms of the implication that 'if event-B happened after event A, maybe it happened because of event-A'.

There are non-compositional relationships of both types which have not yet been found to correlate with polysemy. For instance, if something is PART OF X, it follows that THERE IS ONE OTHER PART OF X; similarly, if something is ONE KIND OF X, it follows that THERE IS ONE OTHER KIND OF Y. It would be premature to assume that these entailment-like links cannot yield polysemy. Perhaps we may soon come across a language in which the exponent of PART OF or of KIND OF is identical in form to ONE OTHER. (In Arrernte there is overlap between OTHER and a portmanteau involving KIND OF, namely, 'different kinds of', which has

the form *arrpenhe-ante-arrpenhe* 'other-only-other'; in Yankunytjatjara a similar portmanteau has the form *kutjupa-kutjupa* 'other-other'.)

Implicational relationships not yet known to be associated with polysemy include those between GOOD and WANT (if X thinks this is good, then X wants this), between SEE (and HEAR) and KNOW (if X sees/hears something, then X comes to know something), and between MAYBE and the combination NOT KNOW IF (if X thinks: maybe it will happen, X doesn't know if it will happen).

One non-compositional polysemy which came to light in the SLU volume appears to have a different origin than any discussed so far. In Longgu, the primitive THINK is realised by means of a word which can also mean LIKE THIS. The link between THINK and LIKE THIS, and also between SAY and LIKE THIS, is clearly that both predicates may take quotational or quasi-quotational complements. Thus, when people want to 'hedge' a little in giving an opinion, they may combine the meanings THINK and LIKE THIS ('I think about it like this: ---'). Similarly, the semantic phrases SAY and LIKE THIS will often occur together ('she said something like this: ---'). In both cases, this semantic frame enables the speaker to add vividness by letting them 'act out' the content of the reported dictum or opinion as if it were his or her own.

'Obviously', as Wierzbicka (1994a: 457) says, the link 'is not compositional; the two meanings ... are likely to co-occur, but they have no common part and cannot be reduced to one another in any way'. But the fact that this link surfaces in THINK/LIKE THIS polysemy cannot be put down solely to collocation, because there are many collocations which do not give rise to polysemy. The crucial factor is that LIKE THIS appears at the boundary of a clause which is within the semantic scope of THINK OF SAY. The LIKE THIS element can thus readily become grammaticalised as a complementiser, after which it is easy to imagine that the verb of saying or of thinking could come to be omitted, giving rise in due course to polysemy. Romaine and Lange (1991) report that English *like* is being grammaticalised in the speech of American teenagers as a marker of reported dicta or opinions (as in *And I'm like 'Nothing much'* or *And I saw him coming, and I'm like 'Noooo.'*)

What applies to LIKE THIS, and to LIKE, applies also to THIS itself, which is well-known to be a diachronic source of complementisers in many languages. According to Frajzyngier (1991: 225) it is a typological generalisation that 'if a language has a complementizer derived from a demonstrative, such a complementizer will mark the complements of

the verbs of saying and thinking'. This suggests that we may one day come across SAY/THIS or THINK/THIS polysemy.

### 5. *The syntax of semantic metalanguage*

The syntax of the universal semantic metalanguage is understood in purely combinatorial terms, the hypothesis being that there is a set of universal patterns according to which the lexical primitives can combine. Lack of explicitness on this score has been the cause of justifiable criticism of the NSM approach. McCawley (1983), for instance, called on Wierzbicka to supplement her list of primitives with a list of their universal patterns of combination (cf. also Goddard 1989). Over the past five or so years Wierzbicka has concentrated an intense research effort into formulating specific hypotheses about such patterns (cf. Wierzbicka 1991b; 1996), which are currently being subjected to cross-linguistic testing.

An important research tool in the current search to discover a universal semantic syntax is the notion of the 'canonical context'. By this is meant a sentence or sentence fragment composed predominantly (or exclusively) of semantic primitives which is hypothesised to be expressible in any language. For instance, given the primitives SOMEONE, SOMETHING, SAY, BAD and YOU, it is hypothesised that one could put them together to say, in any language, the equivalent of SOMEONE SAID SOMETHING BAD ABOUT YOU.

Some examples of sentences embodying hypotheses about NSM syntax are given below. The 'Universal Syntax of Meaning' workshop at the Australian National University (July 1994) used a list of about 200 such sentences as a focus for discussion in a range of languages including Japanese, Mangan-Mbula, Thai, Malay, Arrernte, Yankunytjatjara, Marrithiyel, French, Russian, Mandarin Chinese and Hawaiian Creole English. One volume containing studies of this kind has appeared (Goddard Ed. 1997) and another is underway (Goddard and Wierzbicka Eds, to appear).

13. I don't know where he is now.  
I think maybe something bad happened.  
I want you to do it.  
She said something to me, but I didn't hear it.  
Something happened in this place.

There are many kinds of nut (yam, bat).  
These people lived for a long time.  
He/she did it like this.  
This happened before/after you were born.  
It is far from here to that place.  
This thing has two parts.  
If I do this, maybe people will think something bad about me.

It can be helpful to think of the emerging picture of NSM syntax in terms of 'kernel sentences' of early generative grammar. What are the basic kernel sentences of the semantic metalanguage? How can these basic sentences be elaborated and combined? The following 'sketch grammar' of the natural semantic metalanguage is based on Wierzbicka (1996), who stresses that her proposals are highly tentative.

The basic unit of NSM syntax is analogous to the clause, namely, a combination of a 'substantive' (I, YOU, SOMEONE, PEOPLE, SOMETHING) with any one of a range of 'predicates' and some additional elements determined by the nature of the predicate. The substantives can combine with THIS, and with the other determiners (THE SAME, OTHER), quantifiers (ONE, TWO, ALL, MANY/MUCH, SOME) and 'attributes' (GOOD, BAD, BIG, SMALL) to form units analogous to noun-phrases.

The elements which may function as predicates are as follows:

'Personal' predicates:	THINK, KNOW, WANT, FEEL, SEE, HEAR, SAY
Actions, events, movement:	DO, HAPPEN, MOVE
Life:	LIVE, DIE
Location:	(BE) SOMEWHERE
Spatial relators:	(BE) ABOVE, BELOW, NEAR, FAR, INSIDE, ON (ONE) SIDE
Other relators:	(BE) LIKE, PART OF
Evaluators & descriptors:	(BE) GOOD, BAD, BIG, SMALL
Existence:	THERE IS

These predicative elements can be classified into groups in various ways according to their properties. For instance, although all of them can combine with a 'time adjunct' such as AT THIS TIME OR SOMETIME BEFORE NOW (see below), only some require such an adjunct, be it explicit or understood. These (DO, HAPPEN, MOVE, LIVE, DIE, SEE, HEAR, SAY) may be regarded as analogues of prototypical verbs. A larger group, including

all those just listed as well as KNOW and THINK, and perhaps WANT, can combine with the 'meta-predicate' CAN.

All the 'personal predicates' can take a complement (THIS or SOMETHING), and in the case of KNOW and THINK the complement can be sentence-like, e.g. 'I know that...' , 'I think that ...'. The predicates designated 'relators' are a mixed bunch, consisting of some spatial elements (ABOVE, BELOW, NEAR, FAR, INSIDE), similarity (LIKE), and 'part-hood' (PART OF). What they have in common is that they necessarily involve two referring substantive expressions.

It is notable that the evaluators and descriptors (GOOD, BAD, BIG and SMALL) can function both as predicates and as attributes (i.e. combining directly with substantives, e.g. SOMEONE GOOD/BAD, SOMETHING BIG/SMALL). They can be regarded as analogues of prototypical adjectives.

The predicate THERE IS should also be singled out, firstly, because as long noted by linguists and philosophers, 'existence' is a predicate unlike any other, and secondly, because it can form a proposition without a substantive as 'subject', e.g. 'there are many kinds of bird'.

As mentioned, all predicates can occur with 'time adjuncts', that is, with expressions such as AT THIS TIME, AT THE SAME TIME, BEFORE NOW, AFTER NOW. Most predicates can also combine with the 'time-period' elements A SHORT TIME and A LONG TIME. The syntactic bond in this case is closer than that of temporal adjuncts, but this has yet to be investigated closely. Locational adjuncts also occur, but with a narrower range of predicates than may take time adjuncts. The predicates which take locational adjuncts are DO, HAPPEN, MOVE and THERE IS. Locational adjuncts may also occur in the complement of SEE (as in 'I see something there'). Locational phrases are similar to temporal phrases in composition, e.g. IN THIS PLACE, IN THE SAME PLACE, ABOVE HERE, BELOW HERE.

All types of clause can combine with the two 'operators': negation (NOT) and possibility (MAYBE). When these co-occur, MAYBE has the wider scope.

Clauses can be combined in NSM syntax in various ways. We have already noted that some predicates, e.g. THINK, SAY, KNOW, can take propositional complements. As well, a very powerful clause linking strategy is made possible by the capacity of the element THIS to refer back to the content of a preceding clause. This enables entire clauses to function as temporal adjuncts (AT THE SAME TIME AS THIS, AFTER THIS, BEFORE THIS), and also to participate in another kind of adjunct structure, not mentioned so far, the causal adjunct (BECAUSE OF THIS). Finally, there are two purely interclausal constructions in the current NSM meta-

language, the conditional (introduced by IF) and the counterfactual (IF ... WOULD).

NSM thinking on clause-level grammar as outlined above has many affinities with conceptions of the layered structure of the clause developed in Role and Reference Grammar (RRG); cf. Foley and Van Valin (1984), Van Valin (1990). Over-simplifying, one could say there is agreement in seeing the predicate as the innermost layer of the clause and subject to modification by aspect, which in NSM terms is presumably constructed from duration (FOR SOME TIME) and quantifier elements. The predicate along with its core arguments (valency options) constitutes a larger unit (known in RRG as the core) which is the realm over which internal negation (NO) and root modality (CAN) operate.

Attached to the core is a 'periphery' which may contain temporal and spatial adjuncts. The resulting 'clause' is the domain over which such operators as 'status', 'evidentials' and 'illocutionary force' have their effect. In NSM terms MAYBE is the only semantically simple operator at this level. The other categories are complex, being decomposable in terms of elements such as (for evidentials, cf. Wierzbicka 1994b) I, KNOW, SEE, HEAR, and SAY, and (for illocutionary force, cf. Wierzbicka 1991a) I, YOU, WANT, BECAUSE, and (pre-eminently) SAY. The fact that such complex operators have their scope over an entire clause will follow from the fact that the basic predicates involved (SAY, KNOW) are those which take clausal complements.

In her latest work on NSM syntax Wierzbicka has embraced one new theoretical concept which is a radical departure from her earlier work. This is the notion that semantic primitives may have optional or alternative 'valency options'. For example, it is now assumed that the primitive DO, in addition to its obligatory subject and complement (as in 'someone did something'), may also take a 'patient' (as in 'someone did something TO someone'). In other words, from the point of view of their realisation in English, it is now proposed that DO and DO TO are manifestations of a single primitive. This interpretation supersedes long-standing attempts to explicate the 'patient' notion in lexical (compositional) terms, as proposed in Wierzbicka (1980); cf. Bogusławski (1991). It is furthermore assumed that DO may manifest an addition 'instrumental' option, or a 'comitative' option.

14a. X did something

14b. X did something to someone [i.e. 'patient' option]

- 14c. X did something to someone with something [i.e. 'instrument' option]  
 14d. X did something with someone [i.e. 'comitative' option]

The primitive SAY also has an array of valency options. As well as an obligatory subject and complement, it may also take an optional 'addressee' and/or an optional 'topic'. That is, in an English-based NSM, SAY has extra valency options SAY TO and SAY ABOUT, which may occur separately or both at once, as in (15).

- 15a. someone said something  
 15b. someone said something to someone [optional "addressee" slot]  
 15c. someone said something about something [optional "topic" slot]  
 15d. someone said something to someone about someone [both optional slots]

DO and SAY are far from the only primitives now believed to have alternative or optional valency options. For instance, it is now thought that GOOD has a 'beneficiary' option (GOOD FOR); that HAPPEN has an optional 'patient' slot (HAPPEN TO) and possibly an alternative optional 'locus' slot (HAPPEN IN THIS PLACE); and that, like SAY, THINK has an optional 'topic' slot (SAY ABOUT, THINK ABOUT).

Two other kinds of valency option are contemplated in current work. 'Reference point' options would apply to the primitives THE SAME, OTHER and MORE, introducing a second argument giving the point of comparison, as in (16). 'Subset' options would apply to quantifiers, as in (17).

- 16a. I did the same as you  
 16b. someone other than me; in another place than here; a kind other than this  
 16c. more than this; more than two things of the same kind; more than before  
 17. one/two/many of these people

It should perhaps be emphasised that when Wierzbicka claims universality for the various valency options and complementation possibilities of semantic primes adumbrated above, she is not claiming that the formal realisation of these structures in different languages will

be identical. But formal differences do not necessarily compromise semantic equivalence. To make this point more concrete, consider the formal marking of argument structure. In English, the 'topic' arguments of SAY, THINK, and KNOW happen to be marked by the same formal means, namely, the preposition *about*. But it is not expected that the same marking will be used in all languages, or even that an adposition will necessarily be involved. For instance, in Mbula there is only a single general oblique preposition *pa*, which, as shown in (18), marks both 'locutionary topic' and 'addressee' (which are distinguished by order). In Yankunytjatjara, the 'topic' of SAY is marked by a suffix *-tjara* (which in other constructions can express the meaning 'have'), while the 'topic' of THINK appears as a direct object; see (19).

## 18. Mbula:

<i>Ni</i>	<i>i-so</i>	<i>piam</i>	<i>pa</i>	<i>mbulu</i>	<i>tiam</i>	<i>ta</i>	<i>na.</i>
he	3SG-say	REF.US	REF	behavior	our	SPEC	GIVEN

'He spoke to us about our behaviour.'

- 19a. *Ngayulu ara kutjupa-tjara wangka-nytji-tkitja mukuri-nganyi.*  
 I matter other-having SAY-noml-COMPL WANT-PRES  
 'I want to talk about another matter'.

- 19b. *Ngayulu mani kuli-ni.*  
 I money think-pres  
 'I'm thinking about money.'

Similarly, languages differ in formal means used to constitute the determiner and quantifier constructions. In English, a relatum of the prime THE SAME is marked by the preposition *as*, while a relatum of OTHER is marked by *than*. In Malay, a relatum of *sama* (THE SAME) is marked by the preposition *dengan* which in other constructions can mean, roughly, 'with'; e.g. *sama dengan saya* 'the same as me'. A relatum of *lain* (OTHER) is marked by *dari/daripada*, which is elsewhere 'from'; e.g. *lain daripada saya* 'other than me'. Examples (20) and (21) show two different strategies by which the 'subset' valency options of the quantifiers can be realised. In Lao (Enfield Forthcoming), to express the same meaning as the English phrase *two of these people*, one uses a construction which has the form 'these people, two people'. In Mbula, the corresponding construction has the form 'these people, two of

them'. But none of these differences in formal realisation interferes with the semantic equivalence between the constructions.

20. Lao:  
*Khon nii sòdng khoon jaak vaw kap too.*  
 people this two people want say with you  
 '(Of) these people, two people want to speak to you.'

21. Mbula:  
*Wal tana kizin ru ...*  
 people that LOC.3pl two ....  
 'These people, two of them ....'

Complement structures also differ from language to language in their formal realisation. For example, depending on whether the subject of the complement clause is the same as or different to that of the main verb *mukuringanyi* (WANT), Yankunytjajara complements will receive different complementisers (*-kitja* or *-ku*, respectively). In many languages, complements of WANT assume the same form as complements of THINK and KNOW – unlike English which utilises a different form (with *to*) as oppose to THINK and WANT (with *that*). In some languages, complement clauses may not be formally marked as such, that is, they may be identical in form with independent clauses. But, again, none of these variations means that one cannot express, in all languages, sentences equivalent in meaning to, for example, I WANT TO DO SOMETHING, I DON'T WANT THIS TO HAPPEN, I WANT YOU TO DO THIS, and so on. For a more comprehensive discussion of the range of formal realisations of 'metalanguage syntax', see Goddard and Wierzbicka (Eds, Forthcoming).

The new NSM work on valency options of semantic primitives has a certain affinity with conceptions of 'semantic roles' which have descended from the work of Gruber (1970) and Fillmore (1968, 1977). Many contemporary approaches to grammar incorporate an inventory of semantic (or 'thematic') roles, such as agent, patient, theme, experiencer, beneficiary, and so on. As Blake (1994: 68) remarks, however: 'there are no agreed criteria and there is certainly no consensus on a universal inventory'. Current NSM work offers the prospect of characterising semantic roles precisely and universally, by tying them to argument roles of semantically primitive lexical universals. Roughly, 'agent' is the obligatory argument of DO and

'patient' is its optional argument (i.e. the second argument of DO TO), 'theme' is the argument of MOVE, 'beneficiary' is the argument of GOOD FOR. Interestingly, Jackendoff (1990: 46-50) has also proposed that semantic (thematic) roles should be re-conceptualised as argument roles of conceptual primitives.

Of course, in the literature the terminology of semantic roles is often used inconsistently. For example, 'experiencer' sometimes means someone who FEELS and at other times someone who SEES or HEARS ('perceiver'), or even KNOWS ('cogniser'). The use of lexical primitives would facilitate standardisation in this respect (cf. Wierzbicka 1995). As another example, the term 'agent' is sometimes used assuming 'volition' or 'control' and sometimes not. Using semantic primitives (especially WANT and THINK) one could define these complex notions and ensure they are used in a precise and standardised fashion.

It must be conceded that the summary undertaken in this section has been sketchy not only on account of limitations of space, but also because of our present limited state of knowledge. There has been over 30 years of research into lexical primitives, but only a few into their syntax. Almost every area of the NSM syntax (substantives, mental predicates, space, time, quantification, conditionals, etc.) could sustain a doctoral study if approached thoroughly from a cross-linguistic point of view. Only a single such study has yet been undertaken, Jean Harkins' (1995) inquiry into the semantic syntax of WANT. On the other hand, it should be clear that considerable progress is being made toward the goal of constructing a universal metalanguage for semantic representation based on the shared grammatical and lexical core of all languages.

### 7. Concluding remark

Looking back over the past twenty years, one can see a gradual convergence of interests between Wierzbicka and her associates, and what one might call the mainstream of modern linguistics. Wierzbicka's project has moved from a preoccupation with lexical decomposition to a new focus on the deep, universal syntax of meaning. The mainstream (or at least, important currents within it) has moved from a preoccupation with non-semantic – or even anti-semantic – accounts of syntax to a new interest in the grammatical aspects of lexical semantics.

Though much remains to be done, the NSM metalanguage is the most comprehensive and empirically-based system of semantic analysis in modern linguistics. It will be surprising if the next decade does not see it attract increasing attention from mainstream linguistics.

*Department of Linguistics  
University of New England  
Armidale, NSW 2351 Australia*

## References

- Ameka, Felix. 1994. Ewe. In Goddard and Wierzbicka (eds), 57-86.
- Andrews, Avery. 1985. The major functions of the noun phrase. In T. Shopen (ed.) *Language Typology and Linguistic Description*. Cambridge: Cambridge University Press. 62-154.
- Benveniste, Emile. 1971[1966]. *Problems in General Linguistics*. Translated by Mary Elizabeth Meek. Coral Gables: University of Miami Press.
- Blake, Barry J. 1994. *Case*. Cambridge: Cambridge University Press.
- Bogusławski, Andrzej. 1991. Semantic primes for agentive relations. *Lingua Posnaniensis* 32/33. 339-64.
- Bolinger, Dwight. 1989. Extrinsic possibility and intrinsic potentiality: 7 on May and Can +1. *Journal of Pragmatics* 13.1-23.
- Bugenhagen, Robert D. 1994. The exponents of semantic primitives in Mangap-Mbula. In Goddard and Wierzbicka (eds), 87-108.
- Diller, Anthony. 1994. Thai. In Goddard and Wierzbicka (eds), 149-170.
- Dixon, R.M.W. 1979. Ergativity. *Language* 55. 59-138.
- Enfield, Nick. Forthcoming. The syntax of natural semantic metalanguage expressions in Lao. In Goddard and Wierzbicka (eds), Forthcoming.
- Evans, Nicholas. 1994. Kayardild. In Goddard and Wierzbicka (eds), 203-228.
- Fillmore, Charles J. 1968. The case for case. In E. Bach and R. T. Harms (eds), *Universals in Linguistic Theory*. Holt Rinehart & Winston. 1-88.
- Fillmore, Charles J. 1977. The case for case reopened. In P. Cole and J. M. Sadock (eds), *Grammatical Relations*. New York: Academic Press. 59-83.
- Foley, William A. 1986. *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.
- Foley, William A., and Robert D. Van Valin Jr. 1984. *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.
- Frajzyngier, Zygmunt. 1991. The *De Dicto* Domain in Language. In Traugott, E. and Heine, B. (eds) *Approaches to Grammaticalization* (Vol 1). Amsterdam: John Benjamins. 219-251.
- Goddard, Cliff. 1989. *Issues in Natural Semantic Metalanguage*. *Quaderni di Semantica*, 10 (1). 51-64.
- Goddard, Cliff. 1991. Testing the Translatability of Semantic Primitives into an Australian Aboriginal language. *Anthropological Linguistics* 33(1). 31-56.
- Goddard, Cliff. 1994. Lexical primitives in Yankunytjatjara. In Goddard and Wierzbicka (eds), 229-262.
- Goddard, Cliff. 1996. Cross-linguistic research on metaphor. *Language and Communication* 16(2). 145-151.
- Goddard, Cliff. 1997. (Ed.) *Studies in the Syntax of Universal Semantic Primitives*. Special Issue of *Language Studies* 19(3).
- Goddard, Cliff. In press a, 1997. Grammatical categories and semantic primes. *Australian Journal of Linguistics* 17(1).
- Goddard, Cliff. In press b, 1998. *Semantic Analysis: A Practical Introduction*. Oxford: Oxford University Press.
- Goddard, Cliff. Forthcoming. The syntax of semantic primitives in Malay (Bahasa Melayu). In Goddard and Wierzbicka (eds.), forthcoming.
- Goddard, Cliff and Anna Wierzbicka (eds). 1994. *Semantic and Lexical Universals - Theory and Empirical Findings*. Amsterdam: John Benjamins.
- Goddard, Cliff and Anna Wierzbicka. Forthcoming. *Meaning and Universal Grammar*. New York: Oxford University Press.
- Gruber, Jeffrey. 1970. *Studies in Lexical Relations*. Bloomington, Indiana: Indiana University Linguistics Club.
- Harkins, Jean. 1995. *Desire in language and thought: A cross-linguistic and cross-cultural study*. PhD thesis. Australian National University.
- Harkins, Jean and David P. Wilkins. 1994. Mparntwe Arrernte and the Search for Lexical Universals. In Goddard and Wierzbicka (eds), 285-310.
- Hill, Deborah. 1994. Longgu. In Goddard and Wierzbicka (eds), 311-330.
- Ishiguro, Hide. 1972. *Leibniz's Philosophy of Logic and Language*. London: Duckworth.
- Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, MA: MIT Press.
- Mosel, Ulrike. 1987. Subject in Samoan. In Laycock, Donald C. and Winter, Werner (eds) *A world of language*. (Pacific Linguistics, C-100). Canberra: Pacific Linguistics. 455-479.
- Mosel, Ulrike. 1994. Samoan. In Goddard and Wierzbicka (eds), 331-360.
- McCawley, James D. 1983. Review of Anna Wierzbicka's *Lingua mentalis: The semantics of natural language*. *Language*, 59(3). 654-659.



- Pawley, Andrew. 1994. Kalam Exponents of Lexical and Semantic Primitives. In Goddard and Wierzbicka (eds), 387-422.
- Romaine, Suzanne and Deborah Lange. 1991. The use of *like* as a marker of reported speech and thought: a case of grammaticalisation in progress. *American Speech* 66(3).227-279.
- Sørensen, Holger Steen. 1963. The meaning of proper names. Copenhagen: Gad.
- Van Valin Jr., R. D. 1990. Layered syntax in role and reference grammar. In J. Nuyts, A. M. Bolkestein and C. Vet (eds), *Layers and levels of representation in language theory*. Amsterdam: John Benjamins.193-232.
- Wierzbicka, Anna. 1972. *Semantic Primitives*. Translated by Anna Wierzbicka and John Besemeres. Frankfurt: Athenäum Verlag.
- Wierzbicka, Anna. 1976. In defense of YOU and ME. In W. Girke and H. Jachnow (eds), *Theoretische Linguistik in Osteuropa*. Tübingen: Max Niemeyer Verlag.1-21.
- Wierzbicka, Anna. 1980. *Lingua Mentalis: The Semantics of Natural Language*. Sydney: Academic Press.
- Wierzbicka, Anna. 1985. *Lexicography and Conceptual Analysis*. Ann Arbor: Karoma.
- Wierzbicka, Anna. 1987. *English Speech Act Verbs: A semantic dictionary*. Sydney: Academic Press.
- Wierzbicka, Anna. 1988. *The Semantics of Grammar*. John Benjamins.
- Wierzbicka, Anna. 1989. Semantic primitives: The expanding set. *Quaderni di Semantica*, 10(2).309-332.
- Wierzbicka, Anna. 1991a. *Cross-cultural Pragmatics: The semantics of social interaction*. Berlin: Mouton de Gruyter.
- Wierzbicka, Anna. 1991b. Lexical universals and universals of grammar. In M. Kefer and J. v. d. Auwera (eds), *Meaning and Grammar*. Berlin: Mouton de Gruyter.
- Wierzbicka, Anna. 1992a. *Semantics, Culture, and Cognition*. Oxford: Oxford University Press.
- Wierzbicka, Anna. 1992b. The search for universal semantic primitives. In M. Pütz (ed), *Thirty years of linguistic evolution*. Amsterdam: John Benjamins.215-242.
- Wierzbicka, Anna. 1994a. Semantic primitives across languages: a critical review. In Goddard and Wierzbicka (eds), 445-500.
- Wierzbicka, Anna. 1994b. Semantics and epistemology: The meaning of 'evidentials' in a cross-linguistic perspective. *Language Sciences* 16(1).81-137.
- Wierzbicka, Anna. 1995. A semantic basis for grammatical typology. In W. Abraham, T. Givón and S. Thompson (eds), *Discourse, Grammar, and Typology*. Amsterdam/Philadelphia: John Benjamins.179-209.

- Wierzbicka, Anna. 1996. *Semantics, Primes and Universals*. Oxford: Oxford University Press.