

**Semantic primes, semantic molecules, semantic templates:**

**Key concepts in the NSM approach to lexical typology<sup>1</sup>**

*Cliff Goddard, Griffith University*

The Natural Semantic Metalanguage (NSM) approach has a long track record in cross-linguistic lexical semantics (Wierzbicka 1992, 1996, 1999; Goddard 1998, 2005, 2006, 2008; Harkins and Wierzbicka 2001; Goddard and Wierzbicka 2002; Peeters 2006; Gladkova 2010; Ye 2007a, 2007b, 2010; Bromhead 2009, 2011; Wong 2005, 2010; and other works). It is therefore not surprising that it has a clear theoretical position on key issues in lexical semantic typology and a well-developed set of analytical techniques. From a theoretical point of view, the overriding issue concerns the *tertium comparationis*. What are the optimal concepts and categories to support the systematic investigation of lexicons and lexicological phenomena across the world's languages? To this question, the NSM approach offers the following answer: the necessary concepts can—and must—be based on the shared lexical-conceptual core of all languages, which NSM researchers claim to have discovered over the course of a thirty-five year program of empirical cross-linguistic semantics. This shared lexical-conceptual core is the mini-language of semantic primes and their associated grammar. In addition, over the past 10 or so years, NSM researchers have developed certain original analytical constructs which promise to enhance the power and systematicity of the approach: in particular, the notions of semantic molecules and semantic templates. This paper sets out to explain and illustrate these notions, to report some key analytical findings (updated, in many cases, from previously published accounts), and to extrapolate their implications for the further development of lexical typology.

**1. General principles and approach**

The NSM approach differs from most other work in cross-linguistic semantics in two fundamental ways. First and foremost, NSM semantics is based on reductive paraphrase, in a very strict and literal sense. An NSM explication of a sentence or sentence frame is a systematic reductive paraphrase, i.e. an attempt to “say the same thing” in a paraphrase composed of maximally simple, intelligible and translatable words (semantic primes), thereby laying bare the semantic content of the original sentence or sentence frame. NSM researchers do not attempt in the first instance to classify lexical meanings, but rather to paraphrase them without circularity. Classifications may emerge inductively, generalizations of other kinds

may emerge – but the first process is always paraphrase. A corollary to the reductive paraphrase technique is that no specialist or technical terms are allowed in formal NSM semantic explications, because to do so inevitably leads to unacceptable abstractness and obscurity and/or to circularity.

When semantic description is carried out in accordance with these principles, it can be viewed both as linguistic analysis and as conceptual analysis. In other words, the NSM claim is that a successful reductive paraphrase which predicts and/or explains natural usage (including distribution, collocation, entailments, implications, and so on) and which satisfies native speaker intuitions can be viewed as a conceptual model. Because it is carried out in terms that are known to speakers and that form part of their everyday linguistic competence, a paraphrase analysis can have a *prima facie* claim to conceptual authenticity, in the sense of representing what anthropologists call an “insider perspective”. At the same time, the constraint that reductive paraphrase be carried out in the language concerned (or, equivalently, in words which have semantic counterparts in the language concerned) safeguards against terminological Anglocentrism, i.e. the imposition of Anglo conceptual categories onto the concepts of other languages.

The most fundamental NSM concept is the concept of semantic primes, i.e. meanings which cannot be paraphrased in simpler terms: the bedrock of linguistic meaning. To the extent that semantic primes can be identified and can be shown match up across languages, they provide a stable and language-neutral metalanguage for lexical typology, at least on its semantic side; for mapping out patterns of polysemy, patterns of structuring in the lexicon, the general architecture of semantic domains and fields, for investigating lexicon-grammar interactions, and so on (Lehrer 1992; Koch 2001; Koptjevskaja-Tamm 2008). Framing semantic analyses (explications) in semantic primes ensures that they are clear, translatable, and intuitively accessible, which of course make them more predictive and easier to test.

The current model of 64 primes is the result of an incremental program of empirical/analytical research that began with Wierzbicka (1972). Major benchmarks since then have included the edited volumes Goddard and Wierzbicka (1994), Goddard and Wierzbicka (2002), Peeters (2006), and Goddard (2008), along with numerous other publications.<sup>2</sup> The Table of semantic primes below (Table 1) is presented using English exponents (a form which represents a prime meaning in a given language is known as an ‘exponent’ of that prime). Tables of primes in Spanish, Russian, and Japanese are included in the Appendix to this article. Comparable tables have been drawn up for many languages,

including French, Polish, Danish, Chinese, Korean, Lao, Malay, Mbula/Mangaaba-Mbula, East Cree, Amharic, Arabic, and others.

Though valuable for a summary presentation, it must be acknowledged that a simple listing of exponents, as in Table 1, has its limitations. In particular, since exponents can be polysemous, as discussed below, merely listing a form is not necessarily adequate to indicate the intended primitive sense. A full account of each prime includes also a description of its basic combinatorial properties, however, which indicates the range of use and normally suffices to make it perfectly clear which sense of a polysemous exponent is the intended primitive sense.<sup>3</sup>

Table 1: Semantic primes (English exponents), grouped into related categories.

I, YOU, SOMEONE, SOMETHING~THING, PEOPLE, BODY	substantives
KIND, PART	relational substantives
THIS, THE SAME, OTHER~ELSE	determiners
ONE, TWO, MUCH~MANY, LITTLE~FEW, SOME, ALL	quantifiers
GOOD, BAD	evaluators
BIG, SMALL	descriptors
KNOW, THINK, WANT, FEEL, SEE, HEAR	mental predicates
SAY, WORDS, TRUE	speech
DO, HAPPEN, MOVE, TOUCH	actions, events, movement, contact
BE (SOMEWHERE), THERE IS, HAVE (SOMETHING), BE (SOMEONE/SOMETHING)	location, existence, possession, specification
LIVE, DIE	life and death
WHEN~TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME, MOMENT	time
WHERE~PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE	space
NOT, MAYBE, CAN, BECAUSE, IF	logical concepts
VERY, MORE	intensifier, augmentor
LIKE~AS~WAY	similarity

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 combinatorial variants or “allolexes” (indicated with ~) – Each prime has well-specified syntactic (combinatorial) properties.

There is no space here to review or justify this inventory in detail, as has been done extensively in the publications previously mentioned (some examples whose lexicalisation in particular languages has been disputed will be discussed in Section 2). It bears repeating, however, that to be a plausible candidate as an NSM semantic prime, a word (strictly speaking, word-meaning) must be indefinable, i.e. ultimately simple, in addition to being well attested in a wide range of languages. A word like ‘eat’, for example, would be a non-starter on both counts, since it is clearly not undecomposable (it involves ‘doing’, ‘mouth’, etc.) and it is known not to have equivalents in some languages (Wierzbicka 2009; cf. Newman 2009).

The same applies to many other impressionistically basic items of English vocabulary, such as ‘go’, ‘hot’, and ‘bird’ (cf. Goddard 2001, 2002).

The Natural Semantic Metalanguage consists not just of a lexicon, but also of a syntax. Semantic primes are hypothesised to have certain universal combinatorial properties, and the available evidence indicates that these properties also manifest themselves in all or most languages. Space precludes an adequate treatment here, so the reader is referred to Goddard and Wierzbicka (2002) and Goddard (2008). To give a very brief indication of the kinds of properties involved, it can be mentioned that they include: (a) basic combinatorics: e.g. that substantives can combine with specifiers – ‘this something~thing’, ‘someone else’, ‘one place’, ‘two parts’, ‘many kinds’; (b) basic and extended valencies of predicates and quantifiers, e.g. that DO has patient and instrument valencies such as ‘do something to something’ and ‘do something with something’, and that ONE allows a partitive option, in expressions such as ‘one of these things’; (c) the complement options of the mental primes, KNOW, THINK and WANT.

For illustrative purposes, the proposed universal valency and complement frames for three semantic primes (HAPPEN, DO, THINK) are displayed in Table 2 using English exponents.

Table 2. Valency frames for three semantic primes: HAPPEN, DO and THINK

something HAPPENS	[minimal frame]
something HAPPENS to someone/something	[undergoer frame]
<hr/>	
someone DOES something	[minimal frame]
someone DOES something to someone else	[patient <sub>1</sub> frame]
someone DOES something to something	[patient <sub>2</sub> frame]
someone DOES something with something	[instrument frame]
someone DOES something with someone	[comitative frame]
<hr/>	
someone THINKS about someone/something	[topic of cognition frame]
someone THINKS something (good/bad) about someone/something	[topic + complement frame]
someone THINKS like this: “ — ”	[quasi-quotational frame]
(at this time) someone THINKS that [ — — ] <sub>s</sub>	[propositional complement frame]
<hr/>	

Naturally, the entries in Table 2 include various English-specific morphosyntactic devices, most obviously, the prepositions *to*, *with*, and *about*. Importantly however, the claim that semantic equivalents of these frames are available in all languages does not entail that other languages realise the frames using analogous morphosyntactic devices. For example, the patient and instrument roles of DO can be indicated in other languages by case-marking,

postpositions, verb serialisation, or purely by word-order. Even if prepositions are used in a given language, there is no requirement or expectation that they will pattern in a fashion analogous to English; for example, there is no requirement that the instrument and comitative roles will be marked by the same formal means, as they are in English, i.e. by means of preposition *with*. An NSM based on any language L1 (say, English) can be transposed into an NSM based on any other language L2 (say, Spanish, Russian, Japanese) via a finite set of specifiable morphosyntactic rules. (In this connection, interested readers may consult the ‘English-Spanish NSM Translator’ devised by Francesco Zamblera, which automatically converts between English-based and Spanish-based NSMs using a javascript program. It is available at [www.vilnergoy.org/nsm](http://www.vilnergoy.org/nsm), along with the associated documentation.)<sup>4</sup>

## **2. The lexicalisation issue: semantic primes across languages**

Needless to say, the finding that the 64 semantic primes appear to be present as lexical meanings in all languages (the so-called Strong Lexicalisation Hypothesis<sup>5</sup>) is itself a substantial claim about lexical typology. In assessing counter-claims in the literature it is essential to bear in mind that polysemy frequently complicates the task of identifying exponents of primes and matching them up across languages. Often the range of use of exponents of the same prime do not coincide, because as well as the identical shared meaning, the words in question also have additional meanings which differ from language to language, i.e. there is a match-up between the meanings of lexical units but not between whole lexemes. Though much remains to be done, over the past 15 years NSM researchers have accumulated a lot of data about common patterns of polysemy involving exponents of semantic primes. A selection of widely attested patterns is summarized, with some inevitable over-simplification, in Table 3.

In NSM studies, language-specific evidence is always adduced to support claims for semantic primes which depend on a polysemy analysis. Of course, to establish polysemy requires a principled method of semantic analysis. The conventional wisdom (if one can call it that), according to which it is often difficult or impossible to separate polysemy from semantic generality, or to separate lexically encoded information from contextual inference, is really just a symptom of the lack of an adequate systematic method of semantic description (Goddard 2006). If one does not have a method of stating even a single meaning, it is hardly surprising that one can make no headway when faced with multiple meanings.<sup>6</sup>

Table 3. Selected common polysemies of exponents of semantic primes (data from studies in Goddard and Wierzbicka 1994, 2002; Peeters 2006; Goddard 2008; and Gladkova 2010).

Semantic prime	Additional meaning(s)	Language and relevant lexical item
DO	'make'	Amharic ( <i>adərrege</i> ), Ewe ( <i>wɔ</i> ), Italian ( <i>fare</i> ), Kalam ( <i>g-</i> ), Malay ( <i>buat</i> ), Mbula ( <i>-kam</i> ), Russian ( <i>delat'</i> ), Spanish ( <i>hacer</i> ), Swedish ( <i>göra</i> ), Yankunytjatjara ( <i>palyani</i> )
FEEL	'taste and/or smell'	Ewe ( <i>se le lāme</i> ), Italian ( <i>sentire</i> ), Kalam ( <i>nn</i> ), Malay ( <i>rasa</i> ), Russian ( <i>čuvstvovat'</i> ), Spanish ( <i>sentir</i> )
	'hear'	Amharic ( <i>tə-semma-</i> ), Italian ( <i>sentire</i> ), Kalam ( <i>nn</i> ), Spanish ( <i>sentir</i> )
	'feel by touch'	Acehnese ( <i>rasa</i> ), English ( <i>feel</i> ), Italian ( <i>sentire</i> ), Spanish ( <i>sentir</i> )
BEFORE	'first'	Kalam ( <i>nd</i> ), Kayardild ( <i>ngariija</i> ), Lao ( <i>kòon'</i> ), Mbula ( <i>mungu</i> ), Samoan ( <i>muamua</i> )
	'ahead of and/or in front of'	Kalam ( <i>nd</i> ), Kayardild ( <i>ngariija</i> ), Russian ( <i>do</i> ), Samoan ( <i>muamua</i> )
WORDS	'what is said and/or message'	Amharic ( <i>k'al</i> ), English ( <i>words</i> ), Malay ( <i>perkataan</i> ), Mbula ( <i>sua</i> ), Russian ( <i>slova</i> )
	'talk and/or language'	Amharic ( <i>k'al</i> ), Kayardild ( <i>kangka</i> ), Korean ( <i>mal</i> ), Mandarin ( <i>huà</i> ), Mbula ( <i>sua</i> )

As mentioned, the lexicalised status of some primes has been challenged for some languages, and a couple of these claims are not yet resolved. Before commenting on one or two of these, I would like to make the point that in too many cases, as it seems to me, apparently definitive claims that “Language Y has no word for semantic prime X” are advanced on the basis of unreliable or inadequate data and/or weak argumentation. We find an example of unreliable or inadequate data in a much-discussed article titled “The myth of language universals” (Evans and Levinson 2009). The authors declare (oddly enough, without reference to NSM work) that there are “vanishingly few” semantic universals, and they illustrate by saying that, for example, the Australian Aboriginal language Guugu-Yimidhirr has no word for ‘if’. To back this claim, they cite John Haviland’s (1979) short grammar of Guugu-Yimidhirr, now out of print and unavailable to most readers. In fact, however, Haviland (1979: 151-2) had stated: “The related particle *budhi* ‘if’ signals uncertainty, or questions the possibility of some outcome, sometimes very much like a subordinate conjunction, sometimes in a more modal sense”. Haviland cited the apparently conditional sentence *Nyundu budhu dhadaa nyundu minha maanaa bira* [2sgNOM if go.NONPAST 2sgNOM meat get.NONPAST indeed] with two glosses: ‘If you go, you’ll get meat for sure’ and ‘Should you go, you’ll get meat for sure’.<sup>7</sup>

As an example of weak argumentation, one can refer to Everett's (2005) much publicised description of Pirahã, a language spoken in the Brazilian rainforests. On Everett's account, Pirahã language and culture appears "extreme" in many ways, most notably in lacking various grammatical and semantic features that are widely viewed as universal. In relation to the semantic claims, I will discuss only the supposed lack of any expressions for the semantic prime ALL. The "closest expressions Pirahã can muster", according to Everett, are examples such as the following, where word 'ogi' 'big' (or a nominalised version 'ogíáagaó' 'bigness') appears to convey the meaning ALL.

- (1) *Hiaitíihí hi' ogi-áaga-ó pi-ókaobii.*  
 pirahã.people he big-be(permanence)-direction water-direction entered  
 'All the people went to swim/went swimming/are swimming/bathing, etc.'
- (2) *'igihí hi 'ogiáagaó 'oga hápií ...*  
 man he bigness field went  
 'The men all went to the field.'

Despite the free translations, Everett insists that even in these contexts, the Pirahã word 'ogi' means 'big' and not ALL, even though this requires him to assert that the true meaning of the expression *hiaitíihí hi 'ogi* in (1) above is "people's bigness", not 'all the people'. The sentence as a whole, he is saying, means something like 'People's bigness went to swim'. Unfazed by the bizarre, not to say incoherent, quality of this interpretation, Everett refuses to consider the possibility that 'ogi' is polysemous and can express two meanings, i.e. either BIG or ALL. According to him, because 'ogi' means 'big' in some contexts, it means 'big' in all contexts.

Consistent with the obvious counter-hypothesis that 'ogi' is polysemous between BIG and ALL, there is the fact that Pirahã plural pronouns are formed by adding 'ogi' to the singular (or unspecified) form. It is well known that forms such as 'I-all' for 'we', 'you-all' for 'you (pl)', and 'he-all' for 'they' are common in creoles, and there is evidence that Pirahã pronominal roots have been borrowed from a nearby Tupi-Guarani language. Everett maintains, however, that the Pirahã word that corresponds to English 'we' truly means something like "my bigness". When challenged by Wierzbicka (2005: 641) on the issue of polysemy, Everett simply declined to say anything on the subject,<sup>8</sup> asserting instead that "much of Pirahã is largely incommensurate with English" (Everett 2005: 624, Note 5).

To be sure, the lexicalisation status of some primes has sometimes been challenged on the basis of better data and argumentation. Bohnemeyer (1998a, 1998b) has argued, for example, that Yucatek Maya lacks words corresponding to AFTER and BEFORE, but he does not deny that relations of temporal sequence can be clearly conveyed in the language. His argument is that this effect is achieved via pragmatic inference based on the combination of aspectual operators (such as the “terminative” or “post-state” *ts’o’k*, roughly similar to English *finish* or *end*), the linear order of clauses, and resumptive topicalisation (cf. Bohnemeyer 1998a: 213–215). From an NSM point of view, however, the crucial thing is that if a subordinate clause marked with *ts’o’k* and depicting event A, is followed by a main clause depicting event B, then the meaning conveyed corresponds to English ‘After A, B’. The fact that the interpretation is different if *ts’o’k* appears instead in the main clause (where it will correspond roughly to ‘finish’) suggests that *ts’o’k* is polysemous (cf. Goddard 2001). Bohnemeyer (2003) remains unconvinced, insisting that *ts’o’k* has a uniform Yucatek-specific meaning, unstatable in ordinary English,<sup>9</sup> and that Yucatek Maya and English are radically incommensurate in their temporal semantics. To take a second example, in a careful study Junker (2008) reports an apparent lexical gap for the prime PART in East Cree. Her treatment leaves some open questions, however, because although body-part terminology is arguably the canonical lexical domain for “part-hood” relations, 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 express components such as ‘one part of someone’s body’ (for ‘head’) or ‘two parts of someone’s body’ (for ‘hands’). In the Yucatek Maya and East Cree cases, the debate has not been settled conclusively.

Other exchanges about the supposed non-lexicalisation of primes in particular languages, include: Khanina (2008) and Goddard and Wierzbicka (2010) on WANT; van Brakel (2002: 151) and Wierzbicka (2007b) on SEE; Shi-xu (2000) and Chappell (2002: 270–271) on FEEL in Chinese; Myhill (1996) and Durst (1999) on BAD in Biblical Hebrew; and Dixon and Aikhenvald (2002) and Goddard (2011) on WORDS. In all these cases, in my view, the lexicalised status of the primes in question has been sustained.

In short, despite the existence of a handful of unresolved cases, the balance of evidence clearly indicates that semantic primes are expressible by words, phrases or affixes in all or most of the world’s languages. Furthermore, even if future research should establish beyond doubt that some languages lack exponents of certain primes, this would not invalidate the current inventory of semantic primes for the majority of the world’s languages, nor would it necessarily invalidate NSM analyses based on the current metalanguage.<sup>10</sup>



After this outline of NSM assumptions, we can proceed to the main concern of the present study, namely, to explore the applications of the NSM approach to lexical semantic typology. The general advantages of the method are: first, that the metalanguage of semantic primes provides a vehicle for extremely fine-grained resolution of meaning; second, that it wards off implicit circularity and excessive abstractness; and third, that it safeguards against (or at least minimises) terminological ethnocentrism, thereby enabling the analyst to produce cognitively plausible semantic descriptions. In addition, however, recent research in the NSM program has developed two notions of special interest to lexical typology, namely, the notions of semantic molecules and semantic templates.

### 3. Semantic molecules

An extensive body of published work shows that lexical meanings in many domains (including emotion terms, speech-acts, value terms, and discourse particles) can be explicated directly into semantic primes. Informally speaking, these domains can be characterised as “abstract” (non-concrete) areas of the lexicon, but there are also some items of concrete vocabulary that yield to this approach. I will illustrate with two English nouns from different semantic domains: *hands* (body-parts) and *children* (social categories). These examples have not been chosen at random. They will be relevant to subsequent argumentation.

Explications [A] and [B] below are taken from Wierzbicka (2007a), with slight adjustments,<sup>11</sup> and Goddard and Wierzbicka (to appear), respectively. For present purposes, it is not necessary to argue for the details, but rather to draw out some general points about structure and nature of the representations. First, although the wording of the individual components may be relatively simple, an explication taken as a whole is a rather complex structure. It appears to be an empirical fact that many human concepts have this kind of intricate structure. It is also worth noting the range and diversity of semantic primes that typically occur in explications. Between them, explications [A] and [B] use nearly half the prime inventory – over 30 primes – drawn from all divisions of the prime lexicon.<sup>12</sup>

[A] *hands (someone’s hands)*

- a. two parts of someone’s body
- b. these parts are parts of two other parts of this someone’s body on two sides of the body
- c. these parts of someone’s body can move as this someone wants
- d. these parts of someone’s body have many parts
- e. these parts can move in many ways as this someone wants
- f. because people’s bodies have these two parts, people can do many things with many things as they want

[B] *children*

- a. people of one kind
- b. all people are people of this kind before they can be people not of this kind
- c. when someone is someone of this kind, this someone has lived for a short time, not a long time
- d. the bodies of people of this kind are small
- e. when people are like this, they can do some things, they can't do many other things
- f. because of this, if other people don't do some good things for them, bad things can happen to them

As mentioned, a successful explication must be consistent with the distribution of the expression being explicated (including its referential range, when appropriate), with its collocational possibilities and preferences, and with its entailments and implications when substituted into contexts of use. It should also satisfy native speaker intuitions. It is not possible here to justify explications [A] and [B] fully against these criteria, but an abbreviated discussion of certain points – in particular, certain objections raised by a reviewer – may be helpful. It may seem an obvious point, but it is apparently necessary to say that in assessing the viability of an explication, the explication must be taken as whole. For example, when assessing [A] against the referential range of the word *hands* in English, it is no counterargument to say that from components (a)–(d) alone, the explication could refer to feet, rather than to hands. The restriction to hands is clearly achieved by the subsequent components, which describe the intended referents as having multiple parts which can move ‘in many ways’ and as enabling people to ‘do many things with many things as they want’. Equally, it is no counterargument to the presence of the element ‘two’ in component (a) to point out that some people have only one hand; first, because the explication is for the word *hands* (not *hand*), and, second, because to say that someone has one hand clearly implies that one other hand is missing, i.e. unlike as with, say, head or nose.<sup>13</sup>

While it is true that words from many domains (especially “abstract” domains) can be explicated directly into semantic primes, and that the same applies to some non-abstract words (as just shown), NSM researchers have long recognized (Wierzbicka 1991; Goddard 1998: Ch. 6) that for most words in the concrete lexicon, it is not possible to produce plausible explications directly using semantic primes alone. Rather, such explications typically require a combination of semantic primes and complex lexical meanings known in NSM theory as semantic molecules. That is, semantic molecules are complex meanings which are decomposable into combinations of semantic primes but which function as units in the structure of other, more complex concepts. For example, explications for *sparrow*, *owl* and *eagle* include ‘bird’ as a semantic molecule; explications for *fork*, *spoon* and *plate* include

‘eat’; explications for *walk* and *run* include ‘feet’ and ‘ground’. The concept of semantic molecules is similar to that of intermediate-level concepts in the semantic practice of the Moscow School of Semantics (Apresjan 1992, 2000; Mel’čuk 1989, 2006), but with the important additional constraint that NSM semantic molecules must be meanings of lexical units in the language. It appears that most of the concrete lexicon – nominal, verbal, adjectival – relies on semantic molecules. The exploration of semantic molecules promises to contribute much to a general theory of vocabulary structure; in particular, new ways to represent semantic complexity and new ways to depict semantic dependencies and relationships. In this section, I will illustrate these contentions with some concrete examples, starting with simple examples from the domains of body-parts and social categories.

Wierzbicka (2007a) is an extensive study of body-part semantics, including over 40 explications. For the most part, she found that body-part explications require components of three different kinds: a partial characterisation of the shape of the body-part, its “location” on the body, and an indication of its function. The aspect of interest at this moment is the need for a shape specification, because shape descriptors (such as ‘long’, ‘round’, and ‘flat’) are not semantic primes and if they are required in explications, this amounts to recognising them as semantic molecules. For example, explication [C] for *legs* utilises the molecule ‘long’; explication [D] for *head* utilises the molecule ‘round’.<sup>14</sup> When semantic molecules appear in explications, they are marked as such by the notation [m].

[C] *legs (someone’s legs)*

- a. two parts of someone’s body
- b. these parts are long [m]
- c. these parts are below all the other parts of the body
- d. these 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

[D] *head (someone’s head)*

- a. one part of someone’s body
- b. this part is like something round [m]
- c. this part is above all the other parts of the body
- d. when someone thinks about something, something happens in this part of this someone’s body

How then can shape descriptors be analysed? Can we be certain that they can be used safely in body-part explications without incurring circularity? Wierzbicka (2006a) provides a general treatment of shape descriptors, including explications for English *long*, *round*, *flat*, and *straight* (among others), and an account of the considerable polysemy of each of these

words. The full details need not concern us. The key point can be drawn out from a single example, namely, the explication in [E] for ‘long’. (It is important to note that this explication applies only to the “shape descriptor” sense of *long*, i.e. as when one describes a kind of physical object such as a tail, a stick, or a cucumber as ‘something long’.<sup>15</sup>)

[E]     *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 is immediately obvious that one body-part – ‘hands’ – plays a crucial role in this explication, and indeed, in all Wierzbicka’s explications for shape descriptor concepts. This is because shape descriptors designate properties that are both visual and “tangible”, and to spell out the nature of the latter concept requires both semantic prime TOUCH (contact) and semantic molecule ‘hands [m]’. As established earlier, ‘hands’ itself can be explicated directly into semantic primes, so there is no circularity here. Rather, there is a chain or hierarchy of semantic dependency that can be represented as follows:

$$\{\text{‘legs’}, \text{‘arms’}, \text{‘head’}\} < \{\text{‘long’}, \text{‘round’}\} < \{\text{‘hands’}\} < \{\text{semantic primes}\}$$

This diagram is intended to indicate that each word set enclosed in curly brackets depends semantically on all the word sets to the right of it. This dispels any assumption that the impressionistically basic words of a particular semantic domain (in this case, ‘parts of the body’) are more or less the same in their degree of semantic complexity.

Let us work through a second example from another domain, that of social categories, such as *men*, *women*, *children*, *boys* and *girls* (Goddard and Wierzbicka to appear). We have already established that *children* can be explicated directly into semantic primes. Now consider explication [F], noting that in the final line the word ‘child [m]’ appears as a semantic molecule. Essentially, the proposal is that the concept of *women* depends on the idea that there are two kinds of people’s bodies, *women* being people of the kind whose body type allows them to have children. In other words, the concept of ‘women’ depends semantically on the concept of ‘child’.<sup>16</sup>

[F] *women*

- a. people of one kind
- b. someone can be someone of this kind after this someone has lived for some time, not for a short time
- c. there are two kinds of people's bodies, people of this kind have bodies of one of these two kinds
- d. some parts of bodies of this kind are not like parts of bodies of the other kind
- e. the bodies of people of this kind are like this:
  - at some times there can be inside the body of someone of this kind a living body of a child [m]

Taking the analysis a step further, Goddard and Wierzbicka (to appear) argue that the meaning of *men* incorporates 'women' as a semantic molecule. Subsequently, all three of these basic social categories, i.e. 'men', 'women' and 'children', are needed in the explications of numerous other words; for example, in the domain of kinship (Wierzbicka to appear). Some of these relationships can be depicted as follows:

{ 'father', 'mother', 'husband', 'wife' } < { 'men' } < { 'women' } < { 'children' } < { semantic primes }

Lest this conclusion seem unremarkable, it is worth reminding ourselves that in the old structuralist canon 'men' and 'women' were analysed as [+MALE, +ADULT] and [-MALE, +ADULT], respectively. This depicted these two words as symmetrical in semantic structure and made no reference to 'children' whatever.

It will be evident by now that many complex concepts have multiple "nestings" of molecule within molecule. In explications 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 nesting can occur 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).

As suggested by the last point, it is evident that semantic molecules can differ in their degree of productivity and in how widely they range across the lexicon. How many productive semantic molecules are there?, it may be asked. At the current early stage of research, the answer is not very clear. For English, the number is probably between 150 and 250. It is known that for English productive semantic molecules can come from at least the categories listed in Table 4 (the examples given are non-exhaustive).

Table 4: Selection of semantic molecules of English, grouped into categories (Goddard 2010)

parts of the body:	‘hands’, ‘mouth’, ‘eyes’, ‘ears’, ‘head’, ‘legs’, ‘feet’
physical descriptors:	‘long’, ‘round’, ‘flat’, ‘straight’, ‘hard’, ‘sharp’
physical activities:	‘sit’, ‘stand’, ‘lie’, ‘eat’, ‘drink’, ‘hold’
physical acts:	‘kill’, ‘pick up’, ‘bite’
expressive actions:	‘laugh’, ‘make sounds’, ‘sing’, ‘write’, ‘read’
topological:	‘top’, ‘bottom’, ‘front’, ‘edges’, ‘ends’, ‘hole’
life-forms:	‘creature’, ‘animal’, ‘bird’, ‘fish’, ‘tree’
environmental:	‘ground’, ‘sky’, ‘sun’, ‘water’, ‘fire’
times and places:	‘year’, ‘day’, ‘country’, ‘home’, ‘school’
materials:	‘wood’, ‘stone’, ‘metal’, ‘glass’, ‘paper’
mechanical:	‘wheel’, ‘pipe’, ‘wire’, ‘engine’, ‘electricity’, ‘machine’
social and family:	‘men’, ‘women’, ‘children’, ‘mother’, ‘father’
major cultural concepts:	‘money’, ‘book’, ‘colour’, ‘number’

To wind up this section, I will itemise some of the implications of the theory of semantic molecules for lexical typology. First, there may well be some universal or near-universal semantic molecules, particularly for 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 paid to language-specific polysemies. Other candidates are certain other body-parts such as ‘eyes’ and ‘ears’ (Wierzbicka 2007a), basic social categories like ‘men’, ‘women’ and ‘children’ (Goddard and Wierzbicka to appear), some basic kin concepts such as ‘mother’, ‘father’, ‘husband’ and ‘wife’ (Wierzbicka to appear), and some environmental molecules, such as ‘sky’, ‘ground’, ‘fire’ and ‘water’ (Goddard 2010).<sup>17</sup>

Second, it is equally 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), but the possibility that lower-level molecules such as shape descriptors and topological terms can also vary somewhat from language to language is more surprising. However, Wierzbicka (2006a) argues that English ‘long [m]’ does not exactly match the comparable Polish molecule ‘podłużny [m]’ ‘elongated, oblong’, and Brotherson (2008) argues that English ‘ends [m]’ differs subtly from its nearest counterpart ‘tapu [m]’ in Makasai (East Timor). The implications of these claims remain to be explored.

Major cultural concepts can also have profound implications for the vocabulary structure of particular languages, sometimes in non-obvious ways. For example, Wierzbicka (2006b, 2007b, 2008) argues that ‘colour [m]’ functions as a semantic molecule in English words like

*red, blue, green*, etc., but that many other languages lack “colour words” in the true sense, because their visual descriptors do not involve any comparable molecule. The semantic molecule ‘number’ also has huge significance in English (and in many other languages), both in helping to constitute the productive lexical domain of number words (Goddard 2009a), and, less obviously, in contributing to diverse other concepts connected with quantification and measurement (such as, for example, categories like ‘age’, ‘temperature’, ‘weight’, units of measurement, words for measuring devices, arithmetical concepts, etc.).

#### **4. Semantic templates: the example of physical activity verbs**

In NSM theory, 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). It has since been elaborated and applied to adjectives of emotion, shape, colour, and physical qualities (Wierzbicka 1999, 2006a, 2006b, 2007b; Goddard and Wierzbicka 2007). In recent years, the semantic template concept has been extended to verbs. This section describes these new developments. Section 5 illustrates their application to lexical typology.

NSM researchers have developed proposals for the template structure of several subclasses of physical activity verbs, including: (i) intransitive verbs of bodily locomotion, such as *walk* and *run* (Goddard, Wierzbicka and Wong, to appear); (ii) routine bodily activities, such as *eat* and *drink* (Wierzbicka 2009, 2010; Ye 2010); and (iii) complex physical activity verbs typically involving instruments, such as *cut* and *chop* (Goddard and Wierzbicka 2009).<sup>18</sup> At the present time, these templates have been worked out and tested only for English and a small selection of other languages (Polish, German, Japanese, Chinese, Kalam, Warlpiri), but NSM researchers believe that the same or similar template structures will be appropriate for many languages, hence their potential significance for lexical typology.

The proposed template structures for these three subclasses are very similar, as shown in Figure 1. In each case, the two top-most sections are termed, respectively, Lexico-Syntactic Frame and Prototypical Motivational Scenario, and the final section is termed Potential Outcome. The major differences lie in the penultimate section(s). For bodily locomotion and routine physical activities, the Manner section describes a coordinated set of body-part movements. For complex physical activities, the Instrument section describes an Instrument and how it is used, and the nature of the incremental effect that this exercises on the object (WHAT IS HAPPENING TO THE OBJECT).

Locomotion and routine physical activities, e.g. <i>walk, run, eat, drink</i>	Complex physical activities involving instruments, e.g. <i>cut, chop, grind</i>
LEXICO-SYNTACTIC FRAME	LEXICO-SYNTACTIC FRAME
PROTOTYPICAL MOTIVATIONAL SCENARIO	PROTOTYPICAL MOTIVATIONAL SCENARIO
MANNER	INSTRUMENT (incl. WHAT IS HAPPENING TO THE OBJECT)
POTENTIAL OUTCOME	POTENTIAL OUTCOME

Figure 1: Template structure for three subclasses of physical activity verbs.

Let us review these proposed template structures. Lexico-Syntactic Frame refers to the top-most section, with different macro-classes having different frames. Figure 2 displays Lexico-Syntactic Frames for three subclasses of physical activity verbs. The details in the frame determine the mapping from lexical semantics to morphosyntactic expression. The frames define core argument structure, inherent aspect, causal notions, and the controlled nature of the activities. Notice that no technical linguistic terms (‘agent’, ‘patient’, ‘duration’, ‘control’, or the like) are used in stating the frame.

Locomotion, e.g. <i>walk, run</i>	someone is doing something somewhere for some time because of this, this someone’s body is moving in this place at the same time as this someone wants
Routine physical activities, e.g. <i>eat, drink</i>	someone is doing something to something for some time because of this, something is happening to this something at the same time
Complex physical activities, e.g. <i>cut, chop</i>	someone is doing something to something for some time because of this, something is happening at the same time to this something as this someone wants this someone is doing it with something else

Figure 2: Lexico-Syntactic Frames for three subclasses of physical activity verbs.

A notable feature of the Lexico-Syntactic Frames displayed in Figure 2 is that they are phrased in the imperfective (note the durative component ‘for some time’). Many treatments in other frameworks assume without discussion that perfective uses (*walked, ran, cut, chopped, etc.*) are basic, but NSM analysts agree with the tradition in Russian lexicology that, for physical activity verbs, the imperfective forms and uses are semantically simpler. This is because their perfective counterparts involve extra semantic components, e.g. ‘at one time’, and the specification that an outcome has been achieved. Though we cannot go through the details here, the claim is that this analytical strategy enables a solution to the so-called imperfective paradox<sup>19</sup> and to the problem of how to specify the semantic relationships between constructional variants (alternations) of a single verb (Goddard and Wierzbicka 2009; Goddard to appear).



The next section is Prototypical Motivational Scenario. Running against the extensionalist methodology of some work in lexical typology, e.g. Majid and Bowerman (2007), NSM researchers maintain that speakers conceptualize human activities partly by reference to their prototypical motivations. For example, the proposed Prototypical Motivational Scenario associated with English *walk* states that a person often does something like this (i.e. *walks*) when they want to be somewhere after some time, not far from the place where they are. (This does not imply that people only ever *walk* with this motivation. Obviously, one can walk for exercise or for pleasure, or for other reasons.) Likewise, the prototypical scenario associated with *eat* and with *drink* is that someone wants something to be inside their body. Complex physical activity verbs (*cut*, *chop*, *grind*, *knead*, etc.) have a richer cognitive structure than locomotion and other routine activities, because they prototypically involve something like conscious intention: an actor forming a “preparatory thought” directed towards changing the current state of some object. For example, for English *cut*, the proposed Prototypical Motivational Scenario involves wanting something not to be one thing anymore, but instead to be two things, and as well, wanting to control the separation process with some precision. Examples of prototypical scenarios for representative verbs of three physical activity subclasses are given in Figure 3.

<i>walk</i> (locomotion)	at many times when someone does this, this someone does it because it is like this: – this someone is somewhere at some time – this someone wants to be somewhere else after some time – this other place is not far from the place where this someone is
<i>drink</i> (routine physical activity)	at many times when someone does this, this someone does it because it is like this: – this something is something like water [m] – this someone wants this something to be inside their body
<i>cut</i> (complex physical activity)	at many times when someone does this, this someone does it because it is like this: a short time before this someone thought like this about this something: “I don’t want this something to be one thing anymore, I want it to be two things because of this, I want to do something to this something for some time when I do this, I want something to happen to this something all the time as I want”

Figure 3: Prototypical Motivational Scenarios for sample verbs from three subclasses.

The next section of the template for physical activity verbs is Manner: how the activity is carried out. Given the rational, goal-directed nature of human action, it is not surprising that the details are closely linked to the Prototypical Motivational Scenario. People do things in a certain way in order to get the desired result. For English locomotion verbs like *walk* and *run*,

the Manner section includes a lot of detail about how the feet and legs move in relation to one another and in relation to the ground (in other languages with more general motion verbs, much less detail is included). For routine physical activities like *eat* and *drink*, the Manner section details how the parts of the mouth and (in some cases) the hands are used. For complex physical activities like *cut*, *chop*, and *grind*, the section that corresponds to “manner” is more complex: a description of an Instrument and how it used, which can involve several interrelated sub-events. Then comes a section describing how the object is incrementally affected by the action of the instrument (What is Happening to the Object). In the Manner section, there is commonly provision for an iterative structure of repeated episodes (‘the same thing happens many times; it happens like this ...’).

The final section is Potential Outcome, i.e. an indication of what the final result can be if the process continues long enough. This component is introduced as follows: ‘if it happens like this for some time, after this, ...’. From a logical (or teleological) point of view, the Potential Outcome component completes the explication by linking the projected process with the motivation of the prototypical actor.

## **5. Cross-linguistic comparison using semantic templates**

The proposed templates for physical verbs were not preconceived notions, but emerged as a consistent organisational format during the painstaking and iterative process of semantic analysis (drafting and re-drafting multiple explications, testing them against range of use in natural discourse and against native speaker intuitions, checking the coherence and well-formedness of the metalanguage, and so on). They seem to have a natural “internal logic” by which the causal and temporal interconnections between the various components can be ordered in a coherent fashion. On this account it seems likely that these templates will be similar across many languages. Albeit that the sample of languages is very small, this supposition seems to be borne out across the several non-English languages on which detailed work of this nature has already been done: Japanese, Polish, Kalam, and Warlpiri. Space permits only two cross-linguistic examples here: English *drink* compared with Kalam *ñb* ‘eat/drink’, and English *cut* compared with Japanese *kiru* “cut”. The exposition will be abbreviated. Further justification and comparisons with other related verbs are given in Goddard and Wierzbicka (2009) and Wierzbicka (2009, 2010). (Note that the Prototypical Motivational Scenario sections in the following four explications differ slightly from the previously published versions.)

*English drink vs. Kalam ñb ‘eat/drink’.* Explication [G] below is for English *drink*. In terms of its overall structure, most of the relevant details have been introduced already. A couple of notable points are as follows: (i) the Prototypical Motivational Scenario includes a characterisation of the object as ‘something like water [m]’, i.e. a liquid;<sup>20</sup> (ii) the Manner section depicts an iterative structure; more specifically, it involves doing something with the mouth that causes some of the “water-like” substance to be inside the mouth for a very short time, following which a further action of the mouth causes it to be somewhere else inside the person’s body.

[G] *Someone is drinking something:*

- |    |   |                                       |
|----|---|---------------------------------------|
| a. | someone is doing something to something for some time<br>because of this, something is happening to this something at the same time   | LEXICO-SYNTACTIC FRAME                |
| b. | at many times when someone does this to something, this someone does it<br>because it is like this:<br>– this something is something like water [m]<br>– this someone wants this something to be inside their body  | PROTOTYPICAL<br>MOTIVATIONAL SCENARIO |
| c. | when someone does this to something, the same thing happens many times<br>it happens like this:<br>– this someone does something to this something with their mouth [m]<br>– because of this, after this, part of this something is for a very short time inside this someone’s mouth [m]<br>– after this, this someone does something else to it with their mouth [m]<br>– because of this, after this, this part of this something is not inside this someone’s mouth [m] anymore,<br>it is somewhere else inside this someone’s body for some time | MANNER                                |
| d. | if it happens like this for some time, after this, all (parts) of this something can be<br>somewhere inside this someone’s body for some time   | POTENTIAL<br>OUTCOME                  |

It may be instructive to respond to several objections to explication [G] that were raised by a reviewer of this paper. As mentioned earlier, when assessing an explication it is essential to evaluate the explication as a whole. For example, it would be misguided to object to explication [G] on the grounds that many other activities aside from *drinking*, such as *eating* or *kissing*, for example, involve ‘doing something to something with the mouth’. Equally misguided would be to reject the explication on the grounds that it allegedly applies to highly atypical situations such as “rapidly eating spoonfuls of crushed ice”. Section (b) of the explication characterises *drinking* in terms of what someone does ‘at many times’ when someone wants ‘something like water’ to be inside their body: this hardly matches the crushed ice situation. Likewise, it would be otiose to object that the expression ‘something like water’ could extend to glass or to ice. First, it seems doubtful that ordinary people think of glass or ice as ‘something like water’; second, there is no typical scenario in which

someone wants ice or glass to be inside their body; third, the actions described in the Manner section would not work as a way of introducing ice or glass into one's body.

If we were to compare explication [G] with that for English *eat*, we would see a slightly different Prototypical Motivational Scenario (involving ‘something not like water [m]’) and, consequently, a more elaborate Manner section, with more detail about actions of parts of the mouth (related to chewing) and about how these actions affect the substance in the mouth. As well, the Manner section for *eat* involves some preliminary action with the hands related to bringing the food item to the mouth. (For a detailed study of “eat” and “drink” verbs in Mandarin and Shanghainese, see Ye (2010).)

The Papuan language Kalam (Pawley and Bulmer in press) has no words equivalent in meaning to the English *eat* and *drink*.<sup>21</sup> Instead, both activities (as it seems from an English point of view) are designated by the verb *ñb*-, roughly, ‘consume’. According to Pawley and Bulmer’s dictionary, *ñb* is general in its semantics, rather than ambiguous; a sentence like *Tap etp nbsay*? ‘What are they eating/drinking?’ is genuinely vague.

Explication [H] shows how such an undifferentiated “eat/drink” meaning can be constructed (Wierzbicka 2009, 2010). It follows the same semantic template as for English *drink*, and many of the details also remain the same, while others differ. Notably (i) the Prototypical Motivational Scenario does not characterise the substance as either ‘something like water [m]’ (as with *drink*) or as ‘something not like water [m]’ (as with *eat*); and (ii) the period of time for which each mouthful of the substance remains in the mouth is described as ‘a short time’ (rather than ‘a very short time’, as with *drink*). Naturally, the elaborated manner details for *eat* are not appropriate.

[H]     *Someone X is ñb-ing something Y* [Kalam *ñb*]:

- |    |  |                                       |
|----|--|---------------------------------------|
| a. | someone is doing something to something for some time<br>because of this, something is happening to this something at the same time  | LEXICO-SYNTACTIC FRAME                |
| b. | at many times when someone does this to something, this someone does it<br>because it is like this:<br>this someone wants this something to be inside their body   | PROTOTYPICAL<br>MOTIVATIONAL SCENARIO |
| c. | when someone does this to something, the same thing happens many times<br>it happens like this:<br>– this someone does something to this something with the mouth [m]<br>– because of this, after this, part of this something is for a short time inside this someone’s mouth [m]<br>– after this, this someone does something else to it with their mouth [m]<br>– because of this, after this, this part of this something is not inside this someone’s mouth [m] anymore,<br>it is somewhere else inside this someone’s body for some time | MANNER                                |

- |  |                      |
|--|----------------------|
| d. if it happens like this for some time, after this, all (parts) of this something can be<br>somewhere inside this someone's body for some time | POTENTIAL<br>OUTCOME |
|--|----------------------|

*English cut vs. Japanese kiru.* Explication [I] is a full explication for English *cut*, following the semantic template outlined in the previous section for complex physical activity verbs (cf. Goddard and Wierzbicka 2009). In the Lexico-Syntactic Frame and Prototypical Motivational Scenario sections, we see that there is an emphasis on the “controlled” nature of the intended separation effect on the object. How this is achieved is spelt out in the Instrument section. This specifies that ‘some parts of this something are sharp [m]’ and then describes how the actor’s hand holds and guides the instrument so that it moves ‘as this someone wants’. The next section states that this results in the sharp edge enacting a controlled effect on the object at the point of contact leading to a permanent change in the object.<sup>22</sup>

[I] *Someone is cutting something (e.g. some paper, a cake) with something else:*

- |  |   |
|--|---|
| a. someone is doing something to something for some time<br>because of this, something is happening at the same time to this thing as this someone wants<br>this someone is doing it with something else   | LEXICO-SYNTACTIC FRAME                                |
| b. at many times when someone does this to something, this someone does it<br>because it is like this:<br>– a short time before, this someone thought like this about it:<br>“I don’t want this something to be one thing anymore, I want it to be two things<br>because of this, I want to do something to this something for some time<br>when I do this, I want something to happen to this something all the time as I want”   | PROTOTYPICAL<br>MOTIVATIONAL SCENARIO                 |
| c. when someone does this to something, it happens like this:<br>– this someone holds [m] part of something else with one hand [m] all the time<br>– some parts of this other something are sharp [m]<br>– this someone’s hand [m] moves for some time as this someone wants<br>– because of this, the sharp [m] parts of this other thing touch this thing in some places<br>as this someone wants<br>– because of this, something happens to this thing in these places as this someone wants<br>– because of this, after this, part of this thing is not like it was before | INSTRUMENT (incl. WHAT IS HAPPENING<br>TO THE OBJECT) |
| d. if it happens like this for some time, after this, this something can be two things   | POTENTIAL OUTCOME                                     |

The comparable scenario for English *chop* has a somewhat different prototypical scenario (it involves wanting ‘something hard [m]’ not to be one thing anymore, but instead to be many small things), with associated differences in the instrument-related sections (roughly, the sharp-edged instrument having a long handle, and being repeatedly raised above the object and brought down on it).

For an example of a similar-yet-different word in another language, we can turn to Japanese. The closest Japanese counterpart of *cut* is the verb *kiru*. Like *cut*, *kiru* usually refers to an activity performed with either a knife or scissors, and as in the case of *cut*, the prototypical intention appears to consist in dividing an object into two things in a controlled fashion. What is different is that in some situations Japanese *kiru* can refer to an action performed with one's fingers rather than with an instrument with a sharp edge. This applies in particular to paper, or objects made from paper, such as a sachet of powdered soup or sugar. Opening such a sachet with one's fingers can be described as a case of *kiru*, as in a sentence like: *Suupu-no fukuro-o te-de kitta* [soup-GEN sachet-ACC hand-INS *kiru*:PRT] '(He/she) opened the soup sachet with [his] hands.' Japanese dictionaries, e.g. Morita (1989), specify 'hand or cutlery' as the "instruments" associated with the verb *kiru*. When one folds a piece of paper, and then separates it along the crease with one's hand, the natural way to describe this in Japanese is with *kiru* – and, of course, this particular procedure is very commonly done in Japan on account of the widespread practice of origami, which involves the folding of paper and separating it into pieces by hand and with precision (Buisson 1992; Honda 1965).

Explication [J] below gives a unified interpretation of *kiru*. Despite the seemingly radical difference to *cut*, the explication follows the same template. The Instrument section presents the instrument of *kiru* as simply 'part of something else', without specifying whether or not the part belongs to an instrument or is part of the agent's body. This unspecified part can therefore stand equally well for the blade of a knife or for a person's fingers. Needless to say, this means there can be no mention of any sharp parts or of the agent holding anything. It can be stated, however, that as the agent's hand moves, correspondingly 'this part' moves at the same time 'as this someone wants'.

[J] *Someone-ga thing-o (thing-de) ki-tte-i-t* [Japanese *kiru*]:

- |    |  |   |
|----|--|---|
| a. | someone is doing something to something for some time<br>because of this, something is happening at the same time to this something as this someone wants<br>this someone is doing it with something else  | LEXICO-SYNTACTIC FRAME                                |
| b. | at many times when someone does this to something, this someone does it<br>because it is like this:<br>– this something is not something very hard [m]<br>– a short time before, this someone thought like this about this something:<br>"I don't want this something to be one thing anymore, I want it to be two things<br>because of this, I want to do something to this something for some time<br>when I do this, I want something to happen to this something all the time as I want" | PROTOTYPICAL<br>MOTIVATIONAL SCENARIO                 |
| c. | when someone does this to something, it happens like this:<br>– this someone does something for some time with part of something else  | INSTRUMENT (incl. WHAT IS HAPPENING<br>TO THE OBJECT) |

- during this time one of this someone's hands [m] moves as this someone wants
  - because of this, this part moves at the same time as this someone wants
  - because of this, this part touches this thing in some places as this someone wants
  - because of this, something happens to this other thing in these places as this someone wants
  - because of this, after this, thing is not like it was before
- d. if it happens like this for some time, after this, this something can be two things      POTENTIAL OUTCOME

Again, the four-part template for complex physical activity verbs provides a guiding framework, despite the considerable differences in the content of individual sections.

Since we are focused in this section on physical activity verbs, it seems appropriate to draw out two specific conclusions pertaining to physical activity verbs, before turning to some more general reflections about the NSM contribution to the project of lexical typology. The first conclusion is that, unlike much of the nominal lexicon, the verbal lexicon generally has a non-hierarchical structure.<sup>23</sup> Although informally it may be acceptable to speak of a set of verbs like *chop*, *slice*, and *mince* as “verbs of cutting”, it is not true that *chop*, *slice*, and *mince* are true semantic hyponyms of ‘cut’, i.e. *chopping* is not a ‘kind of cutting’, *slicing* is not a ‘kind of cutting’, and so on. This follows because semantic analysis shows that *chopping*, *slicing*, *mincing*, and so on, each have their own distinctive prototypical motivations and (related) manner specifications, and that these are not simply elaborated versions of the prototypical motivation and manner components for ‘cutting’ (Goddard and Wierzbicka 2009).

Second, the most promising avenue for exploring the semantic typology of human activity verbs is to focus on the Prototypical Motivational Scenario. This is because most of the other features of explications for verbs of this kind have their rationale in the nature of the prototypical motivation. For example, we can investigate whether all, or most, languages have a verb including in the following component: ‘I want this something not to be one thing anymore’. If a verb with this component is found, we can ask if it also includes one of the following two components: (i) ‘I want it (= this something) to be two things’ (as with *cut*), (ii) ‘I want it to be many small things’ (as with *chop*). Concurrently, we can ask whether the prototypical ‘something’ involved in such scenarios is characterized in any way, and if so, in what way, e.g. as ‘hard [m]’ or ‘very hard [m]’, and whether the instrument is characterized in any way, e.g. as ‘heavy [m]’. Proceeding in this way, it should be possible to build a lexico-semantic typology of verbs of physical activity based on universal semantic primes (such as WANT, SOMETHING, PART, ONE, TWO, MANY, SMALL, and so on) and on universal or widely attested semantic molecules, such as ‘sharp [m]’, ‘hard [m]’, and ‘heavy [m]’.

## 6. Concluding remarks

In a recent overview, Koptjevskaja-Tamm (2008) identifies three major fronts for research in lexical typology: (i) patterns of structuring in the lexicon, (ii) the architecture of lexical fields (semantic domains), and (iii) lexicon-grammar interactions. I believe that the recently developed NSM analytical concepts of semantic molecules and semantic templates have something to offer on each of these fronts.

In relation to patterns of structuring in the lexicon, the concept of semantic molecules leads to new ways of understanding lexical semantic complexity. Many semantic structures appear to have a kind of “gangly and lumpy” quality – lengthy strings of simple semantic primes interspersed with semantically dense molecules. Semantic molecules enable an incredible compression of semantic complexity, but at the same time this complexity is disguised by its being encapsulated and telescoped into lexical units embedded one in the other, like a set of Russian dolls. As far as I can see, these qualities are unlike anything envisaged in structuralist, cognitivist or generativist approaches to lexical semantics.

Regarding the architecture of lexical fields, the relevance of the notions of semantic templates should be obvious. In establishing the semantic template for words of a particular semantic class in any language, one creates a framework for cross-linguistic comparison. On present indications it seems likely that many template structures will be the same or similar across languages. When comparing related words in different languages, it becomes a simpler and more tractable task to seek out the locus of differences between languages. To add an example to that of physical activity verbs, one can mention Goddard and Wierzbicka’s (2007) study of physical quality adjectives, i.e. words like *sweet, hot, hard, heavy, rough, sharp*, in cross-linguistic perspective. The authors propose a semantic template that seems to work not only for English, but also for several other languages. Needless to say, whether and to what extent it would work for a much broader range of languages remains to be explored; but the point is that the groundwork has been done to make such a project possible. Presumably some semantic templates are widely attested across the languages of the world, while others are found only in languages of particular families, linguistic areas, or grammatical types.

The idea that many semantic templates begin with a Lexico-Syntactic Frame relevant to macro-grammatical properties is obviously highly pertinent to lexicon-grammar interaction. We have seen examples of Lexico-Syntactic Frames for three subclasses of English verbs based on semantic prime DO, in combination with other components; of course there are many others, corresponding to other subclasses of action and activity verbs. Classes and subclasses



of verbs from other divisions of the lexicon have their own distinctive Lexico-Syntactic Frames. For example, verbs and adjectives of emotion and sensation are based on FEEL, using components like (for emotions) ‘someone felt something (good/bad)’ or (for sensations) ‘someone felt something (good/bad) in part of his/her body’. Speech-act verbs are based on SAY, using components like ‘someone said something (good/bad/like this) to someone’. Yet other verbs are based on HAPPEN, using components like ‘something happened (to something)’ or ‘something happened somewhere’, in combination with other elements. Although much remains to be done, the concept of Lexico-Syntactic Frame would seem to offer a promising tool both for segmenting the verbal lexicon of individual languages and for comparing the organization of verbal lexicons across languages.

Overall, one can only agree with Koptjevskaja-Tamm (2008: 43) that the most urgent problems of lexical typology are methodological, including the need to “improve standards in cross-linguistic identification of studied phenomena and their (semantic) analysis”, and the need to “achieve a reasonable consensus on the meta-language used for semantic explications and on ways of representing meanings”. In my view, there is sufficient evidence in existing work to establish that the NSM metalanguage is an effective tool for analyzing and representing meanings across languages. From a practical point of view, wider use of the NSM vocabulary of semantic description will help reduce the proliferation of hard-to-reconcile individual terminologies.<sup>24</sup> Even if lexical typologists of other persuasions do not want to sign up to all the theoretical commitments of the NSM approach, I would urge them to consider the advantages of adding the reductive paraphrase methodology to their descriptive toolkits.

## Appendix: Exponents of semantic primes in three additional languages

Japanese	Spanish	Russian
WATASHI <i>I</i> , ANATA <i>you</i> , DAREKA <i>someone</i> , NANIKA~MONO~KOTO <i>something/thing</i> , HITO~HITOBITO <i>people</i> , KARADA <i>body</i>	YO <i>I</i> , TU <i>you</i> , ALGUIEN <i>someone</i> , ALGO~COSÁ <i>something~thing</i> , GENTE <i>people</i> , CUERPO <i>body</i>	JA <i>I</i> , TY <i>you</i> , KTO-TO <i>someone</i> , ČTO-TO~VEŠČ' <i>something~thing</i> , LJUDI <i>people</i> , TELO <i>body</i>
SHURUI <i>kind</i> , BUBUN <i>part</i>	TIPO <i>kind</i> , PARTE <i>part</i>	ROD~VID <i>kind</i> , ČAST' <i>part</i>
KORE <i>this</i> , ONAJI <i>the same</i> , HOKA <i>other</i>	ESTO <i>this</i> , LO MISMO <i>the same</i> , OTRO <i>other</i>	ÉTOT <i>this</i> , TOT ŽE <i>the same</i> , DRUGOJ <i>other</i>
HITO~ICHI- <i>one</i> , FUTA~NI- <i>two</i> , TAKUSAN <i>many~much</i> , SUKOSHI <i>little~few</i> , IKUTSUKA <i>some</i> , MINNA <i>all</i>	UNO <i>one</i> , DOS <i>two</i> , MUCHO <i>much~many</i> , POCO <i>little~few</i> , ALGUNOS <i>some</i> , TODO <i>all</i>	ODIN <i>one</i> , DVA <i>two</i> , MNOGO <i>much~many</i> , MOLO <i>little~few</i> , NEKOTORYE <i>some</i> , VSE <i>all</i>
II <i>good</i> , WARUI <i>bad</i>	BUENO <i>good</i> , MALO <i>bad</i>	XOROŠIJ~XOROŠO <i>good</i> , PLOXOJ~PLOXO <i>bad</i>
OOKII <i>big</i> , CHIISAI <i>small</i>	GRANDE <i>big</i> , PEQUEÑO <i>small</i>	BOL' ŠOJ <i>big</i> , MALEN' KIJ <i>small</i>
OMOU <i>think</i> , SHIRU <i>know</i> , HOSHII~TAI~NOZOMU <i>want</i> , KANJIRU <i>feel</i> , MIRU <i>see</i> , KIKU <i>hear</i>	PENSAR <i>think</i> , SABER <i>know</i> , QUERER <i>want</i> , SENTIR <i>feel</i> , VER <i>see</i> , OÍR <i>hear</i>	DUMAT' <i>think</i> , ZNAT' <i>know</i> , XOTET' <i>want</i> , ČUVSTVOVAT' <i>feel</i> , VIDET' <i>see</i> , SLYŠAT' <i>hear</i>
IU <i>say</i> , KOTOBA <i>words</i> , HONTOO <i>true</i>	DECIR <i>say</i> , PALABRAS <i>words</i> , VERDAD <i>true</i>	GOVORIT'~SKAZAT' <i>say</i> , SLOVA <i>words</i> , PRAVDA <i>true</i>
SURU <i>do</i> , OKORU~OKIRU <i>happen</i> , UGOKU <i>move</i> , FURERU <i>touch</i>	HACER <i>do</i> , PASAR <i>happen</i> , MOVERSE <i>move</i> , TOCAR <i>touch</i>	DELAT' <i>do</i> , PROISXODIT'~SLUČAT'SJA <i>happen</i> , DVIGAT'SJA <i>move</i> , KASAT'SJA <i>touching</i>
(DOKOKA) IRU~ARU <i>be (somewhere)</i> , IRU~ARU <i>there is</i> , MOTSU <i>have</i> , (DAREKA/NANIKAI) DEARU <i>be (someone/something)</i>	ESTAR <i>be (somewhere)</i> , HAY <i>there is</i> , TENER <i>have</i> , SER <i>be (someone/something)</i>	BYT' (GDE-TO) <i>be (somewhere)</i> , BYT'~EST' <i>there is</i> , BYT' U <i>have</i> , BYT' (KEM-TO/ČEM-TO) <i>be (someone/something)</i>
IKIRU <i>live</i> , SHINU <i>die</i>	VIVIR <i>live</i> , MORIR <i>die</i>	ŽIT' <i>live</i> , UMERET' <i>die</i>
ITSU~TOKI <i>when~time</i> , IMA <i>now</i> , MAE <i>before</i> , ATO <i>after</i> , NAGAI AIDA <i>a long time</i> , MIJIKAI AIDA <i>a short time</i> , SHIBARAKU NO AIDA <i>for some time</i> , SUGUNI <i>moment</i>	CUÁNDO~TIEMPO <i>when~time</i> , AHORA <i>now</i> , ANTES <i>before</i> , DESPUÉS <i>after</i> , MUCHO TIEMPO <i>a long time</i> , POCO TIEMPO <i>a short time</i> , POR UN TIEMPO <i>for some time</i> , MOMENTO <i>moment</i>	KOGDA~VREMJA <i>when~time</i> , SEJČAS <i>now</i> , DO <i>before</i> , OSLE <i>after</i> , DOLGO <i>a long time</i> , KOROTKOE VREMJA, <i>a short time</i> , NEKOTOROE VREMJA <i>for some time</i> , MOMENT <i>moment</i>
DOKO~TOKORO <i>where~place</i> , KOKO <i>here</i> , UE <i>above</i> , SHITA <i>below</i> , CHIKAI <i>near</i> , TOOI <i>far</i> , MEN <i>side</i> , NAKA <i>inside</i>	DÓNDE~SITIO <i>where~place</i> , AQUÍ <i>here</i> , ARRIBA <i>above</i> , DEBAJO <i>below</i> , CERCA <i>near</i> , LEJOS <i>far</i> , LADO <i>side</i> , DENTRO <i>inside</i>	GDE~MESTO <i>where~place</i> , ZDES' <i>here</i> , NAD <i>above</i> , POD <i>below</i> , DALEKO <i>far</i> , BLIZKO <i>near</i> , STORONA <i>side</i> , VNUTRI <i>inside</i>
-NAI <i>not</i> , TABUN <i>maybe</i> , DEKIRU <i>can</i> , -KARA <i>because</i> , MOSHI (BA) <i>if</i>	NO <i>not</i> , TAL VEZ <i>maybe</i> , PODER <i>can</i> , PORQUE <i>because</i> , SI <i>if</i>	NE <i>not</i> , MOŽET BYT' <i>maybe</i> , MOČ' <i>can</i> , POTOMU ČTO <i>because</i> , ES LI <i>if</i>
SUGOKU <i>very</i> , MOTTO <i>more</i>	MUY <i>very</i> , MÁS <i>more</i>	OČEN' <i>very</i> , BOL' ŠE~EŠČE <i>more</i>
YOO~DOO~YOONI <i>like/how/as</i>	COMO <i>like</i>	KAK~TAK <i>like</i>

– 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, i.e., morphologically, complex – They can have combinatorial variants or allolxes (indicated with ~) – Each prime has well-specified syntactic (combinatorial) properties.

## Notes

1. I am grateful to Anna Wierzbicka for valuable input to this paper. Earlier versions were presented at the ALT Workshop on Lexical Typology held in Paris, September 2007, and at departmental seminars at Stockholm University and Uppsala University in the same month. I received stimulating feedback from participants at all three gatherings. The present paper has benefited considerably from suggestions from several anonymous reviewers and from the editor of this special issue. This research was supported by the Australian Research Council.
2. A bibliography and downloads can be found at the NSM Homepage: [www.uned.edu.au/bcss/linguistics/nsm](http://www.uned.edu.au/bcss/linguistics/nsm)
3. The most polysemous/multi-functional exponent in the English listing is probably HAVE. Its intended primitive sense of “possession” should be clear, however, from basic combinatorial contexts such as: ‘this someone has many things’ and ‘someone has many things of this kind’. In English, furthermore, semantic prime HAVE has a converse in *belong*, i.e. ‘if someone X HAS something Y’, then ‘something Y belongs to someone X’. These and related issues concerned with isolating and identifying semantically primitive meanings are discussed in Wierzbicka (1996: Ch 2-3), Goddard and Wierzbicka (2002), and Goddard (2008).
4. Would it be possible to devise a language-independent representation of NSM, with morphosyntactic spell-out rules into various different natural languages? In my opinion, this prospect is not theoretically coherent, because any sufficiently rich representational system, including computer languages, necessarily depends on a relationship with a known natural language. To pursue this issue further would take us too far from the focus of the present study.
5. The Strong Lexicalisation Hypothesis is stated in Goddard (1994: 7) as follows: “Apart from demonstrable cases of polysemy, every semantically primitive meaning can be expressed through a distinct word or morpheme in every language.” Immediately afterwards, it was stated that the hypothesis “does not entail that there should be a single unique lexical form for each primitive. Some languages have several word-forms (allomorphs of the same lexical item) functioning as contextual variants expressing the same primitive meaning. Conversely, it sometimes happens that the same form serves as an exponent of different primitives, although their distinct syntactic frames makes it appropriate to recognise polysemy. Also, the lexical exponent of a semantically primitive

meaning may be formally complex, including elements which function elsewhere as full morphemes.”

6. Khanina (2008) is a case in point. The author concludes after a survey of 73 languages that WANT cannot be a semantic prime in the NSM sense, because in many languages the exponent of WANT expresses other meanings as well, especially ‘like’, ‘love’, and ‘seek’ or, less commonly, modal-like meanings, such as ‘can’. The conclusion does not follow, however, because Khanina adopts a “no polysemy” assumption. Declining to engage in language-internal semantic analysis, she simply accepts the range of dictionary glosses for a given item in any language as representing a single general meaning or ‘macrofunction’. For a defence of WANT as a conceptual and linguistic universal, see Goddard and Wierzbicka (2010).
7. According to Stephen Levinson (p.c. email, 4 April 2011), this conclusion was not actually based mainly on Haviland’s (1979) grammar, but on Levinson’s own fieldwork (not mentioned in the published paper). Levinson states in his email: “there is no coded IF in GY, *budhu* is a dubitative particle, which can implicate but does not code IF”. Needless to say, the existence of unpublished field notes does not improve the quality of publically available data. Further, the wording “implicate[s] but does not code” seems to imply analytical assumptions that are open to question; for example, an assumption that *budhu* has a single abstract meaning, i.e. “dubitative”, in both monoclausal and biclausal constructions. Similar particles in the Australian languages Arrernte and Yankunytjatjara (*peke* and *tjinguru*, respectively) have been analysed as exhibiting polysemy between MAYBE (in single clauses) and IF (in biclausal constructions) (Harkins and Wilkins 1994; Goddard 1994). In view of the sentence cited by Haviland (1979), I would be inclined to think that the same applies to Guugu Yimidhirr.
8. All Everett (2005: 643) has to say about semantic methodology is: “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”. His implication is that Pirahã words for ‘all’ do not imply absolute exhaustiveness, but this overlooks the obvious point that ‘all’ in natural languages is not the same as universal quantification in logic. Ordinary English ‘all’ does not entail absolute exhaustiveness either.
9. Unlike English *finish* and *end*, Yukatek *ts’o’ok* is compatible with punctual verbs such as ‘die’ and ‘wake up’. Hence: “*Ts’o’ok* must be assumed to represent a type of phasal operator unattested and probably unparalleled in Indo-European languages ... This

underlines the status of *ts'o'ok* as an operator of temporal coherence rather than merely a lexical verb ...” (Bohnenmeyer 1998b: 270). It is notable that quasi-aspectual verbs have also been reported to be the usual exponents of BEFORE and AFTER in some Austronesian languages, such as Samoan (Mosel 1994: 349-354) and Acehnese (Durie, Daud, & Hasan 1994: 191-192); cf. Wierzbicka (1994: 485-487).

10. It is not true to say, then, that the usefulness of the NSM method depends entirely on the universality hypothesis. However, it is true that if it were to turn out that some languages lack exponents of certain primes, this would call into question the “meta-semantic adequacy” of those languages, i.e., whether they provide the resources for explicating their own meanings in language-internal terms. The expressive power of such a language would of course be redeemed 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 would imply semantic incommensurability in the areas concerned.
11. In earlier versions of explication [A], the (b) component was: ‘they are on two sides of the body’. The elaborated version presented here (‘these parts are parts of two other parts of this someone’s body on two sides of this someone’s body’) avoids the possibility that the explication will be initially misconstrued as referring to the arms, while recognising that the hands are themselves in a paronymic (meronymic) relationship with another pair of body parts. Component (e) is phrased in a slightly simpler form than its counterpart in Wierzbicka (2007).
12. They are: SOMEONE, PEOPLE, BODY, SOMETHING/THING; PART, KIND; THIS, THE SAME, OTHER; ONE, TWO, MANY, SOME, ALL; GOOD, BAD; SMALL; WANT, DO, HAPPEN, MOVE, LIVE; TIME, FOR SOME TIME, A LONG TIME, A SHORT TIME, BEFORE; SIDES; LIKE/AS; CAN, IF, NOT, BECAUSE. There is a lesson here: a fairly well developed metalanguage is needed to address even a small selection of examples.
13. Criticism of this calibre can hardly be taken seriously. One must also be wary of accepting tendentious counter-claims. For example, the same reviewer rejects explication [B] by saying: “How do we account for the fact that children, especially in non-Western cultures, are not necessarily conceptualised as persons?”. No checkable references, either linguistic or anthropological, were supplied in support of this alleged fact (or factoid). Although people in “non-Western cultures” can have radically different attitudes and values concerning children (for example, allowing them little freedom of action and/or opinion), it would be ethnocentric to assume that children in these cultures are not conceptualised as persons. Likewise, the documented culture-specificity of child-directed

speech (e.g. Schieffelin & Ochs 1986) in no way entails that children in different cultures are conceptualised as “non-persons”.

14. Since it is framed as ‘one part of someone’s body’, explication [D] obviously does not apply to the ‘heads’ of various kinds of animals; furthermore, an animal’s head is typically not positioned above the other parts of an animal’s body. The NSM view is that body-part words when applied to animals involve a polysemic extension from an anthropocentric prototype. The pattern governing the extension is of some interest in itself, because it depends on an analogical relation; e.g. *head*<sub>2</sub> (e.g. *a cat’s head* or *a snake’s head*): ‘one part of the body of a living thing of one kind; this part is like one part of people’s bodies; this part of people’s bodies is the head [m]’.
15. Explication [E] does not apply to the use of *long* in respects of places, such in expressions such as *a long river* or *long road*. Furthermore, even in relation to objects, explication [E] is not intended to cover dimensional *long*, e.g. *long arms*, *a long stick*. In these uses *long* describes a contingent attribute of an object, rather than its inherent “shape”. Furthermore, unlike shape descriptor *long*, dimensional *long* is gradable and it also contrasts with *short*. In general, shape descriptor words (including *round*, *flat*, *wide*, and others) exhibit considerable regular polysemy. These issues have been discussed at length in Wierzbicka (2006a), where explications for the extended meanings are presented.
16. The explication for the singular ‘child [m]’ is the same as for ‘children [m]’, except that the initial component is ‘someone of one kind’, rather than ‘people of one kind’.
17. Goddard (2001) expressed reservations about the universality of ‘father’, especially in comparison with ‘mother’. See Wierzbicka (to appear) for arguments that appear to vindicate the universal status of the meaning ‘father’. Goddard (2001) also claimed on the basis of Japanese, which appears to make a categorical distinction between *misu* (roughly) ‘water generally’ and *yu* ‘hot water’, that ‘water’ was not a universally lexicalised concept. This claim is now retracted (cf. Goddard 2010).
18. Sibly (2010) is a study of English verbs of physical contact, such as *hit*, *punch*, *slap*, *kiss*, and *kick*.
19. The paradox (cf. Levin & Rappaport Hovav 1992; Jackendoff 1990; Parsons 1990) turns on the fact that for a certain class of verbs (Vendler’s (1967) accomplishment verbs) the simple past tense entails a result-state, yet this entailment fails when the verb appears in the imperfective (progressive). For example, to say that *John cleaned the table* entails

that the table ended up clean, whereas to say that *John was cleaning the table* does not carry this entailment.

20. It would not be justifiable to employ the word ‘liquid’ as a semantic molecule, in preference to the expression ‘something like water [m]’, for three reasons. First, for a word-meaning to be considered for the status of a semantic molecule it must be intuitively plausible that it functions as a semantic unit in the conceptualisation of ordinary speakers, including children. It seems unlikely that ‘liquid’ enjoys this status. Second, words comparable in meaning to English *drink* are found in many languages that lack any expression corresponding to ‘liquid’; as far as we know, however, all languages have a word for ‘water’. Third, the meaning ‘water’ is known to be required as a semantic molecule in numerous words across the lexicon, e.g. *rain*, *river*, *wash*, and it has been explicated (Goddard 2010); hence, it is to be preferred on the basis of parsimony.
21. For discussions of Kalam from an NSM perspective, see Pawley (1994), Goddard (2001); see also Wierzbicka (1996: 200–202), Goddard (2001: 27–28).
22. The Prototypical Motivational Scenario component ‘I want this something to be two things’ might meet with the objection that (obviously) one can *cut* an object into more than two pieces. However, the reference to ‘two things’ is embedded in a prototypical frame (‘at many times someone does something like this when it is like this: ...’), so there is no claim that *cutting* is always intended to produce two pieces, and just as importantly, the first “cut” would normally separate two pieces from one another, and the same can be said about every subsequent “cut”. On the other hand, there are various lexicalized English expressions with generic mass noun objects (*cutting grass*, *cutting hair*), which do not sufficiently match the prototypical motivational scenario. Such expressions often do not have literal counterparts in other languages, even when those languages have verbs which approximate English *cut* in meaning. For example, in Polish one cannot say *ciąć trawę* (for ‘to cut grass’) or *ciąć włosy* (for ‘to cut hair’); one says *kosić trawę* and *obcinać włosy*, respectively. This indicates that the English expressions are lexicalised items which do not fully reflect the semantics of *cut* as an independent verb. It might also be objected that *cut* can be used in scenarios that do not involve a hand-held instrument, e.g. a guillotine. Such uses, however, normally occur in “perfective/resultative” contexts and often in particle-verb construction, such as in an English Simple Past sentence like *They cut off his head with a guillotine*. They are decidedly odd as “imperfectives”, cf. \**They were cutting his head with a guillotine*. As mentioned, different-but-related explications are required for these uses, and for a variety of other constructions; for

example, with an inanimate subject, such as *The glass cut his hand*. These matters are discussed at some length in Goddard and Wierzbicka (2009) and Goddard (to appear).

23. Notwithstanding the preponderance of taxonomic ('kind of') and meronymic ('part of') relations in the nominal lexicon, it is well to remember that there are many non-taxonomic macro-categories as well (Wierzbicka 1985; Goddard 2009b).
24. As an example of how divergent descriptive metalanguages "create obstacles for evaluating cross-linguistic connections even between studies of high semantic and lexicographic quality", Koptjevskaja-Tamm (2008) adduces Enfield's (2003) and Viberg's (2002, 2006) studies of Lao *daj* and Swedish *få*, respectively. Though the two verbs would appear to be similar in their basic meanings (something like 'get' or 'acquire') and in their patterns of grammaticalization, it is impossible to directly compare the studies on account of their divergent descriptive terminologies.



## References

- Apresjan, Jurij D. 1992. *Lexical semantics: User's guide to contemporary Russian vocabulary*. Ann Arbor: Karoma. [Orig. published 1974 as *Leksiceskaja semantika – sinonimeckie sredstva jazyka*, Moskva: Nauka.].
- Apresjan, Jurij D. 2000. *Systematic lexicography*. Trans. Kevin Windle. Oxford: Oxford University Press.
- Bromhead, Helen 2009. *The Reign of truth and faith: Epistemic expressions in 16th and 17th century English*. Berlin: Mouton de Gruyter.
- Bromhead, Helen 2011. Ethnogeographical categories in English and Pitjantjatjara/Yankunytjatjara. *Language Sciences* 33(1). 58–75.
- Brotherson, Anna 2008. The ethnogeometry of Makasai East Timor. In Cliff Goddard (ed.), *Cross-linguistic semantics*, 259–276. Amsterdam: John Benjamins.
- Buisson, D. 1992. *The art of Japanese paper*. Paris: Terrail.
- Enfield, N. J. 2003. *Linguistic epidemiology*. London: Routledge.
- Everett, Daniel L. 2005. Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology* 46(4). 621–646.
- Evans, Nicholas & Steven C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32. 249–492.
- Gladkova, Anna. 2010. *Russkaja kul'turnaja semantika: émocii, cennosti, žiznennye ustanovki* [Russian cultural semantics: Emotions, values, attitudes.]. Moscow: Languages of Slavic Cultures.
- Goddard, Cliff. 1994. Lexical primitives in Yankunytjatjara. In Cliff Goddard & Anna Wierzbicka (eds.), *Semantic and Lexical Universals*, 229–262. Amsterdam: John Benjamins.
- Goddard, Cliff. 2001. Lexico-semantic universals: A critical overview. *Linguistic Typology* 5(1). 1–66.
- Goddard, Cliff. 2006. Natural Semantic Metalanguage. In Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd edn, 544–551. Elsevier.
- Goddard, Cliff. 2008. Natural Semantic Metalanguage: The state of the art. In Cliff Goddard (ed.), *Cross-Linguistic Semantics*, 1–34. Amsterdam: John Benjamins.
- Goddard, Cliff. 2009a. The conceptual semantics of numbers and counting: An NSM analysis. *Functions of Language* 16(2). 193–224.

- Goddard, Cliff. 2009b. *Vegetables, furniture, weapons*: Functional macro-categories in the English lexicon. Paper presented at the *2009 Conference of the Australian Linguistic Society*. La Trobe University, Melbourne, July 2009.
- Goddard, Cliff. 2010. Semantic molecules and semantic complexity (with special reference to “environmental” molecules). *Review of Cognitive Linguistics* 8(1). 123–155.
- Goddard, Cliff. 2011. *Semantic Analysis* [Revised 2<sup>nd</sup> edition]. Oxford: Oxford University Press.
- Goddard, Cliff. To appear. Verb classes and valency alternations (NSM approach), with special reference to English physical activity verbs. In Andrej Malchukov & Bernard Comrie (eds.), *Valency Classes in the World’s Languages* [provisional title].
- Goddard, Cliff (ed.). 2006. *Ethnopragmatics: Understanding discourse in cultural context*. Berlin: Mouton de Gruyter.
- Goddard, Cliff (ed.). 2008. *Cross-Linguistic semantics*. Amsterdam: John Benjamins.
- Goddard, Cliff & Anna Wierzbicka. 2002. Semantic primes and universal grammar. In Cliff Goddard & Anna Wierzbicka (eds.), *Meaning and universal grammar – Theory and empirical findings, Vol I*, 41–85. Amsterdam: John Benjamins.
- Goddard, Cliff & 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(1). 675–800.
- Goddard, Cliff & Anna Wierzbicka. 2009. Contrastive semantics of physical activity verbs: ‘Cutting’ and ‘chopping’ in English, Polish, and Japanese. *Language Sciences* 31. 60–96.
- Goddard, Cliff & Anna Wierzbicka. 2010. WANT is a linguistic and conceptual universal (response to Khanina). *Studies in Language* 34(1). 108–123.
- Goddard, Cliff & Anna Wierzbicka. To appear. *Men, women and children*: The conceptual semantics of basic social categories. In *Words and meanings: Lexical semantics across domains, languages and cultures*.
- Goddard, Cliff & Anna Wierzbicka (eds.). 1994. *Semantic and lexical universals – Theory and empirical findings*. Amsterdam: John Benjamins.
- Goddard, Cliff & Anna Wierzbicka (eds.). 2002. *Meaning and universal grammar – Theory and empirical findings. Vols I and II*. Amsterdam: John Benjamins.
- Harkins, Jean & David P. Wilkins. 1994. Mparntwe Arrernte and the search for lexical universals. In Cliff Goddard & Anna Wierzbicka (eds.), *Semantic and Lexical Universals*, 285–310. Amsterdam: John Benjamins.

- Harkins, Jean & Anna Wierzbicka (eds.). 2001. *Emotions in cross-linguistic perspective*. Berlin: Mouton de Gruyter.
- Haviland, John. 1979. Guugu Yimidhirr. In R. M. W. Dixon & Barry J. Blake (eds.), *Handbook of Australian languages*, 27–182. Canberra: Australian National University Press.
- Honda, I. 1965. *The world of origami*. Tokyo & San Francisco: Japan Publications Trading Company & Elmsfor.
- Jackendoff, Ray. 1990. *Semantic structures*. Cambridge, MA: MIT Press.
- Khanina, Olesya. 2008. How universal is *wanting*? *Studies in Language* 32(4). 818–865.
- Koch, Peter. 2001. Lexical typology from a cognitive and linguistic point of view. In Martin Haspelmath, Ekkehard König & Bernard Comrie (eds.), *Language typology and language universals, Vol 2*, 1142–1178. Berlin: Walter de Gruyter.
- Koptjevskaja-Tamm, Maria. 2008. Approaching lexical typology. In Martine Vanhove (ed.), *From polysemy to semantic change: A typology of lexical semantic associations*, 3–54. Amsterdam: John Benjamins.
- Lehrer, Adrienne. 1992. A theory of vocabulary structure: Retrospectives and prospectives. In Martin Pütz (ed.), *Thirty years of linguistic evolution*, 243–256. Amsterdam: John Benjamins.
- Levin, Beth & Malka Rappaport Hovav. 1992. Wiping the slate clean: A lexical semantic exploration. In Beth Levin & Steven Pinker (eds.), *Lexical & conceptual semantics*, 123–151. Cambridge, MA: Blackwell. (Reprinted from *Cognition* 41. 1991, 123–151)
- Majid, Asifa & Melissa Bowerman (eds.). 2007. Cutting and breaking events across languages. [Special issue] *Cognitive Linguistics* 18(2).
- Mel'čuk, Igor. 1988. Semantic description of lexical units in an explanatory combinatorial dictionary: Basic principles and heuristic criteria. *International Journal of Lexicography* 1. 165-188.
- Mel'čuk, Igor. 2006. Explanatory Combinatorial Dictionary. In Giandomenico Sica (ed.), *Open problems in linguistic and lexicography*, 225-355. Monza, Italy: Polimetrica.
- Morita, Y. C. 1989. *Kiso Nihongo jito [Basic Japanese dictionary]*. Tokyo: Kadokawa Shoten.
- Newman, John (ed.). 2009. *The linguistics of eating and drinking*. Amsterdam: John Benjamins.
- Parsons, Terence. 1990. *Events in the semantics of English: A study in subatomic semantics*. Cambridge, MA: MIT Press.

- Pawley, Andrew. 1994. Kalam exponents of lexical and semantic primitives. In Cliff Goddard & Anna Wierzbicka (eds.), *Semantic and lexical universals – Theory and empirical findings*, 387–422. Amsterdam: John Benjamins.
- Pawley, Andrew & R. N. H. Bulmer. In press. *A dictionary of Kalam with ethnographic notes*. Canberra: Pacific Linguistics.
- Peeters, Bert (ed.). 2006. *Semantic Primes and universal grammar: Empirical evidence from the romance languages*. Amsterdam: John Benjamins.
- Schieffelin, Bambi B. & Elinor Ochs (eds.) 1986. *Language socialization across cultures*. Cambridge: Cambridge University Press.
- Sibly, Anne. 2010. Harry slapped Hugo, Tracey smacked Ritchie: The semantics of *slap* and *smack*. *Australian Journal of Linguistics* 30(3). 323–348.
- Vendler, Zeno. 1967. *Linguistics in philosophy*. Ithaca, NY: Cornell University Press.
- Viberg, Åke. 2002. Polysemy and disambiguation cues across languages. The case of Swedish *få* and English *get*. In Bengt Altenberg & Sylviane Granger (eds.), *Lexis in contrast*, 119–150. Amsterdam: John Benjamins.
- Viberg, Åke. 2006. Towards a lexical profile of the Swedish verb lexicon. *STUF* (Sprachtypol. Univ. Forsch.) 59(1). 103–129.
- Wierzbicka, Anna. 1972. *Semantic primitives*. Frankfurt: Athenäum.
- Wierzbicka, Anna. 1985. *Lexicography and conceptual analysis*. Ann Arbor: Karoma.
- Wierzbicka, Anna. 1991. Semantic complexity: Conceptual primitives and the principle of substitutability. *Theoretical Linguistics* 17. 75–97.
- Wierzbicka, Anna. 1992. *Semantics, culture, and cognition*. Oxford: Oxford University Press.
- Wierzbicka, Anna. 1996. *Semantics: Primes and universals*. Oxford: Oxford University Press.
- Wierzbicka, Anna. 1999. *Emotions across languages and cultures: Diversity and universals*. Cambridge: Cambridge University Press.
- Wierzbicka, Anna. 2005. Comments on Daniel L. Everett, Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology* 46(4). 641.
- Wierzbicka, Anna. 2006a. Shape in grammar revisited. *Studies in Language* 30(1). 115–177.
- Wierzbicka, Anna. 2006b. There are no ‘color universals’, but there are universals of visual semantics. *Anthropological Linguistics* 47(2). 217–244.
- Wierzbicka, Anna. 2007a. Bodies and their parts: An NSM approach to semantic typology. *Language Sciences* 29. 14–65.

- Wierzbicka, Anna. 2007b. Shape and colour in language and thought. In Andrea C. Schalley & Drew Khlenzoz (eds.), *Mental states: Vol 2: Language and cognitive structure*, 37–60. Amsterdam: John Benjamins.
- Wierzbicka, Anna. 2008. Why there are no ‘colour universals’ in language and thought. *Journal of the Royal Anthropological Institute* 14. 407–425.
- Wierzbicka, Anna. 2009. All people eat and drink: Does this mean that ‘eat’ and ‘drink’ are universal human concepts? In John Newman (ed.), *The linguistics of eating and drinking*, 65–89. Amsterdam: John Benjamins.
- Wierzbicka, Anna. 2010. ‘Eating’ and ‘drinking’ in Kalam. In John Bowden, Nikolaus Himmelmann & Malcolm Ross (eds.), *A Journey through Austronesian and Papuan linguistic and cultural space: Papers in honour of Andrew K. Pawley*, 651–662. Canberra: Pacific Linguistics.
- Wierzbicka, Anna. To appear. Universal of kinship: Overcoming Eurocentrism in kinship semantics using seven universal semantic molecules.
- Wong, Jock. 2005. “Why you so Singlish one?” A semantic and cultural interpretation of the Singapore English particle *one*. *Language in Society* 34(2). 239–275.
- Wong, Jock. 2010. The triple articulation of language. *Journal of Pragmatics* 42(11). 2932–2944.
- Ye, Zhengdao. 2007a. ‘Memorisation’, learning and cultural cognition: The notion of *bèi* (‘auditory memorisation’) in the written Chinese tradition. In Mengistu Amberber (ed.), *The language of memory in a crosslinguistic perspective*, 139–180. John Benjamins: Amsterdam.
- Ye, Zhengdao. 2007b. Taste as a gateway to Chinese cognition. In Andrea C. Schalley & Drew Khlenzoz (eds.), *Mental states: Volume 2: Language and cognitive structure*, 109–132. Amsterdam: John Benjamins.
- Ye, Zhengdao. 2010. Eating and drinking in Mandarin and Shanghainese: A lexical-conceptual analysis. In E. Christensen, E. Schier & J. Sutton (eds.), *ASCS09: Proceedings of the 9th Conference of the Australasian Society for Cognitive Science*, 375–83. Sydney: Macquarie Centre for Cognitive Science.