# 1. Natural Semantic Metalanguage: The state of the art

**Chapter** · January 2008

DOI: 10.1075/slcs.102.05god

CITATIONS
46

READS
8,099

**1 author:**

Cliff Goddard
Griffith University
**139** PUBLICATIONS **3,180** CITATIONS

SEE PROFILE

CHAPTER 1

# Natural Semantic Metalanguage: The state of the art

Cliff Goddard

This chapter gives an overview of the Natural Semantic Metalanguage approach to semantic analysis, originated by Anna Wierzbicka. It enumerates the main theoretical and methodological arguments underpinning the approach, and reviews the 35 year program of conceptual analysis and cross-linguistic testing which has led to the current model of 63 semantic primes and their associated syntax. Identification issues concerning polysemy, allolexy, portmanteau exponents and apparent lexical gaps are discussed with reference to specific examples from a variety of languages. The chapter then turns to the methodology and practice of semantic description, discussing and illustrating how NSM techniques can be applied to different areas of the lexicon. The relatively new concepts of semantic molecules and semantic templates are discussed in some detail. The chapter concludes with a brief review of the contents of the volume.

## 1. The NSM system

The Natural Semantic Metalanguage (NSM) is a decompositional system of meaning representation based on empirically established universal semantic primes, i.e., simple indefinable meanings which appear to be present as identifiable word-meanings in all languages (Wierzbicka 1996; Goddard and Wierzbicka eds 2002). The words and grammar of the metalanguage can be thought of as a highly disciplined and standardised subset of natural language: a small subset of word-meanings (63 in number, see Appendix 1), together with a subset of their associated grammatical properties. Originating with Wierzbicka (1972), the system has been developed and refined over some 35 years.

The NSM approach remains controversial: many linguists have either a "love it" or "hate it" attitude towards it. Nonetheless, on objective criteria (longevity, range, publication outputs) it is indisputable that NSM is one of the most well developed, productive, and comprehensive systems of semantic analysis in contemporary linguistics. The approach has been experiencing a notable upswing over the past decade or so, beginning with the publication of the

collective volume *Semantic and Lexical Universals* (Goddard and Wierzbicka eds 1994). Since then there have been 12 books, including a textbook (Goddard 1998) and several collective volumes or special issues (Goddard ed. 1997, 2006; Harkins and Wierzbicka eds 2001; Enfield and Wierzbicka eds 2002; Durst ed. 2003; Peeters ed. 2006), and over 200 refereed journal publications. In addition to Anna Wierzbicka, Cliff Goddard, Jean Harkins, Bert Peeters, Felix Ameka and other "old hands", there is a raft of new generation NSM researchers, such as (among others): Catherine Travis, Rie Hasada, Marie-Odile Junker, Uwe Durst, Kyung-Joo Yoon, Zhengdao Ye, Ian Langford, Jock Wong, Anna Gladkova, Yuko Asano, Adrian Tien, Helen Bromhead, Anna Brotherson and Sophie Nicholls. Some of these are represented in the present volume.

Within linguistics, increasing interest in and acceptance of the NSM approach has been aided by the tide slowly turning against the narrow English-based "syntacticocentrism" (cf. Jackendoff 2001) which predominated for most of last century, in favour of semantic and typological studies.[1] At the same time, the NSM approach has increasingly attracted favourable attention in adjacent disciplines such as anthropology (D'Andrade 2001), cultural psychology (Shweder 2004), evolutionary psychology (Jones 1999), and semiotics (Eco 1999: 150–153).

A strong point of NSM work over three decades has been the consistency or stability of its notation, in particular, the use of reductive paraphrases. This stands in sharp contrast to the situation in linguistics at large (especially in syntax), where, as Croft (2001: i) has observed "a continuing kaleidoscope of notations [has] … made even five-year old journal articles—and many reference grammars—difficult to decipher". This is not to say, of course, that the program has been static over the years. On the contrary, it is possible to discern several periods in which research activity was focused on different goals. Briefly, the 1970s to mid-1980s can be termed the "early development" period of the semantic primitives approach. The key publications (Wierzbicka 1972, 1980a, 1980b) worked with a very austere set of 13 or 14 semantic primes. Even so, valuable analytical work was done on emotion terms, body parts, stage of life terms, speech acts, cases, and a number of other areas. Wierzbicka (1985) was an unprecedented investigation into the semantics of "concrete" vocabulary. The mid-1980 to late 1990s has been dubbed the "expanding set" phase. Over this period the inventory of semantic primes was expanded almost five-fold, and a new research emphasis on cross-translatability and lexical universals entered the program. From the mid-1990s onwards, the emphasis shifted to the grammar of the metalanguage and to "whole metalanguage" studies, embracing both lexicon and grammar. The publication of the two volume *Meaning and Universal Grammar* (Goddard and Wierzbicka eds 2002) marked a new highpoint of empirical detail and theoretical explicitness. More recently one can discern a trend towards increased systematisation and formalisation. Nonetheless, over this whole span of three and half decades work, the fundamental conviction behind the NSM program has remained the same.

This fundamental conviction is the conviction that ordinary natural languages are adequate to represent their own semantics via language-internal paraphrase; that is, belief in the "meta-semantic adequacy" of natural languages. This entails the view that every language has an irreducible semantic core with a language-like structure, with a mini-lexicon of indefinable expressions (semantic primes) and associated syntax.

Logicians and formal semanticists usually adopt the opposite assumption, essentially because they regard natural language as too unsystematic, vague and grammatically idiosyncratic to be relied upon as a precise and unambiguous mode of expression. In a classic essay, Tarski (1935) expressed the view that to pursue the semantics of colloquial language using "exact methods" would necessitate the "thankless task of a reform of this language … to define its structure, to overcome the ambiguity of the terms which occur in it"; and in the end it may be doubted "whether the language of everyday life, after being 'rationalised' in this way, would still preserve its naturalness and whether it would not rather take on the characteristic features of the formalised languages" (p. 267). Now that the NSM metalanguage exists as a practical reality, we are in a position to reassess Tarski's view. We can now see that the tiny vocabulary and narrow range of grammatical patterns in the NSM indeed compromise its "naturalness" in certain ways, but—crucially—not to the point that it cannot be understood via ordinary natural language. Equally, we can now see that in this disciplined, constrained and "rationalised" version of natural language, it is possible to formulate testable hypotheses about a multitude of domains in lexical, grammatical and illocutionary semantics.

To step away from natural language as the language of semantic representation, on the other hand, is not only unnecessary, it is ultimately futile. This is because even a technical or artificial semantic language of description (composed of features, logical symbols, and what-not) still has to be interpretable by the users of the representation. Even if these users are trained experts, such as professional semanticists, they necessarily rely on their native natural languages as they are learning to use their "expert system", and they necessarily fall back on their native natural languages when they communicate between themselves about the intended meaning or interpretation of the technical analyses (as when logicians and formal semanticists routinely introduce and discuss the "intuitions" behind their formalisms in ordinary language paraphrases). On this basis Keith Allan (1986: 268) has argued that any artificial semantic formalism is really "a degenerate form of a natural language", because in order to read and interpret an artificial metalanguage what one effectively does is to mentally translate it back into one's native language. A standardised and regularised version of a natural language is preferable to an artificial creation which is in a covertly parasitic relationship with ordinary language.[2]

There are also methodological reasons for preferring a metalanguage of semantic description based as transparently as possible on natural language.

As pointed out by Ruth Kempson (1977), among others, to test and revise hypotheses about meanings depends on our being able to generate predictions about ordinary language use on the basis of the semantic description. The more transparent the relationship between the two, the clearer and more testable the descriptions become. Conversely, as the relationship between the semantic description and the "real" language being described becomes more and more tenuous, the prospect of any real empirical testing evaporates. For example, take the perennial analysis of *x kills y* as [CAUSE *x* (DIE *y*)]. One obvious empirical objection to this analysis is that the range of use of *X caused Y to die* does not match that of *X killed Y*, because the version with *kill* demands a more "direct" connection between the agent's action and the fatal outcome. The proper reaction to such evidence is to revise the analysis, but all too frequently the response has been to claim that the semantic component CAUSE is not identical in meaning to the English verb *cause* (e.g., McCawley 1972). Allowing a disconnect between the language of description and the language being described makes the analysis immune from empirical disconfirmation.

A third reason for rejecting technical terms in the metalanguage relates to the goal of achieving a description which is conceptually realistic, in the sense of capturing the indigenous or "native" conceptualisation. A description framed in indigenous categories can have a *prima facie* case for conceptual reality, because it is framed in terms which are used by speakers themselves on a daily basis, which are recognisable to them, and which are directly accessible to their intuitions. In contrast, a description framed in "exotic" terms, unrecognisable to the people concerned, at the very least requires some kind of special pleading before it can be accepted as modelling the conceptualisation of these people. Yet such special pleading is seldom provided by advocates of exotic, technical analyses. Note that I am not saying that abstract technical analyses are necessarily wrong or that they have no place in semantics—only that they require special justification. Other things being equal, a simpler representation framed in natural language is always preferable to a technical one.

The principle that semantic hypotheses about indigenous ways of thinking ought to be framed in indigenous terms (Wierzbicka's (1992: 331) "principle of indigenisation") disqualifies the analyst from using even ordinary language words if these words have no equivalents in the language being described. For example, it would violate the principle of indigenisation to describe the causatives of Yankunytjatjara using a putative semantic component CAUSE, because, like most languages of the world, Yankunytjatjara has no verb corresponding to *cause* (though it has an equivalent to the semantic prime BECAUSE Goddard (1991)). Violating the principle of indigenisation can legitimately be described as "terminological ethnocentrism", because like ethnocentrism in general it involves imposing alien cultural categories upon people of other languages and cultures, with an inevitably distorting effect.

On the basis of the discussion to this point, we can draw the following conclusion for cross-linguistic semantics: an optimal semantic metalanguage must be based as transparently as possible on ordinary natural languages, and it must consist only of elements whose meanings are present in all natural languages, i.e., of universally lexicalised meanings. In short, it ought to be based on "the intersection of all languages".

### 1.1 Identifying semantic primes within and across languages

As mentioned, semantic primes are, by definition, those meanings in any language which cannot be paraphrased in simpler terms; but the linguistic status of the exponents of semantic primes is sometimes misunderstood. It is critical to recognise that exponents of primes are not lexemes in a given language, but rather "lexical units" (Cruse 1986; Mel'čuk 1989), i.e., pairings of a single specifiable sense with a lexical form. Matching semantic primes across languages means aligning lexical units (across languages) which share a given primitive meaning. When exponents of a given prime have different polysemic extensions (as they frequently do), there is a match-up between lexical units, but not between whole lexemes.

Just as a single lexical form sometimes serves to express two or more meanings (polysemy or homonymy), so a single meaning is sometimes expressed by two or more forms. Applied to semantic primes, this means that we can expect that a single prime will sometimes have two or more allolexes, i.e., alternative lexical realisations.

## 2. Polysemy and allolexy of semantic primes

After some fifteen years of research into how semantic primes manifest themselves across languages, NSM researchers have accumulated a lot of data about common patterns of polysemy. Some widely attested patterns are summarised in Table 1. Language-specific evidence is always required to support claims for semantic primes which depend on a polysemy analysis.

**Table 1.** Some common polysemies involving exponents of semantic primes (data from studies in Goddard and Wierzbicka eds 1994, 2002, and this volume Ch 2)

| | | |
|---|---|---|
| SAY | 'speak', 'make sounds' | Thai, Mandarin, Yankunytjatjara, Kalam |
| | 'do' | Bunuba, Kalam |
| THINK | 'worry', 'long for', 'intend' | Amharic, Mandarin, Swedish |
| WANT | 'like', 'love', 'seek' | Spanish, Ewe, Ulwa |
| HAPPEN | 'arrive', 'appear' | French, Ewe, Mangaaba-Mbula |
| DO | 'make' | Malay, Arrernte, Samoan, Kalam |
| BEFORE | 'first', 'go ahead', 'front' | Lao, Samoan, Kayardild, Ewe |
| FEEL | 'taste', 'smell', 'hold an opinion' | Acehnese, Ewe, French, Mandarin |
| BECAUSE | 'from' | Yankunytjatjara, Arrernte |

The converse of polysemy is allolexy, i.e., a situation in which there are multiple lexical realisations of a single prime. Consider, for example, the alternation between indefinite substantives (*something*, *someone*, *somewhere*, *sometime*), generic (or "light") nouns (*person*, *thing*, *place*, *time*),[3] and interrogative pronouns (*what*, *who*, *where*, *when*) used as embedded complements of KNOW. In normal idiomatic English, indefinites are used when they are not combined with a modifier (except for postnominal *else*: see below); generic nouns are used with modifiers; and the interrogative pronouns are used as embedded complements:

(1) a. something happened.
b. the same thing happened again.
c. I don't know what happened.

(2) a. she said something about him.
b. she said two things about him.
c. I don't know what she said about him.

Intuitively, the meaning expressed by the items *something*, *thing*, and *what* is the same across the examples in (1) and (2) above, and it seems impossible to state any difference between the three sets of expression classes in the form of a substitutable paraphrase. Furthermore, there is a systematic pattern of mutual entailments between the three expression sets. On this basis, we can regard SOMETHING, WHAT, and THING as allolexes. Likewise, semantic prime OTHER has a postnominal allolex ELSE, which can be used as the sole modifier of an indefinite or interrogative/ignorative; for example, *something/what else*, *someone/who else*, *somewhere/where else*, *sometime/when else*.

Like all patterns of allolexic variation, these patterns are language-specific. They are part of the English-specific realisation of the natural semantic metalanguage, but they are not essential or integral to the natural semantic metalanguage as such. It is possible to violate these particular "allolexical rules" of English without losing intelligibility, and with only partial loss of acceptability: expressions such as *this someone*, *the same someone*, *one other someone*, and the like, are quite intelligible and are even attested, albeit rarely, in corpora of English usage. In the interests of consistency and simplicity, there is a case to be made for using them in semantic explications. Other allolexical rules are stricter, in the sense that violating them results in outright ungrammaticality or unintelligibility. For example, the English realisation of the first-person pronoun exhibits "case allolexy"; i.e., there are two allolexical forms *I* and *me*, with *I* appearing preverbally and *me* elsewhere (Goddard and Wierzbicka 2002). Swapping these forms creates a bizarre effect which would be hardly tolerable in explications. In heavily inflectional languages, such as the Slavic

languages, case allolexy is extensive, and usually near obligatory. In isolating languages, on the other hand, it can be completely non-existent.

In some languages, certain combinations of semantic primes are normally expressed via portmanteau forms; for example, Cantonese has a portmanteau *móuh* of NOT and THERE IS, Polish has the portmanteau *dawno (temu)* of A LONG TIME and BEFORE. Occasionally the portmanteau may be identical with the exponent of one of the primes involved (Roberts 2005); for example, in Russian the combination 'two someones' is expressed by the word for TWO, e.g., *ėtix dvoe* [these two] 'these two (someones)'.

Allolexy often seems to serve a "structure-indexing" function. That is to say, the formal patterns of variation show—in an iconic-indexical fashion—something about the combinatorial context in which the element occurs (cf. Goddard 2002). Nevertheless, there is no paraphrasable meaning difference; nor can language-specific allolexical effects be preserved under translation.

The twin and complementary phenomena of polysemy and allolexy have troubled some observers, either because (with Tarski, and Leibniz for that matter) they crave for an ideal "one form, one meaning" metalanguage, or because they are suspicious that polysemy and allolexy allow NSM researchers too much room to wriggle out of difficult situations. Be that as it may, there is no denying the reality of these phenomena.

As far as we know, there is no human language in which exponents of semantic primes are unaffected by polysemy and allolexy, i.e., in no human language is there a transparent one-to-one mapping between universal primitive meanings and surface lexical forms. Rather than regretting or fretting about this, I suggest that we try to think about the situation a little differently, and in particular I would like to suggest an analogy with the human body. The human body is clothed and decorated differently in different cultures and societies around the world. To the naked eye, people's physical forms look different in different parts of the world—so much so that we can often guess where someone comes from by how they look. But beneath the clothing, headgear, jewellery, tattoos, scarification, and other superficial alternations, the essential human body is the same all around the world, and nobody is much troubled by this.

It is interesting and important to study and to understand different forms of clothing and body decoration, which are frequently linked with history and with culture, and the same goes for the study of language-specific patterns of polysemy and allolexy. They constitute an essential part of the "individuality" of each human language, and are worthy of study in their own right. But at the same time, it would be absurd to put aside the study of human physiology and anatomy on account of surface variations, and, indeed, to deny the universality of the human body would be to deny the physiological and anatomical unity of humankind. To deny *a priori* the shared common conceptual core of languages is to deny the "psychic unity" of humankind.

## 2.1 "Missing" exponents

The hypothesis that "every semantically primitive meaning can be expressed through a distinct word, morpheme or fixed phrase in every language" has been termed the Strong Lexicalisation Hypothesis (Goddard 1994: 13). After more than a decade's subsequent work, NSM researchers continue to affirm and defend this hypothesis. Yet how often does one hear NSM theory dismissed with the casual assertion that "there's no word for semantic prime X in such-and-such language"? Such assertions can have quite an impact when they come from apparently authoritative sources, such as fieldworkers, and when they concern languages and cultures from distant places. It is always easier to believe that culturally distant peoples have "fundamentally" different ways of thinking (and the word "fundamentally" always seems to turn up in these contexts) than to believe the same about one's friends and neighbours (cf. Keesing 1994). But when examined closely, these claims have not (so far) stood up. Let us run quickly through some examples.

*No exponent of TIME in Hopi?*
The classic example is Benjamin Lee Whorf's (1956) claim that Hopi is a "timeless language". Yet as shown in Malotki's (1983) monumental study, the Hopi language has an interrogative/indefinite word for WHEN (namely, *hisat*), and expressions such as 'at this time' and 'at the same time' are readily expressible in Hopi via an allolex *-sat/saq* TIME (cf. Goddard 2003b for more details). The following examples are taken from Malotki (1983: 305, 146, 144):

(3) *Pam hisat nima?*
that when go.home
'When did he go home?'

(4) *Taavok yàa-sat=haqam ay nu' tsöng-moki.*
yesterday this-time=APPROX ASSR I hunger-die
'Yesterday at about this time I got really hungry.'

(5) *Pam sú-'inùu-saq nakwsu.*
that the.same-me-time start.out
'He started out at the same time as me.'

*No exponent of FEEL in Nepali?*
In his idiosyncratic "review" of *Meaning and Universal Grammar*, van Driem (2004) preferred not to comment on the detailed studies documenting exponents of semantic primes in Lao, Chinese, Malay, Polish, Mbula and Spanish.

He chose instead to bring forward language data of his own, which readers would presumably find hard to challenge.

> Yet what does NSM make of the ubiquitous inconveniences of life such as the fact that Nepali has no word or expression that means the same thing as English *feel*? … The mental sense of the English-inspired prime FEEL finds functional translational equivalents in Nepali through impersonal constructions with the verb *laagnu* 'impinge upon, begin to, appear to be, make itself felt'. Yet this does not change the fact that there is no Nepali equivalent for English *feel* and no English meaning equivalent to *laagnu*. (van Driem 2004: 163–164)

The final sentence of this quotation makes it clear that van Driem does not acknowledge any difference between lexemes and lexical units. In reality, the Nepali verb he mentions—*laagnu*—is a perfectly good exponent of semantic prime FEEL when it is used in an impersonal construction, i.e., with an accusative subject and without subject verb-agreement, as in the examples below.

(6) *Ma-laai khushi/dukha/bhog laagyo.*
1SG-ACC happiness/sadness/hunger feel.PAST.3SG
'I feel happy/sad/hungry.'

To be sure, *laagnu* is a polysemous verb, and in other grammatical constructions, it can convey different meanings, but the meaning it expresses in the impersonal construction appears to match semantic prime FEEL perfectly.

*No exponent of ALL in Pirahã?*

Everett (2005) insists that Pirahã, spoken in the Brazilian rainforests, lacks any expressions equivalent to quantifier primes such as ALL, ONE, TWO, MUCH/MANY, and SOME. An overriding cultural principle is supposedly responsible: "Pirahã culture avoids talking about knowledge that ranges beyond personal, usually immediate experience, or is transmitted via such experience. ... Abstract entities are not bound by immediate personal experience and therefore Pirahã people do not discuss them" (pp. 622–623).[4] We will only look into the claim that "there is no word for 'all' in Pirahã". The "closest expressions Pirahã can muster", according to Everett, are examples such as the following (his examples (10) and (12)), where the word *'ogi* 'big' (or a nominalised version *'ogiáagaó* 'bigness') appear to convey something similar to ALL.

(7) *Hiaitíihí hi 'ogi-'áaga-ó*
pirahã.people he big-be(permanence)-direction
*pi-ókaobíi.*
water-direction entered
'All the people went to swim/went swimming/are swimming/
bathing, etc.'

(8) *'igihí hi 'ogiáagaó 'oga hápií ...*
man he bigness field went
'The men all went to the field.'

Crucial to Everett's analysis is a refusal to recognise polysemy. According to him, because the word *'ogi* means 'big' in some contexts, it means 'big' in all contexts. Though an NP like *hiaitíihí hi 'ogi* [lit. people 3p big] appears to express the same meaning as English 'all the people', for Everett its true meaning is something like "people's bigness"; and though an NP like first-person pronoun *ti* followed by *'ogi* appears to express the same meaning as English 'we', again this is a distortion of the true Pirahã meaning, which Everett glosses as "my bigness".[5] To take the edge off these bizarre-sounding glosses, Everett asserts that: "much of Pirahã is largely incommensurate with English and therefore translation is simply a poor approximation of Pirahã intentions and meaning" (p. 624, Note 5).

When challenged by Wierzbicka (2005: 641) on the issue of polysemy, and with the counter-claim that the Pirahã expression *hi 'ogi* means ALL, Everett (2005) declined to say anything on the subject.[6] In my view, there are clear indications in Everett's materials that *'ogi* is a polysemous item, and that it ought to be possible to contrast the meanings expressed by phrases such as *'igihí 'ogi* 'big men' and *'igihí hi 'ogi* 'all (the) men'.[7]

*No exponent of BODY in Tidore, Kuuk Thaayorre and Tiriyó?*

In the Introduction to a recent collection of cross-linguistic studies of body-parts, Enfield, Majid and van Staden (2006: 145) assert that "Several languages do not have a general term meaning 'body'". Upon examining the papers on which such a claim could be based, however, the editors' definitive pronouncement hardly seems justified (cf. Wierzbicka 2007a). In one of the languages, the Indonesian language Tidore, the author (van Staden 2006: 332) states clearly that there is a word for BODY (namely, *badan*), albeit that this word is a loan word from Malay.[8]

As for Kuuk Thaayorre, when Gaby (2006: 206) reports that the word *pam-minj*, literally 'true man', "may be used to refer to the specifically physical presence of a human (including that purely physical human entity, the corpse)", she seems to provide evidence that *pam-minj* indeed has 'body' as one of its meanings.[9]

In the same volume, Meira (2006: 277) claims that the Amazonian language Tiriyó "has no word for 'body'", because although the word *punu* is the main word in the hierarchy (i.e., the apparent superordinate term for individual body-part words), it also means 'flesh, meat', and in Meira's opinion "there seem to always be 'fleshy' overtones ... [it] seems to be a case of semantic generality". But a fieldworker's impressions are no substitute for a thorough semantic investigation; for example, checking whether contrastive readings are possible, whether

any derivational processes select one putative meaning over another, whether any morphosyntactic properties attach to one meaning or the other, whether they have different antonyms (e.g., 'body' – 'spirit', 'flesh' – 'bones'), whether expressions such as the following could be expressed in Tiriyó: 'the body of a man is not like the body of a woman', and so on. The author did none such (or at least, he reported nothing of the sort). Also relevant would be whether hard body-parts such as bones and teeth are regarded as "parts of the *punu*"; if so, one could hardly maintain the notion that *punu* has "fleshy overtones".

*Final comments on identification issues*

For other discussions about apparent "missing" semantic primes in various languages, see Bohnemeyer (1998, 2003) and Goddard (2001) on BEFORE and AFTER in Yucatec Maya, van Brakel (2002: 151) and Wierzbicka (2007b) on SEE, Shi-xu (2000) and Chappell (2002: 270–271) on FEEL in Chinese, and Myhill (1996) and Durst (1999) on BAD in Biblical Hebrew. The chapters in this volume by Marie-Odile Junker on East Cree, Mengistu Amberber on Amharic, Kyung-Joo Yoon on Korean, and Emily Knight on Bunuba are also notable contributions to the literature on cross-linguistic manifestations of primes.

The robustness of the NSM primes is dramatised when they are compared with various non-prime meanings which have been proposed as universals of human experience or cognition. As demonstrated in Goddard (2001), most non-prime terms—including impressionistically "basic" items of English vocabulary such as 'go', 'eat', 'sit', 'hit', 'hot', 'tree', and 'rock'—collapse on even a small sample of languages. On the current evidence, however, semantic primes do appear to "show themselves" in all human languages.

This is not to say that the issue has been settled beyond doubt. Though in my estimation, no convincing counter-evidence against the Strong Lexicalisation Hypothesis has yet been forthcoming, each proposed counter-example deserves to be examined and considered on a case-by-case basis. It is not theoretically inconceivable that some languages could have one or more "lexical gaps" in their inventory of semantic primes. Junker (this volume Ch 6) reports an apparent lexical gap for the prime PART in East Cree. In my view, Junker's treatment, careful though it is, still leaves some open questions,[10] but it is worth pondering what the consequences would be if such a result were to be established beyond doubt. It would bring into question the "meta-semantic adequacy" of the language concerned, i.e., whether the language provides the resources for explicating its own meanings in language-internal terms. The expressive power of the language would be redeemed, of course, if it could be shown to possess one or more language-specific semantic primes which covered the same territory as the missing primes; but this in turn would raise issues of cross-linguistic incommensurability. We cannot pursue these matters further here, but it will be clear that the theoretical stakes are high.

## 3. The universal grammar of semantic primes

Universal semantic primes have an inherent grammar—a "conceptual grammar"—which is the same in all languages; that is, each semantic prime has certain combinatorial properties (Goddard and Wierzbicka 2002: 41–85) by virtue of the particular concept it represents. The formal realisations (marking patterns, word order, constituent structure, etc.) may differ from language to language without these underlying combinatorial properties being disturbed. Because their inherent syntactic properties are manifested in all languages, semantic primes bring with them a substantial slab of universal syntax: the syntactic properties of semantic primes are literally universals of syntax. Although groups of primes do share particular properties and can be regarded as falling into natural classes (Goddard this volume Ch 3), it is equally true that virtually every prime has some idiosyncratic properties, giving each prime a distinctive syntactic signature.

The syntactic properties of semantic primes include the following: (i) basic combinatorics, e.g., the fact that substantives and relational substantives can combine with specifiers—THIS THING, THE SAME SOMEONE, SOMEWHERE ELSE, ONE PART, MANY KINDS; (ii) basic and extended valencies, e.g., the fact that the prime DO can occur not only in its basic SOMEONE DID SOMETHING frame, but also in extended frames such as SOMEONE DID SOMETHING TO SOMETHING (or TO SOMEONE), SOMEONE DID SOMETHING TO SOMETHING WITH SOMETHING; (iii) propositional complement possibilities of primes like KNOW, THINK, and WANT, e.g., the fact that KNOW and WANT can occur in frames like I KNOW THAT SOMETHING HAPPENED or I WANT SOMETHING TO HAPPEN NOW.

The cross-linguistic viability of the current model of NSM grammar has been checked in considerable depth and detail across a range of typologically divergent languages, and this process is continued in the present volume. At the same time, however, it is important to stress that this model is not presented as a "just so" story, complete in itself and immune to empirical disconfirmation. For example, Goddard and Karlsson (this volume Ch 8, 2004; cf. Goddard 2003a) argue on the basis of facts from Swedish and other Scandinavian languages, that earlier NSM proposals about the complementation syntax of THINK need to be constrained.

In the following subsections I will run quickly through some of the main aspects of the NSM grammar. Fuller treatment is given in Goddard and Wierzbicka (eds 2002); and in Chapter 3 of this volume. In the interests of brevity, I will employ some minimal "common coin" grammatical terminology, such as predicate, argument, complement, and adjunct. For an account of how these terms can be characterised within NSM itself, using an exemplars and lexical prototypes approach, see Goddard and Wierzbicka (2002); for cautionary remarks, see Lehrman (2006).

### 3.1 Predicates, complements and valency options

We can think of a simple NSM clause as consisting of a predicate, such as HAPPEN, DO, SAY, THINK, or WANT, together with one or more substantive phrases whose nature is constrained by the identity of the predicate. By substantive phrases is meant either a simple substantive, such as SOMETHING, SOMEONE, SOMEWHERE, or a more elaborated phrase based on a substantive (see next section). In addition, the simple NSM clause may contain adjunct phrases of time or place.

The substantive phrases occurring with particular predicates can be regarded either as arguments or as complements, depending on whether or not they can be referring expressions (detectable in formal terms by whether they can be specified by the determiner THIS). In the following arrays for HAPPEN and DO, the predicates appear in small caps, the argument substantives are underlined, and the complement substantives are unmarked. In addition to their minimal frame, predicates typically also allow extended frames in which additional arguments identify or fill out the aspects of the situation implied by the nature of the predicate. These optional extras are termed "valency options". For example, the predicate HAPPEN allows us to speak not only of 'something happening', but also of 'something happening to someone' or 'something happening to something'. Borrowing from the usual set of semantic role labels, this additional argument can be labelled an "undergoer" valency option.

(9) a. something HAPPENS
b. something HAPPENS to someone/something [undergoer]

Likewise, with predicate DO it is possible to add an additional argument and speak of 'doing something to someone' or 'doing something to something', and the additional argument in this case is conveniently referred to as a "patient" valency option.[11] This frame can be further extended to speak of 'doing something to something with something', and the additional argument can be labelled as "instrument". Another option for DO is the "comitative" option, as when we speak of 'doing something with someone'.

(10) a. someone DOES something
b. someone DOES something to someone/something [patient]
c. someone DOES something to someone/something with something [instrument]
d. someone DOES something with someone [comitative]

So far there will have been few surprises in terms of content, as many linguists accept some notions of "undergoer", "patient" and "instrument". Typically, however, they are thought of as independent entities of some kind (semantic

roles, thematic roles), rather than as argument slots of basic semantic predicates such as HAPPEN and DO. From an NSM point of view, the idea of an instrument, for example, exists only insofar as one can think about DOING something: it is a conceptual possibility that is "opened up" by the nature of DOING itself, and which is implicitly tied to the concept of DOING. (One linguist who has explicitly identified semantic roles as the argument slots of basic semantic predicates is Ray Jackendoff (1990: 127). Compared with his position, the distinctive aspect of the NSM proposal is the claim that the basic predicates are universally lexicalised meanings like semantic prime DO, rather than abstract conceptual functions, such as Jackendoff's predicate AFF "affect").

Every NSM predicate has a set of valency options, ranging from the very simple (as with DIE—one option only: 'someone dies') to the highly elaborate, as with SAY. In some cases, NSM researchers propose valency options which are seldom recognised in mainstream grammars and which may have no standard labels. For example, it is claimed that semantic prime THINK universally allows a "cognitive topic" valency option, such that one can say, in all languages, the semantic equivalent of a sentence like 'I was thinking about this someone (person) (this thing, this place, etc.)'. The full valency array for THINK is shown below. Notice that the third and fourth frames show sentential complement options: ways in which an expression analogous to a full sentence can be embedded inside the scope of THINK. Cross-linguistic studies have shown that in many languages the propositional frame is rather restricted in its distribution and range of application, compared with the "quasi-quotational" frame; see Chapter 8 for more detail.

(11) a. someone THINKS about someone/something [topic of cognition]
b. someone THINKS something (good/bad) about someone/something [topic + complement]
c. someone THINKS like this: " – – " [quasi-quotational thought]
d. (at this time) someone THINKS that [ —— ]$_S$ [propositional complement]

Perhaps the most complex array of syntactic frames is possessed by semantic prime SAY. In addition to its minimal frame with a complement, these include an addressee option, a topic option, and an option for specifying the "verbal means". There is also a direct speech frame. Other frames for SAY are discussed in Chapter 3.

(12) a. someone SAYS something
b. someone SAYS something to someone [addressee]
c. someone SAYS something about something [locutionary topic]
d. someone SAYS something in some (other) words [verbal means]
e. someone SAYS: " – – " [direct speech]

More details about the syntax of predicate primes can be found in Chapter 3. It can be mentioned here, though, that all predicates can combine with the "meta-predicates" or "predicate operators" NOT and CAN.[12]

### 3.2 Substantives and substantive phrases

The substantive primes include the categorical elements SOMEONE/WHO, SOMETHING/WHAT, PLACE/WHERE, and TIME/WHEN, along with more specific items such as PEOPLE, BODY, and WORDS, the relational substantives PART and KIND, and the deictics I and YOU. Except for the deictics, which are syntactically inert, the substantives can combine with a variety of other elements to form what can be termed "substantive phrases". Generally speaking, substantive phrases can occur in any position in which a simple substantive would make sense.

The elements which combine directly with substantives include: (a) determiners, e.g., 'this someone (person)', 'the same thing', 'somewhere else', (b) quantifiers, e.g., 'two things', 'many people', (c) evaluators, e.g., 'something good', 'something bad', and (d) descriptors, e.g., 'something small'. They can also be modified by *like*-phases, e.g., 'people like this', 'someone like me'.

There is much more to the grammar of substantive phrases than can be covered in this brief sketch. For example, the relational substantives KIND and PART each can head phrases *sui generis*, which can be termed "classifier phrases" and "part-hood phrases", respectively, such as 'one thing of this kind' and 'one part of this thing'. Some of the quantifier primes—ONE, TWO, and SOME—have a partitive or "subset" valency option, such that one can say, in all languages, equivalents of phrases such as: 'one of these people/things', 'two of these people/things', 'some of these people/things'.

### 3.3 Complex sentences

There are several ways in which multiple clauses can be included within single sentences in the NSM grammar. Most obviously, there is an inherently biclausal construction connected with the semantic prime IF, exemplified in sentences such as: 'If you do this, people can think something bad about you'.

As mentioned, semantic primes KNOW and THINK both have the capacity to take propositional complements (in English-based terminology, *that*-complements). For example, 'I know that something bad happened to this someone (person)'. The semantic prime WANT is also complement-taking, but it has special characteristics that set it apart from the other complement-taking predicates (Harkins 1995). Although WANT can also take something akin to a full sentential complement (in constructions such as 'I want you to do something' and 'I don't want this to happen'), there is a special construction type for when the subject of a person's 'want' is the wanter him or herself, as in 'I want to do something', 'I want to know something', and so on. In the traditional generative

terminology the construction was termed an "EQUI-clause"; in a more recent terminology it is a "control" structure involving "little PRO". Behind both terms is the idea that the apparently missing subject of the complement clause is actually present in some sense (either in deep structure, or in a form with a null phonological realisation). The NSM view, on the other hand, is that "what you see is what you get", i.e., that what might seem by analogy with a different construction to be incomplete is actually a complete but compact structure.

There are several other complex sentence structures found in the natural semantic metalanguage. There are "quasi-relative clauses" headed by PLACE or TIME; for example, 'I am far from the place where I live'. There are adverbial clauses of time and reason, e.g., 'when something like this happens, ...', 'it happened like this because someone wanted it'. There are "analogy clauses" with LIKE as a linking element, e.g., 'I want to do something bad to this someone (person), like this someone (person) did something bad to me'.

### 3.4 Formal variations between NSMs

Obviously the natural semantic metalanguage manifests itself differently in different individual languages. We have already seen examples of language-specific variation at the lexical level—different patterns of polysemy, allolexy and the possibility of portmanteau realisations of combinations of primes. Similar formal variations are to be found in the language-specific realisation of the grammar of the natural semantic metalanguage, in particular in morphology and morphosyntax. Some differences in marking are clearly superficial. For example, in English the subset valency of the quantifier primes involves a prepositional phrase introduced by *of* (e.g., *two of them*), whereas in Malay the comparable preposition is *daripada* 'from' (e.g., *dua orang daripada mereka* [two people from them]). Likewise, it is no surprise that the addressee valency option of SAY can be marked by an allative marker in one language (e.g., English *to*), a locative marker in another (e.g., Yankunytjatjara *-ngka/-la*), a dative marker in yet another, and so on. In such cases, the relevant expressions can be matched up both formally and, more importantly, semantically: there is no specifiable semantic difference between them.

Perhaps more challenging is the finding that there can be constituent structure differences between semantically identical expressions. For example, the subset valency of quantifiers in Lao involves a "fronted" noun-phrase denoting the larger set: *Khon$^{2}$ nii$^{4}$, sòòng$^{3}$ khon$^{2}$* [people this, two people] 'two of these people' (Enfield 2002). Or to take another example, in Spanish an "equi" complement of WANT appears in an infinitival clause, e.g., *Quiero ir* [I-WANT go:INF] 'I want to go', whereas a full clausal complement of WANT is introduced by complementiser *que* and takes a subjunctive verb, e.g., *Quiero que tu vayas* [I-WANT that you go:SUBJ] (Travis 2003). Despite the formal differences, however, it is not possible to state any paraphrasable difference between the meanings expressed

by the relevant constructions, and one can always transpose a sentence in a Lao-based NSM, for example, into an English-based NSM, or into a Spanish-based NSM, and so on. In this sense, the various language-specific versions of the natural semantic metalanguage are isomorphic. Anything that can be said in one NSM, can be said in any NSM.

To say this is not to minimise the importance or interest of the language-specific formal features of those individual languages. On the contrary, as we identify the language-specific aspects of how the natural semantic metalanguage is realised in a particular local language, we are identifying the essential morphosyntax of that local language; and this can be an important part of an overall grammatical description of that language. But from the point of view of cross-linguistic semantics, the crucial thing is that natural semantic metalanguages are expressively equivalent and isomorphic.

### 3.5 Using the full NSM in explications

In a short sketch like this, it is hard to give an impression of how the different aspects of the semantic metalanguage work together to allow complex semantic explications to be composed inside the small vocabulary of semantic primes. Perhaps the best way to approach it is by way of one or two examples. The explication in [A] below has been advanced to capture the meaning of the English verb *to miss (someone)* (Goddard and Wierzbicka in press b).

[A] *Someone X misses someone else Y:*

a. when this someone X thinks about someone else Y, this someone feels something bad
like people feel when they think like this about someone:

b. "I was with this someone before,
when I was with this someone, I felt some good things
I know that I can't be with this someone now"

Let us track through this explication, one line at a time. Section (a) begins with a when-clause involving THINK ('when this someone X thinks about this someone'), followed by a main clause based on FEEL. Many explications for emotion concepts begin in this fashion. Then comes an analogy clause introduced by LIKE: it echoes the structure of the first line, introducing a prototypical element: a reference to how 'people feel when they think like this about someone'. The occurrence of THINK in its quasi-quotational frame allows the presentation of a set of "subjective" components in first-person format, labelled as section (b) of the explication. First comes 'I was with this someone before', which utilises locational BE with its comitative option ('being with someone'). Then comes a when-clause and a main clause with FEEL; but in this case, the complement of FEEL appears as 'some good things' (the contrast between feeling 'something good' and feeling 'some good things' is subtle but available in

the NSM). Finally, a clause with KNOW and a propositional complement: 'I know that I can't be with this someone now'.

For a second example, we will take not a semantic explication, but a cultural script written in the natural semantic metalanguage. Cultural scripts are the main mode of representation in the theory of ethnopragmatics, which is the pragmatic sister theory of the NSM approach to semantics (Wierzbicka 2003 [1991]; Goddard and Wierzbicka eds 2004; Goddard ed. 2006). Essentially, cultural scripts represent hypotheses about shared understandings of widely held social attitudes pertaining to ways of speaking, and the like. The script presented in [B] below is intended to capture a characteristically Anglo speech strategy, whereby one may go about getting someone to do something by way of making a "suggestion" to them; that is, by putting to the other person that they may want to think about doing something, and may then decide for themselves that they want to do it (Wierzbicka 2006a). This is one Anglo strategy among many which allows one to avoid seeming to "impose" one's will on someone else. For present purposes, we are not so much interested in the content or validity of this script, as in the way it utilises different aspects of the natural semantic metalanguage.

[B] *An Anglo cultural script for a "suggestive" approach to influencing others*

a. people think like this:
b. when I want someone to do something,
    it can be good if I say something like this to this someone:
c. "maybe you will want to think about it
    maybe if you think about it, you will want to do it"

To begin with, the explication as a whole is introduced by a quasi-quotational complement of THINK: 'people think like this'. The when-clause in component (b) presents WANT with a complement clause ('when I want someone to do something'), followed by GOOD as a main predicate, in combination with CAN ('it can be good ...'). The presence of the CAN nuances the proposition, presenting it as an option or possibility (in contrast to the simpler 'it is good ...'). Used as a main predicate in this way, semantic prime GOOD can take a clausal complement (Goddard and Wierzbicka 2002: 65–66), introduced (in English) by 'if': 'it can be good if I say something like this to this someone'. The complement clause has SAY as its main predicate, with its addressee valency option. The complement of SAY appears as SOMETHING LIKE THIS, thereby opening the way for a set of first-person "message" components, presented in (c). Both these messages begin with the prime MAYBE. They offer the addressee the idea of thinking about something, with a conditional if-sentence as part of that possibility: 'maybe if you think about it, you will want to do it'.

Working through explications and scripts like those in [A] and [B] for the first time, newcomers to NSM often remark: "so it's not as simple as it seems".

And indeed it isn't. Despite its small lexicon, the natural semantic metalanguage has a surprisingly rich "texture", especially in comparison with the austere function-argument structure of predicate calculus and other logical systems. As mentioned, virtually every semantic prime has its own "mini-grammar"—its own conceptual grammar—and the possibilities for mixing and matching the allowable combinations enable a great deal of textual complexity in explications. Is this a good thing or a bad thing? In my view, this is not really an appropriate question to ask about an empirical finding. It appears to be an empirical fact that the structure of human concepts has this kind of intricate structure. Does it not make the NSM harder to learn and to use than one would expect? Perhaps so. But again, this appears to be a fact of life.

Moreover, the structure of many words and concepts in the "concrete" lexicon is rendered yet more complex by the fact that their explications involve not only semantic primes, but also semantic molecules.

## 4. Semantic molecules

Despite the central theoretical role of semantic primes in the NSM theory, not all meanings can be resolved simply or directly into semantic primes. Words from the concrete lexicon, in particular, cannot normally be decomposed directly into primes. For example, plausible explications for words like *cat*, *mouse* and *horse* must begin with the component 'animals [M] of one kind', and plausible explications for *oak*, *elm*, and *pine* must begin with the component 'trees [M] of one kind' (molecules are marked with [M] in explications). The concepts of 'animals' and 'trees' are themselves complex and further decomposable into semantic primes, but they function as units in the explications of many other concepts. By the term semantic molecule, then, we understand a complex lexical meaning which functions as a semantic unit (or "chunk") in the structure of other, more complex concepts.[13]

### 4.1 Semantic molecules in the English lexicon

To develop an idea of the role, range and nature of semantic molecules in the English lexicon, it is useful to consider several explications which have been developed independently of these questions. In this, we are following the normal NSM practice of trying to induce generalisations about semantic structure from empirical work. For reasons of space, we will have to confine ourselves to partial explications. Explication [C] below is a partial explication for a natural kind term—*cats* (based on Wierzbicka 1985 and Goddard 1998). It follows the semantic template[14] for natural kind terms: (a) category within the taxonomic hierarchy, (b) habitat, (c) size, (d) appearance, (e) behaviour, (f) relation with people (the last two sections of the template are omitted in explication [C]).

[C] A partial explication for *cats*

| | | |
|---|---|---|
| a. | animals [M] of one kind | CATEGORY |
| b. | animals [M] of this kind can live in places where people live, because people want this | HABITAT |
| | sometimes they live near places where people live | |
| c. | they are not big | SIZE |
| | a person can pick up [M] one with two hands [M] | |
| d. | they have soft [M] fur [M] | APPEARANCE |
| | they have a round [M] head [M] | |
| | they have pointed [M] ears [M] | |
| | their ears [M] are on two sides of the top [M] part of the head [M] | |
| | their eyes [M] are not like people's eyes [M] | |
| | they have some long [M] hairs [M] near the mouth [M], they stick out [M] on two sides of the mouth [M] | |
| | they have a long [M] tail [M] | |
| | they have soft [M] feet [M] | |
| | they have small sharp [M] claws [M] | |

Running through the four sections presented in [C], we see that section (a) includes a taxonomic "life form" category, 'animals [M]', as a semantic molecule. Section (b) gets by without any additional semantic molecules: it is phrased almost exclusively in semantic primes.[15] Regarding section (c), the "size" section, Wierzbicka (1985) has argued that anthropocentrism pervades human construal of the physical world, and, in particular, that the human body furnishes a reference point for judgements of relative size. In the case of *cats*, this component depends on the potential to pick up a cat with two hands, with both 'pick up [M]' (a physical action verb) and 'hands [M]' (a body-part word) functioning as semantic molecules. The subsequent "appearance" section contains numerous molecules: yet more body-part terms ('head', 'ears', 'eyes', 'mouth', 'tail', 'feet'), and terms of two additional kinds: shape descriptors ('round', 'long', 'pointed'), and physical qualities ('soft', 'sharp').

Just as explication [C] can serve as an exemplar of natural kind terms, so explication [D] for *chairs* can stand as a (partial) exemplar of artefact terms. After the top-level "category" component, the semantic template continues with a "purpose" section specifying people's motivation for making such things. In the case of *chairs*, this obviously involves 'sitting'; roughly speaking, *chairs* are things of the kind people make so as to be able to sit comfortably while doing various things. To capture this notion, it is not necessary to employ 'comfortably' as a semantic molecule, because the required meaning can be rendered directly in semantic primes, as shown in the final line of [D]. What is necessary, however, is to employ 'sit [M]' as a semantic molecule, because the explication of 'sit' is much too complex (cf. Wierzbicka 2006b) to be substitutable directly into [D].

[D] A partial explication for *chairs*:

a. something of one kind CATEGORY
there are things of this kind because people want this
b. people want this because they want these things to be in places where people live PURPOSE
people want them to be in these places because they want people to be able to sit [M] on them
when they want to do something in these places for some time
people want people to be able to sit [M] on them at times like these
because they don't want them to feel something bad in their bodies at these times

Other bodily activity verbs such as 'eat' and 'drink' are also known to be prolific semantic molecules. They are needed in words for artefacts such as *spoon*, *plate*, *cup* and *bottle*, for food and drink words such as *meat*, *bread*, *tea* and *coffee*, and in many other contexts. It is not possible for reasons of space to discuss the structure of bodily activity verbs in any detail, but it is obvious that they call for the use of body-part terms as semantic molecules; for example, to explicate *eat* and *drink* one must include the molecule 'mouth'.

Explication [E] below is a partial explication for a complex physical activity verb: English *cutting* (Goddard and Wierzbicka in press a). Verbs of this kind follow the semantic template: (a) lexico-syntactic frame; (b) prototypical motivational scenario; (c) instrument; (d) how the person uses the instrument; (e) what is happening to the object; (f) potential outcome.

[E] A partial explication for *cutting*:
*Someone X was cutting something Y (e.g., some paper, a cake)*
*with something Z*

LEXICO-SYNTACTIC FRAME
a. someone X was doing something to thing Y with thing Z for some time
because of this, something was happening at the same time to thing Y
as this someone wanted

PROTOTYPICAL MOTIVATIONAL SCENARIO
b. people do something like this when they do something to something
because a short time before they thought about this something like this:
"I want this something not to be one thing anymore
I want this something to be two things
I want these two things to have straight [M] edges [M]"

INSTRUMENT
c. when someone does something like this, they do it with something
this something is not a part of this someone's body
this something has a sharp [M] edge [M]

The shared top-level component, the lexico-syntactic frame, can be phrased without recourse to semantic molecules. It characterises such verbs as activities by which a person produces some effect upon an object in a controlled fashion by means of an instrument. The next component, the prototypical motivational

scenario, begins with a certain intention (shared by *cut* and other "verbs of separation"): 'I want this something not be one thing anymore'. In the case of *cutting*, it continues: 'I want this something to be two things, I want these two things to have straight [M] edges [M]'. The word 'straight' is another example of a shape descriptor, but 'edges' represents a different kind of molecule from those we have previously observed. It can be termed an "ethnogeometrical concept" (Brotherson this volume Ch 10). The same molecule is found later in section (c): *cutting* requires an instrument with a 'sharp [M] edge [M]'.

The point of this exposition has been to illustrate the role of semantic molecules in different kinds of explications, and to establish that certain particular kinds of molecules appear to be common: body-parts, shape descriptors, physical qualities, ethnogeometrical terms, and bodily actions. Before moving to other issues about semantic molecules, a matter of clarification should be attended to.

There are many recurrent components across the lexicon which are not semantic molecules, in the sense under discussion here, because they are not encapsulated as the meanings of surface lexical items. For example, many nouns begin with top-level categorical components such as: 'one part of someone's body' (for body-part terms), 'living things of one kind' (for natural kind terms), 'a place of one kind where people live' (for words like *town*, *city*, *village*, etc). Likewise, many verbs contain high-level components related to semantic role or argument structure; for example: 'someone did something', 'someone did something to something/someone else', 'someone did something to something with something', 'something happened to something else because of it', and so on. Many verbs also contain a component like 'this someone did this because this someone wanted to do it', corresponding to the technical notion of volitional action. To take one further example, from a smaller segment of the lexicon, emotion adjectives such as *sad*, *annoyed* and *homesick*, conform to a semantic template which begins: 'someone feels something (good/bad), as people feel when they think like this: – – ' (followed by a prototypical cognitive scenario setting out certain characteristic thoughts and wants).

Such recurrent but "non-molecular" components can be extremely significant for the interface between lexical and grammatical semantics, and for the creation of lexical classes. They are simple enough in their internal semantic structure, however, to be spelt out in relatively short strings composed purely of semantic primes. They are not semantic molecules, because they are not the meanings of surface lexical items.

How many productive semantic molecules are there? At this formative stage of research, the answer is not very clear. Wierzbicka (in press) has hazarded an estimate (for English and Polish) of 100–200. It is known that productive molecules are drawn from at least nine categories. The first six have already been briefly exemplified: (i) parts of the body, such as 'hands', 'mouth', 'legs';

(ii) physical activities, such as 'eat', 'drink', 'sit'; (iii) shape descriptors, such as 'long', 'round', 'flat'; (iv) physical qualities, such as 'hard', 'sharp' and 'straight'; (v) ethnogeometrical terms, such as 'edges' and 'ends'; and (vi) taxonomic concepts, such as 'animal', 'bird' and 'tree'. To this list one can add at least three further categories, which will not be justified here (cf. Wong, Goddard and Wierzbicka to appear; Wierzbicka 2006a; Goddard and Wierzbicka to appear): (vii) macro-terms from the natural environment, such as 'ground', 'sky' and 'sun'; (viii) "elemental" concepts (for want of a better term), such as 'water' and 'fire'; and (ix) basic social categories, i.e., kinds of people, such as 'men', 'women' and 'children'.

## 4.2 Levels of semantic nesting

We have already seen that there can be molecules within molecules. How many levels of nesting are there? A useful starting point is with parts of the body (Wierzbicka 2007a). Consider the word *head* (in the sense of a person's *head*).[16] As explicated in [F], it requires in its explication not only semantic primes establishing its parthood relationship with the body ('one part of someone's body', etc.) and position ('above all the other parts of this someone's body'), but also an indication of its shape—in the form of the semantic molecule 'round'. The meaning 'head' in turn functions as a molecule in certain other body-part meanings, such as *eyes*, *ears* and *hair*, among others.

[F] *head$_1$ (someone's head)*

a. one part of someone's body
b. this part is above all the other parts of this someone's body
c. this part is round [M]

Shape descriptors are needed in many other body-part concepts too. For example, the concept of *legs* requires the descriptor 'long', as shown in [G]. 'Legs' in turn is a molecule in verbs like *walk* and *run*, among others.

[G] *legs (someone's legs)*

a. two parts of someone's body
b. these two parts are below all the other parts of this someone's body
c. these two parts are long [M]
d. these two parts of someone's body can move as this someone wants
e. because people's bodies have these parts, people can move in many places as they want

Already we can discern several levels of semantic nesting: shape descriptors (like 'long' and 'round') are molecules inside body-part concepts (like 'legs' and 'head'); and body-part concepts in turn are molecules inside bodily

action verbs (like 'run' and 'eat'). But what about the semantics of shape descriptors?

Wierzbicka (2003, 2006a, 2007a) has investigated these in depth, and has drawn a remarkable conclusion: that shape concepts depend in part on the concept of 'hands'.[17] The nub of the argument is that shape concepts designate properties which can be detected not only visually, but also by touch—i.e., by the touch of our hands. This aspect is crucial to capturing the tangibility and physicality of shape concepts (as compared with colour concepts, for example, which depend purely on seeing). A sample explication—for one meaning of the word *long*—is given in [H].[18]

[H] *something long (e.g., a tail, a stick, a cucumber)*

a. when someone sees this thing, this someone can think about it like this:
b. "two parts of this thing are not like any other parts,
because one of these two parts is very far from the other"
c. if someone's hands [M] touch this thing everywhere on all sides,
this someone can think about it in the same way

It might seem at this point that we are facing a fatal circularity. How can shape descriptors depend on a body-part concept, namely 'hands', while at the same time body-part concepts depend on shape descriptors? In fact, there is no circularity on account of another remarkable result (Wierzbicka 2003, 2006a, 2007a). Of all the body-part concepts, it seems that 'hands' alone can be explicated purely in terms of semantic primes, without recourse to any shape (or other) semantic molecules. The explication for 'hands' is shown in [I].

[I] *hands (someone's hands)*

a. two parts of someone's body
b. they are on two sides of this someone's body
c. these two parts of someone's body can move as this someone wants
d. these two parts of someone's body have many parts
e. if this someone wants it, all the parts on one side of one of these two parts can touch
all the parts on one side of the other one at the same time
f. because people's bodies have these two parts, people can do many things with many
things as they want
g. because people's bodies have these two parts, people can touch many things as they want

Wierzbicka (2003) argues that the foundational status of the concept of 'hands' makes sense from the point of view of an embodied concept of cognition. "[H]uman hands", she argues, "mediate, to a large extent, between the world and the human mind", because of the crucial role played by "handling" things and by touching things in an exploratory way with the hands. Equally, the hands are our principal "bodily instruments" for making things, for using

things, and for doing things of many other kinds. It makes a lot of sense, therefore, that the concept of 'hands' is a foundational semantic molecule in so many human concepts.

We can conclude that there are up to four levels of semantic nesting within highly complex concepts, such as those for natural kinds and artefacts. In the explication for *cats* or *chairs*, for example, the most complex molecules are bodily action verbs like 'eat [M]' or 'sit [M]'. They contain body-part molecules such as 'mouth [M]' and 'legs [M]'. These in turn contain shape descriptors, such as 'long [M]', 'round [M]' and 'flat [M]', and they in turn harbour the molecule 'hands [M]', composed purely of semantic primes. A further level of nesting occurs when natural kind terms themselves function as semantic molecules at a shallow level of semantic structure. For example, words for unfamiliar species such as *tigers* and *zebras* contain a "likeness" reference to familiar natural kinds, such as 'cats' and 'horses', respectively; endonymic terms like *purr* and *saddle* also contain references to 'cats' and 'horses', respectively (Goddard 1998: 241–242).

### 4.3 Universal vs. language-specific semantic molecules

It seems likely on current evidence that some semantic molecules are universal. This applies to concepts which are foundational for many other concepts and/or for large lexical classes. The molecule 'hands' is a prime candidate, and cross-linguistic surveys appear to support this position, once sufficient attention is focused on questions of language-specific polysemy (Goddard 2001; Wierzbicka 2007a). Other candidates for universal semantic molecules are certain other body-parts such as 'eyes' and 'ears' (Wierzbicka 2007a), basic social categories like 'men' and 'women' (Goddard and Wierzbicka to appear), and perhaps also the sociobiological concept 'mother', given the foundational status of the mother-child relationship in kinship systems (Wierzbicka 1992; Foley 1997: 133).

It is also clear that some semantic molecules are language-specific. This is only to be expected for high-level molecules such as taxonomic categories, since it is well established that there are languages which lack exact equivalents for words like 'animal', 'bird' and 'tree' (Goddard 2001). Likewise, it is not too surprising that some languages employ particular concepts as semantic molecules while others do not. For example, in Polish the word *grzyb* 'mushroom' functions as a semantic molecule: there are many common Polish words for kinds of *grzyby* 'mushrooms', and various other endonymic words which also include 'grzyb [M]' in their meanings.

Perhaps more surprising is the claim that lower-level molecules such as shape descriptors and ethnogeometrical terms can also vary somewhat from language to language. But as mentioned in Note 18, Wierzbicka (2006c, in press) argues that English 'long [M]' does not exactly match the comparable Polish molecule 'podłużny [M]' 'elongated, oblong'; and Brotherson (this volume

Ch 10) argues that English 'ends [M]' differs from its nearest counterpart 'tapu [M]' in Makasai (East Timor). Wierzbicka (2004, 2006b, 2007b) has argued that in English 'colour [M]' functions as a semantic molecule in words like *red*, *blue*, *green*, etc., whereas many other languages lack "colour words" in the true sense, because their visual descriptor words do not involve any comparable molecule.

It should be clear that the concept of semantic molecules is an extremely fertile one, with multiple ramifications for our understanding of the overall structuring of the lexicon, for lexical typology, for language acquisition, and for language and cognition studies. In coming decades we can expect semantic molecules to be one of the most vibrant research fronts in semantics.

## 5. The present volume

Though the focus of this chapter, and of this volume as a whole, is on the linguistic aspects of the natural semantic metalanguage, it is important not to lose sight of the fact that the NSM approach has far-reaching applications in a whole range of practical endeavours, such as intercultural communication, lexicography, language teaching, and language therapy. Because the NSM metalanguage has such broad applications and implications, it is important to get it "just right".

The present volume presents original contributions to the NSM program across a range of research fronts. Part One is about the theory of the metalanguage. Chapter 1 overviews the current state of the NSM system. Chapter 2, by Goddard and Wierzbicka, is the first full exposition of a "new" semantic prime —specificational BE—and of a new syntactic frame for the prime THIS/IT. Goddard's Chapter 3 is an inquiry into the deep syntactic patterning of the natural semantic metalanguage.

The three studies in Part Two of the volume are original contributions to the growing documentation of "whole metalanguage" studies, i.e., studies of how semantic primes and their associated grammar manifest themselves differently in the languages of the world. The three languages—Amharic, Korean and East Cree—are of different language families and come from very different geographical and cultural zones. As well as raising certain theoretical questions for our model of the NSM itself, each chapter can be read as a portrait in miniature of the language concerned. Amberber's chapter is the first full-scale application of NSM to a Semitic language. Yoon's chapter is a condensed version of a book-length treatment of Korean NSM (Yoon 2006). Junker's description of East Cree NSM is the first full description of an NSM based on a polysynthetic language of any language family.

Part Three consists of three studies which target problematic issues about particular semantic primes. In a sense they are problem-solving exercises. Knight's chapter on the Australian language Bunuba is a meticulous language-

internal study of an exponent of a semantic prime which exhibits five-way polysemy. Goddard and Karlsson's study of THINK in Swedish, and Elouazizi and Trnavac's chapter on MOMENT in Tarifyt Berber both propose significant revisions of our understanding of the primes they are concerned with.

Part Four contains descriptive-analytical studies of lexical and grammatical semantics in four more languages—Makasai (East Timor), Koromu (Papua New Guinea), Russian and Japanese. In their introductory sections, each author reports on the exponents of semantic primes in their language, before turning to their particular case study, be it ethnogeometrical terms (Brotherson), inalienable possession (Priestley), or value terminology (Gladkova, Hasada).

Taken together, the volume presents an impressive profile of the diverse research fronts in a flourishing semantic paradigm.

## Abbreviations

| | | | | | |
|---|---|---|---|---|---|
| 1SG | first person singular | APPROX | approximation | p/PL | plural |
| 2SG | second person singular | ASSR | assertive | S | sentential complement |
| 3SG | third person singular | INF | infinitive | SUBJ | subject |
| ACC | accusative case | PAST | past tense | | |

## Notes

1. For an insightful recent review of NSM, see Lehrman (2006). NSM semantics has also been ignored, marginalised and condemned. A few examples will suffice. In the 600 odd pages of the supposedly comprehensive *Handbook of Contemporary Semantic Theory* (Lappin ed. 1996) there is only one, single-line mention of Anna Wierzbicka (in a short paragraph on cognitive linguistics), and no reference to the NSM theory as such. As for condemnations, it is hard to go past Bart Geurts' (2003) proclamations about the futility of reductive paraphrase as a technique of semantic analysis. He finds it particularly galling that anyone would question the use of technical terminology and expert theories to represent ordinary people's cognition: "Frankly, *qua* semanticist, I just don't care that my mother-in-law fails to grasp my theories, and it does not prevent me from testing my predictions against her intuitions as a native speaker". Like many critics, Geurts has not done his homework. He sums up the NSM position as believing "that less than 20 words will suffice for a full-blown semantics of any language" (p. 226).

2. Astonishingly, many mainstream semanticists still miss the fundamental point that any symbols used in a semantic representation can only be understood by virtue of their relationship with some other, previously known, symbols. They seem to think that presenting their units of semantic representation in capital letters and enclosing them in square brackets "detaches" them from natural language, and thereby frees the analyst from any danger of implicit circularity.

3. On the other hand, it is known that English *person* has language-specific peculiarities which can be problematical in some contexts, cf. Wierzbicka (2002: 70–73).

4. On Everett's account, Pirahã also lacks exponents of THINK (since it uses the same verb as SAY to express such meanings; p. 629), and, no doubt, other primes as well.

5. Pirahã plural pronouns are formed by addition of *'ogi* to the singular (or unspecified) form, the same pattern as found in Mandarin Chinese [*wǒ-men* 1SG-PL, *nǐ-men* 2SG-PL, *tā-men* 3SG-PL],

and in many creoles. Given that there is evidence that Pirahã pronominal roots have been borrowed from a nearby Tupi-Guarani language, and that forms such as 'I-all' for 'we', 'you-all' for 'you-PL', and 'he-all' for 'they' are common in creoles, this is additional indirect evidence that *'ogi* is an exponent of ALL (as well as BIG). The prospect of wholesale borrowing of pronouns also suggests, as noted by Levinson (2005: 638), the possibility that Pirahã has undergone creolisation.

6. All Everett (2005) had to say about semantic methodology was that: "All semanticists know that the quantificational properties of a word are revealed by its truth conditions. I have pointed out that Pirahã has no word with the truth conditions of universal quantification" (p. 643). Though I could find no statements about the truth conditions of *hi 'ogi* in Everett's article, he does make some analogous claims about another possible expression of "universal quantification", namely, *báaiso* 'whole'. His point is that this word does not imply absolute exhaustiveness, as one would expect of a true equivalent of universal quantification. But 'all' in natural language is not the same as universal quantification in logic: in ordinary English 'all' does not entail absolute exhaustiveness either.

7. Likewise, it is hard to buy Everett's assertion that Pirahã *hói* is monosemous, so that *'ítíi'isi hói* is vague between 'a small fish' and 'one fish', and *tiobáhai hói* is vague between 'a small child' and 'one child' (2005: 623).

8. Especially in the case of basic concepts like 'body', 'think', 'feel' or 'want', it seems that languages often borrow a new word in order to disambiguate an earlier term which was polysemous. For example, as discussed by Hale (1994: 269), in some Misumalpan languages of Central America the English word for 'want' has been borrowed to replace a polysemous word meaning either 'want' or 'seek'.

9. It is true that in Kuuk Thaayorre, the words for a person's shadow, footprints, voice, and name can occur in certain "inalienable possession" constructions that are largely reserved for body-parts. But this property, which is common in many languages (cf. Chappell and McGregor eds 1996; Priestley this volume Ch 11), does not necessarily mean that the word for BODY doesn't really mean BODY.

10. For example, Junker does not indicate how East Cree speakers could go about constructing Cree-internal explications for words like 'head' and 'hands', i.e., how they could tackle components such a 'one part of someone's body' (for 'head') or 'two parts of someone's body' (for 'hands'), cf. Wierzbicka (2007a). Yet body-part terminology is arguably the canonical lexical domain for "part-hood" relations.

11. On closer examination, it is actually necessary (and theoretically fruitful) to distinguish the two "patient" frames: 'someone did something to someone else' vs. 'someone did something to something'.

12. To forestall misunderstanding, the universal semantic prime CAN does not respect the supposed distinction between "can of ability" and "can of possibility" which is taken for granted by many linguists under the influence of the logical tradition.

13. The Moscow School of semantics (Igor Mel'čuk and colleagues) has long championed the need for intermediate-level components, but without the constraint that they be lexical meanings in the language concerned.

14. A "semantic template" is a structured set of component types shared by words of a particular semantic class. The concept was first employed in explications for artefact and natural kind terms (Wierzbicka 1985), but it has since been elaborated and applied to words of many other kinds. While template structure appears to be fairly constant across languages (Goddard and Wierzbicka in press a), the details of each individual word meaning can differ considerably from those of its closest counterparts in other languages. Each encapsulates a specific cognitive scenario, and these scenarios are often culture-specific and reflect common local cultural practices.

15. *Cats* being domestic animals, the habitat component makes reference to them living with people or near where people live; and PEOPLE is a semantic prime. The comparable component

for some other natural kinds may well require semantic molecules; for example, *whales* and *tuna* presumably 'live in the sea [M]'.

16. Body-parts of other kinds of living things require distinct but related explications, i.e., words like *head*, *legs*, and so on, are polysemous between a primary sense based on the human body, and a secondary sense applying to other species by analogy with the human body. For example, the $head_2$ of a dog or a snake can be explicated as follows: 'one part of the body of a living thing of one kind (e.g., dog, snake); this part is like one part of people's bodies; this part of people's bodies is the $head_1$'.

17. Some physical quality concepts (Goddard and Wierzbicka 2007) also depend on the concept of 'hands'. For example, the concept of physical 'hardness' depends on the idea that if someone wants to "make an impact" on something hard (e.g., to break, deform, or scratch it), they cannot do this with the hands alone: they have to use some kind of instrument.

18. Wierzbicka (2006c) argues that the nearest Polish word *podłużny* 'elongated, oblong' differs slightly but discernibly from English *long* in component (b): *podłużny* requires only that two parts can be construed as 'far' apart (not necessarily as 'very far' apart). The difference accounts for differences in the range of use of the two words; a paperback book or a child's pencil-case, for example, can be described as *podłużny*, but hardly as *long*.

## References


Allan, Keith. 1986. *Linguistic Semantics*, *Vols I and II*. London: Routledge & Kegan Paul.

Bohnemeyer, Jürgen. 1998. Temporal reference from a radical pragmatics point of view. Why Yucatec Maya does not need to express 'after' and 'before'. *Cognitive Linguistics* 9(3): 239–282.

Bohnemeyer, Jürgen. 2003. NSM without the Strong Lexicalization Hypothesis. *Theoretical Linguistics* [Special Issue on "Natural Semantic Metalanguage", edited by Uwe Durst] 29(3): 223–226.

Brotherson, Anna. This volume. The ethnogeometry of Makasai (East Timor). In *Cross-Linguistic Semantics*, Cliff Goddard (ed.), 259–276. Amsterdam: John Benjamins.

Chappell, Hilary. 2002. The universal syntax of semantic primes in Mandarin Chinese. In *Meaning and Universal Grammar – Theory and Empirical Findings*, *Vol. I*, Cliff Goddard and Anna Wierzbicka (eds), 243–322. Amsterdam: John Benjamins.

Chappell, Hilary and William McGregor (eds). 1996. *The Grammar of Inalienability: A Typological Perspective on Body Part Terms and the Part-Whole Relation*. Berlin: Mouton de Gruyter.

Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Cruse, D. A. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.

D'Andrade, Roy. 2001. A cognitivist's view of the units debate in cultural anthropology. *Cross-Cultural Research* 35(2): 242–257.

Durst, Uwe. 1999. BAD as a semantic primitive: Evidence from Biblical Hebrew. *Pragmatics & Cognition* 7(2): 375–403.

Durst, Uwe (ed.). 2003. *Natural Semantic Metalanguage*. Special Issue of *Theoretical Linguistics* 29(3).

Eco, Umberto. 1999. *Kant and the Platypus*. London: Secker & Warburg.

Enfield, N. J. 2002. Combinatoric properties of Natural Semantic Metalanguage expressions in Lao. In *Meaning and Universal Grammar – Theory and Empirical Findings*, *Vol. II*, Cliff Goddard and Anna Wierzbicka (eds), 145–256. Amsterdam: John Benjamins.

Enfield, N. J., Asifa Majid and Miriam van Staden. 2006. Cross-linguistic categorisation of the body: Introduction. *Language Sciences* [Special Issue on "Parts of the body: Cross-linguistic categorisation", edited by Asifa Majid, N. J. Enfield and Miriam van Staden] 28(2/3): 137–147.

Enfield, N. J. and Anna Wierzbicka (eds). 2002. *The Body in the Description of Emotion*. Special Issue of *Pragmatics & Cognition* 10(1/2).

Everett, Daniel L. 2005. Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology* 46(4): 621–646.

Foley, William A. 1997. *Anthropological Linguistics: An Introduction*. Oxford: Blackwell.

Gaby, Alice R. 2006. The Thaayorre 'true man': Lexicon of the human body in an Australian language. *Language Sciences* [Special Issue on "Parts of the body: Cross-linguistic categorisation", edited by Asifa Majid, N. J. Enfield and Miriam van Staden], 28(2/3): 201–220.

Geurts, Bart. 2003. Semantics as lexicography. *Theoretical Linguistics* 29(3): 223–226.

Goddard, Cliff. 1991. Testing the translatability of semantic primitives into an Australian Aboriginal Language. *Anthropological Linguistics*, 33(1): 31-56.

Goddard, Cliff. 1994. Semantic theory and semantic universals. In *Semantic and Lexical Universals – Theory and Empirical Findings*, Cliff Goddard and Anna Wierzbicka (eds), 7–30. Amsterdam: John Benjamins.

Goddard, Cliff. 1998. *Semantic Analysis*. Oxford: Oxford University Press.

Goddard, Cliff. 2001. Lexico-semantic universals: A critical overview. *Linguistic Typology* 5(1): 1–66.

Goddard, Cliff. 2002. Ethnosyntax, ethnopragmatics, sign-functions, and culture. In *Ethnosyntax: Explorations in Grammar and Culture*, N. J. Enfield (ed.), 52–73. Oxford: Oxford University Press.

Goddard, Cliff. 2003a. 'Thinking' across languages and cultures: Six dimensions of variation. *Cognitive Linguistics* 14(2/3): 109–140.

Goddard, Cliff. 2003b. Whorf meets Wierzbicka: Variation and universals in language and thinking. *Language Sciences* 25(4): 393–432.

Goddard, Cliff. This volume. Towards a systematic table of semantic elements. In *Cross-Linguistic Semantics*, Cliff Goddard (ed.), 59–81. Amsterdam: John Benjamins.

Goddard, Cliff (ed.). 1997. *Studies in the Syntax of Universal Semantic Primitives*. Special Issue of *Language Sciences*, 19(3).

Goddard, Cliff (ed.). 2006. *Ethnopragmatics: Understanding Discourse in Cultural Context*. Berlin: Mouton de Gruyter.

Goddard, Cliff and Susanna Karlsson. 2004. Re-thinking 'think': Contrastive semantics of Swedish and English. In *Proceedings of the 2003 Conference of the Australian Linguistic Society*, Christo Moskovsky (ed.), [www.als.asn.au].

Goddard, Cliff and Susanna Karlsson. This volume. Re-thinking THINK in contrastive perspective: Swedish vs. English. In *Cross-Linguistic Semantics*, Cliff Goddard (ed.), 225–240. Amsterdam: John Benjamins.

Goddard, Cliff and Anna Wierzbicka. 2002. Semantic primes and universal grammar. In *Meaning and Universal Grammar – Theory and Empirical Findings*, *Vol. I*, Cliff Goddard and Anna Wierzbicka (eds), 41–85. Amsterdam: John Benjamins.

Goddard, Cliff and Anna Wierzbicka. This volume. New semantic primes and new syntactic frames: "Specificational BE" and "abstract THIS/IT". In *Cross-Linguistic Semantics*, Cliff Goddard (ed.), 35–57. Amsterdam: John Benjamins.

Goddard, Cliff and Anna Wierzbicka. 2007. NSM analyses of the semantics of physical qualities: *Sweet*, *hot*, *hard*, *heavy*, *rough*, *sharp* in cross-linguistic perspective. *Studies in Language* 31(4): 765–800.

Goddard, Cliff and Anna Wierzbicka. In press a. Contrastive semantics of physical activity verbs: 'Cutting' and 'chopping' in English, Polish, and Japanese. *Language Sciences*.

Goddard, Cliff and Anna Wierzbicka. In press b. Universal human concepts as a basis for contrastive linguistic semantics. In *Current Trends in Contrastive Linguistics: Functional and Cognitive Perspectives,* María de los Ángeles Gómez-González, Lachlan Mackenzie, and Elsa Gonzáles Álvarez (eds). Amsterdam: John Benjamins.

Goddard, Cliff and Anna Wierzbicka. To appear. *Men*, *women* and *children*: The semantics of basic social categories.

Goddard, Cliff and Anna Wierzbicka (eds). 1994. *Semantic and Lexical Universals – Theory and Empirical Findings*. Amsterdam: John Benjamins.

Goddard, Cliff and Anna Wierzbicka (eds). 2002. *Meaning and Universal Grammar – Theory and Empirical Findings*, *Vols I and II*. Amsterdam: John Benjamins.

Goddard, Cliff and Anna Wierzbicka (eds). 2004. *Cultural Scripts*. Special Issue of *Intercultural Pragmatics* 1(2).

Hale, Kenneth. 1994. Preliminary observations on lexical and semantic primitives in the Misumalpan languages of Nicaragua. In *Semantic and Lexical Universals – Theory and Empirical Findings*, Cliff Goddard and Anna Wierzbicka (eds), 263–284. Amsterdam: John Benjamins.

Harkins, Jean. 1995. Desire in Language and Thought: A Study in Cross-cultural Semantics. PhD Thesis, Canberra: The Australian National University.

Harkins, Jean and Anna Wierzbicka (eds). 2001. *Emotions in Crosslinguistic Perspective*. Berlin: Mouton de Gruyter.

Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 2001. *Foundations of Language. Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jones, Doug. 1999. Evolutionary psychology. *Annual Review of Anthropology* 28: 553–575.

Junker, Marie-Odile. This volume. Semantic primes and their grammar in a polysynthetic language: East Cree. In *Cross-Linguistic Semantics*, Cliff Goddard (ed.), 163–204. Amsterdam: John Benjamins.

Keesing, Roger M. 1994. Radical cultural difference: Anthropology's myth? In *Language Contact and Language Conflict*, Martin Pütz (ed.), 3–24. Amsterdam: John Benjamins.

Kempson, Ruth. 1977. *Semantic Theory*. Cambridge: Cambridge University Press.

Lappin, Shalom (ed.). 1996. *The Handbook of Contemporary Semantic Theory*. Oxford: Blackwell.

Lehrman, Alexander. 2006. Meaning as grammar. *Language Sciences* 28: 497–507.

Levinson, Stephen C. 2005. Comments on Daniel L. Everett, Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology* 46(4): 637–638.

Majid, Asifa, Enfield, N. J. and van Staden, Miriam (eds). 2006. Parts of the body: Cross-linguistic categorisation. Special Issue of *Language Sciences* 28(2/3): 137–360.

Malotki, Ekkehart. 1983. *Hopi Time: A Linguistic Analysis of the Temporal Concepts in the Hopi Language*. Berlin: Mouton.

McCawley, James D. 1972. A program for logic. In *Semantics of Natural Language*, D. Davidson and G. Hartman (eds), 489–544. Dordrecht: Reidel.

Meira, Sérgio. 2006. Tiriyó body part terms. *Language Sciences* [Special Issue on "Parts of the body: Cross-linguistic categorisation", edited by Asifa Majid, N. J. Enfield and Miriam van Staden] 28(2/3): 262–279.

Mel'čuk, Igor A. 1989. Semantic primitives from the viewpoint of the Meaning-Text Linguistic Theory. *Quaderni di Semantica* 10(1): 65–102.

Myhill, John. 1996. Is BAD a semantic primitive: Evidence from Biblical Hebrew. *Lexicology* 2(2): 99–126.

Peeters, Bert. (ed.). 2006. *Semantic Primes and Universal Grammar. Empirical Findings from the Romance Languages*. Amsterdam: John Benjamins.

Priestley, Carol. This volume. The semantics of "inalienable possession" in Koromu (PNG). In *Cross-Linguistic Semantics*, Cliff Goddard (ed.), 277–299. Amsterdam: John Benjamins.

Roberts, Michael. 2005. Theoretical Issues in Natural Semantic Metalanguage. BA Hons subthesis, Armidale: The University of New England.

Shi-xu. 2000. To feel or not to feel, that is the question. *Culture & Psychology* 6(3): 375–383.

Shweder, Richard A. 2004. Deconstructing the emotions for the sake of comparative research. In *Feelings and Emotions: The Amsterdam Symposium*, Antony S. R. Manstead, Nico Fridja and Agneta Fischer (eds), 81–97. Cambridge: Cambridge University Press.

Tarski, Alfred. 1956[1935]. The concept of truth in formalized languages. In *Logic, Semantics, Metamathematics:Papers from 1923 to 1938*. Translated by J. H. Woodger. 152–278. Oxford: Clarendon Press.

Travis, Catherine A. 2003. The semantics of the Spanish subjunctive: Its use in the natural semantic metalanguage. *Cognitive Linguistics* 14(1): 47–69.

van Brakel, Jaap. 2002. Chromatic language games and their congeners. In *Theories, Technologies, Instrumentalities of Color*, Barbara Saunders and Jaap van Brakel (eds), 147–168. Lanham: University Press of America.

van Driem, George. 2004. Review: Cliff Goddard and Anna Wierzbicka (eds) *Meaning and Universal Grammar – Theory and Empirical Findings*. *Language* 80(1): 163–165.

van Staden, Miriam. 2006. The body and its parts in Tidore, a Papuan language of Eastern Indonesia. *Language Sciences* [Special Issue on "Parts of the body: Cross-linguistic categorisation", edited by Asifa Majid, N. J. Enfield and Miriam van Staden] 28(2/3): 323–343.

Whorf, Benjamin Lee. 1956. *Language, Thought, and Reality. Selected Writings of Benjamin Lee Whorf*. Edited and with an introduction by John B. Carrol. Cambridge, MA: MIT Press.

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

Wierzbicka, Anna. 1980a. *Lingua Mentalis. The Semantics of Natural Language.* Sydney/New York: Academic Press.

Wierzbicka, Anna. 1980b. *The Case for Surface Case.* Ann Arbor: Karoma.

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

Wierzbicka, Anna. 1992. *Semantics, Culture and Cognition: Universal Human Concepts in Culture-Specific Configurations*. New York: Oxford University Press.

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

Wierzbicka, Anna. 2002. Semantic primes and universal grammar in Polish. In *Meaning and Universal Grammar – Theory and Empirical Findings*, *Vol. II*, Cliff Goddard and Anna Wierzbicka (eds), 65–144. Amsterdam: John Benjamins.

Wierzbicka, Anna. 2003. Empirically established universals of language and their importance for cognitive science: The human conceptualisation of shape. Keynote paper at the *4th International Conference on Cognitive Science*, University of New South Wales, Sydney, July 2003.

Wierzbicka, Anna. 2003[1991]. *Cross-Cultural Pragmatics: The Semantics of Human Interaction.* 2nd edition. Berlin: Mouton de Gruyter.

Wierzbicka, Anna. 2004. The semantics of colour: A new paradigm. Keynote paper at the *4th International Conference Progress in Colour Studies* (PICS04), Glascow University, June 29–July 3, 2004.

Wierzbicka, Anna. 2005. Comments on Daniel L. Everett, Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology* 46(4): 641.

Wierzbicka, Anna. 2006a. *English: Meaning and Culture*. New York: Oxford University Press.

Wierzbicka, Anna. 2006b. There are no "color" universals, but there are universals of visual semantics. *Anthropological Linguistics* 47(2): 217–244.

Wierzbicka, Anna. 2006c. Shape in grammar revisited. *Studies in Language* 30(1): 115–177.

Wierzbicka, Anna. 2007a. Bodies and their parts: An NSM approach to semantic typology. *Language Sciences* 29(1): 14–65.

Wierzbicka, Anna. 2007b. The semantics of colour: A new paradigm. In *Progress in Colour Studies, Vol. 1. Language and Culture*, Carole P. Biggam and Christian J. Kay (eds), 1–24. Amsterdam: John Benjamins.

Wierzbicka, Anna. In press. The theory of the mental lexicon. *HSK (Handbücher zür Sprach- und Kommunikationswissenschaft. Handbook of Linguistics and Communication Science*) *Slavic Languages*. Berlin/New York: Walter de Gruyter.

Wong, Jock Onn, Cliff Goddard and Anna Wierzbicka. To appear. *Walk*, *run*, *jump*: The semantics of human locomotion.

Yoon, Kyung-Joo. 2006. *Constructing a Korean Natural Semantic Metalanguage*. Seoul: Hankook Publishing Company.

## Appendices

**Appendix 1.** Semantic primes: English

| | |
|---|---|
| Substantives: | I, YOU, SOMEONE, SOMETHING/THING, PEOPLE, BODY |
| Relational substantives: | KIND, PART |
| Determiners: | THIS, THE SAME, OTHER/ELSE |
| Quantifiers: | ONE, TWO, MUCH/MANY, SOME, ALL |
| Evaluators: | GOOD, BAD |
| Descriptors: | BIG, SMALL |
| Mental predicates: | THINK, KNOW, WANT, FEEL, SEE, HEAR |
| Speech: | SAY, WORDS, TRUE |
| Actions, events, movement, contact: | DO, HAPPEN, MOVE, TOUCH |
| Location, existence, possession, specification: | BE (SOMEWHERE), THERE IS, HAVE, BE (SOMEONE/SOMETHING) |
| Life and death: | LIVE, DIE |
| Time: | WHEN/TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME, MOMENT |
| Space: | WHERE/PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE |
| Logical concepts: | NOT, MAYBE, CAN, BECAUSE, IF |
| Intensifier, augmentor: | VERY, MORE |
| Similarity: | LIKE |

**Notes**: • Primes exist as the meanings of lexical units (not at the level of lexemes) • Exponents of primes may be words, bound morphemes, or phrasemes • They can be formally complex • They can have different morphosyntactic properties, including word-class, in different languages • They can have combinatorial variants (allolexes) • Each prime has well-specified syntactic (combinatorial) properties.

**Appendix 2.** Semantic primes: Japanese and Spanish

| | |
|---|---|
| WATASHI *I*, ANATA *you*, DAREKA *someone*, NANIKA/MONO/KOTO *something/thing*, HITO/HITOBITO *people*, KARADA *body* | YO *I*, TU *you*, ALGUIEN *someone*, ALGO/COSA *something/thing*, GENTE *people*, CUERPO *body* |
| SHURUI *kind*, BUBUN *part* | TIPO *kind*, PARTE *part* |
| KORE *this*, ONAJI *the same*, HOKA *other* | ESTO *this*, LO MISMO *the same*, OTRO *other* |
| HITO-/ICHI- *one*, FUTA-/NI- *two*, TAKUSAN *many/much*, IKUTSUKA *some*, MINNA *all* | UNO *one*, DOS *two*, MUCHO *much/many*, ALGUNOS *some*, TODO *all* |
| II *good*, WARUI *bad* | BUENO *good*, MALO *bad* |
| OOKII *big*, CHIISAI *small* | GRANDE *big*, PEQUEÑO *small* |
| OMOU *think*, SHIRU *know*, HOSHII/ -TAI/NOZOMU *want*, KANJIRU *feel*, MIRU *see*, KIKU *hear* | PENSAR *think*, SABER *know*, QUERER *want*, SENTIR *feel*, VER *see*, OÍR *hear* |
| IU *say*, KOTOBA *words*, HONTOO *true* | DECIR *say*, PALABRAS *words*, VERDAD *true* |
| SURU *do*, OKORU/OKIRU *happen*, UGOKU *move*, FURERU *touch* | HACER *do*, PASAR *happen*, MOVERSE *move*, TOCAR *touch* |
| (DOKOKA) NI IRU/ARU *be (somewhere)*, IRU/ARU *there is*, MOTSU *have*, (DAREKA/NANIKA) DE ARU *be (someone/ something)* | ESTAR *be (somewhere)*, HAY *there is*, TENER *have*, SER *be (someone/something)* |
| IKIRU *live*, SHINU *die* | VIVIR *live*, MORIR *die* |
| ITSU/TOKI *when/time*, IMA *now*, MAE *before*, ATO *after*, NAGAI AIDA *a long time*, MIJIKAI AIDA *a short time*, SHIBARAKU NO AIDA *for some time*, SUGUNI *moment/in one moment* | CUÁNDO/TIEMPO *when/time*, AHORA *now*, ANTES *before*, DESPUÉS *after*, MUCHO TIEMPO *a long time*, POCO TIEMPO *a short time*, POR UN TIEMPO *for some time*, MOMENTO *moment* |
| DOKO/TOKORO *where/place*, KOKO *here*, UE *above*, SHITA *below*, CHIKAI *near*, TOOI *far*, MEN *side*, NAKA *inside* | DÓNDE/SITIO *where/place*, AQUÍ *here*, ARRIBA *above*, DEBAJO *below*, CERCA *far*, LEJOS *near*, LADO *side*, DENTRO *inside* |
| -NAI *not*, TABUN *maybe*, DEKIRU *can*, -KARA *because*, MOSHI (BA) *if* | NO *not*, TAL VEZ *maybe*, PODER *can*, PORQUE *because*, SI *if* |
| SUGOKU *very*, MOTTO *more* | MUY *very*, MÁS *more* |
| YOO/DOO/YOO NI *like/how/as* | COMO *like* |