

Studies in Language 2014, 38(1), 80–127. [Please note that this is a pre-publication manuscript copy and does not correspond to the published version in every exact detail. In particular, the page numbers do not match those of the final publication; the artwork in Figure 1 is different in the final publication; and some minor formatting and referencing details may differ.]

Semantic fieldwork and lexical universals

Cliff Goddard and Anna Wierzbicka

Griffith University, Australian National University

The main goal of paper is to show how NSM findings about lexical universals (semantic primes) can be applied to semantic analysis in little-described languages. It is argued that using lexical universals as a vocabulary for semantic analysis allows one to formulate meaning descriptions that are rigorous, cognitively authentic, maximally translatable, and free from Anglocentrism. A second goal is to shed light on methodological issues in semantic fieldwork by interrogating some controversial claims about the Dalabon and Pirahã languages. We argue that reductive paraphrase into lexical universals provides a practical procedure for arriving at coherent interpretations of unfamiliar lexical meanings. Other indigenous/endangered languages discussed include East Cree, Arrernte, Kayardild, Karuk, and Maori. We urge field linguists to take the NSM metalanguage, based on lexical universals, into the field with them, both as an aid to lexicogrammatical documentation and analysis and as a way to improve semantic communication with consultants.

Keywords: lexical semantics; linguistic fieldwork; endangered languages; NSM; reductive paraphrase; Anglocentrism

1. Introduction

With language endangerment – finally – moving up the mainstream linguistic agenda, recent years have seen a resurgence of interest in linguistic fieldwork and language documentation. A number of important collections have appeared (Austin ed. 2003-2010; Gippert et al. 2006; Ameka et al. 2006; Grenoble and Furbee eds. 2010), along with new textbooks (Crowley 2007; Bower 2008; Sakel and Everett 2012) and handbooks (Thieberger 2012). Reviewing all this new literature, one sees surprisingly little on the topic of semantic fieldwork (cf. Hellwig 2010: 804). Questions such as the following, however, would be of burning interest to most linguistic fieldworkers: How can I work out the meanings of unfamiliar words and grammatical elements? How can I tell when words are polysemous? How can I reduce the danger of imposing categories and meanings from my language onto the language I am

describing? How can I improve communication about meanings with my language consultants and teachers? Equally, linguists generally are increasingly confronted with important – sometimes startling – claims about human languages and cognition based on data from far-flung and little described languages. In this paper we set out to provide some answers and advice about these questions and issues from the perspective of the NSM approach to semantics (Wierzbicka 1996; Goddard and Wierzbicka eds. 2002; Peeters ed. 2006; Goddard ed. 2008; Goddard 2011a; Ye 2007, 2010; Gladkova 2010; Bromhead 2009, 2011a, 2011b; Wong 2005, 2010, in press; Levisen 2012; and other works). See also: the NSM Homepage [<http://www.griffith.edu.au/humanities-languages/school-languages-linguistics/research/natural-semantic-metalanguage-homepage/publications>].

Although it is increasingly recognised that many grammatical phenomena are essentially meaning-based, many linguists still regard lexical semantics as lacking rigorous analytical tools (except perhaps in relation to quantification and other logic-related phenomena that are held to be amenable to formal semantics; cf. Matthewson (2004)). The under-developed state of the literature on semantic fieldwork therefore reflects, in many ways, a broader situation. NSM researchers have been steadily developing techniques and tools for lexical semantic analysis over 30 or so years and the approach has a strong track record in cross-linguistic semantics. If we may say so, we are heartened by the growing acceptance that NSM is receiving; for example, its inclusion in encyclopedias, handbooks, introductory textbooks, and the like. [Note 1] We are hopeful that a slow “seismic shift” is underway from the formalist orthodoxy of the last century to a near future in which meaning has been restored to its rightful central place in linguistic theory and practice.

Endangered languages, and the fieldworkers trying to document them, however, cannot afford to wait another 30 years for this shift to work its way through the discipline as a whole. They need useful techniques for grappling with meaning analysis here and now. Hence the sense of urgency we feel about getting the message out that NSM semantics in general, and in particular, its empirical findings about lexical universals, can “work” for fieldworkers. The NSM technique of reductive paraphrase makes meanings tangible, concrete, and amenable to discussion with consultants and language teachers. The same approach can enable the analyst to cut through perennial confusions about when it is or isn’t justified to recognise polysemy. Skilfully used, it can bring to light culture-specific conceptualisations and ways of thinking that are embedded in the lexicons and grammars of different languages, and thus help to fulfil the promise of field linguistics to document the untold diversity of human ways of knowing, thinking, and feeling.

The paper is structured as follows. Section 2 outlines the NSM approach with particular reference to the typological descriptive perspective. Needless to say, the approach is an intensional (conceptual) one, whose aim is to uncover what words and other expressions mean to the speakers of a language. [Note 2] Section 3 discusses how lexical universals can be used in fieldwork: as a worthy topic for documentation in itself, as a way into the lexicogrammar of the language being studied, and to help with communicating effectively about meaning with native speakers. We argue that the list of semantic primes and their grammar should be in every field linguist's backpack. Section 4 examines some issues in semantic analysis in relation to three languages from far afield: Dalabon (Evans 2007, 2010), East Cree (Junker 2003, 2008), and Pirahã (Everett 2005, 2008). These issues include the relationship between morphological analysis and semantic analysis, the use of reductive paraphrase as a procedure for deciding between monosemy and polysemy, and the roles of experimental data and interpretative analysis. Section 5 argues that conventional semantic approaches depend too heavily on complex English-specific meanings and that this makes it difficult, if not impossible, to faithfully capture indigenous meanings, especially, the meanings of heavily culture-laden words. Section 6 reviews the main conclusions.

2. NSM metalanguage and operating principles

NSM semantics is based on reductive paraphrase into natural language, and reductive paraphrase implies the existence of semantic primes: the irreducible terminal elements of the reductive process. An ideal reductive paraphrase (or explication) goes right down to the level of semantic primes (because this makes the full semantic content explicit), but a less than fully resolved paraphrase can still be very illuminating of semantic relations. The main thing is to ensure that all the words in the explication are semantically simpler than the original expression. This means that the analytical process is moving in the right direction, i.e. resolving complex meanings into simpler ones, and at the same time it ensures that the analysis is free of circularity. A successful reductive paraphrase will predict and/or explain natural usage (including distribution, collocations, entailments, implications, and so on) and it will satisfy native speaker intuitions. Because it is carried out in terms that are known to speakers and that form part of their everyday linguistic competence, a successful reductive paraphrase can have a *prima facie* claim to conceptual authenticity, in the sense of representing what anthropologists call an “insider perspective”.

The inventory of semantic primes – i.e. posited lexico-semantic universals – is given in Table 1 using English exponents. [Note 3] As discussed below, comparable tables have been

drawn up for over 30 languages, including languages from a range of language families, language types, and geographical locations. It is relevant to mention that the NSM metalanguage was not developed solely on the basis of Indo-European languages. The originator of the approach, Anna Wierzbicka, has been based for most of her academic career at the Australian National University, a leading centre for language description, in a milieu which encouraged dialogue with fieldworkers from many parts of the world, particularly Australia, Papua New Guinea, Oceania, and Southeast Asia. High-quality data from little-documented languages was freely available and, naturally therefore, was often drawn upon. For example, in the opening chapter (Goddard 2002) of the two-volume *Meaning and Universal Grammar* collection, almost all the main languages used for data and examples were non-Indo-European: Yankunytjatjara, Yolngu Matha and Bunuba (from Australia), Kalam and Manganaba-Mbula (from PNG), and Malay, Acehnese and Samoan (SE Asia and Oceania). Experts in Asian languages such as Chinese, Japanese, Korean and Thai/Lao have also made contributions in the development of the current version of the metalanguage.

At the bottom of Table 1 (next page), a number of important caveats about semantic primes are given in dot-point form. The chief one is that the relationship between a semantically primitive meaning and a form that represents it in a given language (a so-called ‘exponent’ of the prime) is seldom one-to-one. An exponent in a given language (say, English *move* or *happen*) can be polysemous, i.e. it can have other, non-primitive meanings in addition to its semantically prime meaning. For example, the English word *move* has one meaning (its semantically prime meaning) in a sentence like *I can’t move* and a distinct, semantically complex meaning in a sentence like *I’m moving house tomorrow* (cf. section 3.1 below.) Conversely, a semantically primitive meaning can have more than one exponent, a situation which NSM researchers term *allosemy*. For example, in English the words *other* and *else* express the same meaning, *other* being used as a modifier of nouns and *else* as a modifier of indefinite pronouns.

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
THINK, KNOW, WANT, DON'T WANT, FEEL, SEE, HEAR	Mental predicates
SAY, WORDS, TRUE	Speech
DO, HAPPEN, MOVE, TOUCH	Actions, events, movement, contact
BE (SOMEWHERE), THERE IS, BE (SOMEONE/SOMETHING), BE (SOMEONE'S)	Location, existence, specification, possession
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	Augmentor, intensifier
LIKE	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, i.e., morphologically, complex ■ They can have combinatorial variants or allolexes (indicated with ~) ■ Each prime has well-specified syntactic (combinatorial) properties.

Table 1: Semantic primes (English exponents) (Goddard & Wierzbicka 2002; 2014).

Polysemy and allolexy are language-specific phenomena, albeit certain patterns tend to recur across languages (Goddard 2012). Thus, when we say that English *do* and Spanish *hacer* are exponents of the same prime, this is not the same as saying that English *do* and Spanish *hacer* are “the same” in all their uses, because both words have additional language-specific meanings over and above their shared semantically primitive meaning; for example, *hacer* has an additional meaning that approximates English ‘make’. Equally, when we say that English *other* and Spanish *otro* are exponents of the same prime, this claim is not disturbed by the fact that English *other* has an allolex *else* that has no counterpart in Spanish.

Obviously, in order to be combined into phrases and sentences, semantic primes must have a grammar; or to put it another way, the Natural Semantic Metalanguage consists both of a mini-lexicon and a mini-grammar. The grammar of semantic primes has been explored in a fairly systematic way for over 20 years and, though a couple of murky areas remain, it is now well documented and can be regarded as substantially well understood. As well as simple combinatorial possibilities (e.g. that substantives can combine with specifiers: ‘this something~thing’, ‘someone else’, ‘one place’, ‘two parts’, ‘many kinds’, etc.), the grammar of semantic primes includes basic and extended valencies of predicates and quantifiers, and

complement options for the mental primes, KNOW, THINK and WANT. The available evidence indicates that the combinatorial properties of semantic primes are the same in all or most languages (Goddard and Wierzbicka eds. 2002; Goddard ed. 2008).

NSM grammar is significantly complex. Although groups of primes share particular properties and can be regarded as falling into natural classes, virtually every prime has some idiosyncratic combinatorial properties. For this reason, it is not possible here to give a sketch of the full NSM grammar. The key grammatical properties are displayed in the Chart of NSM Semantic Primes, available at [www.griffith.edu.au/humanities-languages/school-languages-linguistics/research/natural-semantic-metalanguage-homepage/in-brief]. For illustrative purposes, Table 2 displays the valency and complement frames for three semantic primes (HAPPEN, DO, THINK), using English exponents. Naturally, the entries in Table 2 include various English-specific morphosyntactic devices, e.g. the prepositions *to*, *with*, and *about*, but the claim that semantic equivalents of these frames are available in all languages does not entail that other languages realise the frames in an analogous fashion. 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 that they will pattern in a fashion analogous to English; for example, there is no requirement that the instrument and comitative roles of DO will be marked by the same formal means, as they are in English by means of preposition *with*.

Many linguists accept notions of undergoer, agent, patient and instrument, but typically these notions are thought of as independent entities of some kind (semantic or thematic roles), rather than as argument slots of basic predicates such as HAPPEN and DO. In some cases, NSM researchers propose valency options which are seldom recognised in mainstream grammars and which lack standard labels. For example, it is claimed that semantic prime THINK universally allows a “topic of cognition” option, such that one can say, in all languages, the semantic equivalent of a sentence like ‘I was thinking about this someone (this thing, this place, etc.)’. Likewise, it is claimed that LIVE universally allows an option of LIVE WITH (SOMEONE/SOME PEOPLE). The idea that a woman, man or child ‘can live with’ someone else is a key semantic ingredient of concepts like ‘husband’, ‘wife’ and ‘family’ (Wierzbicka to appear). In some cases, NSM forces us to distinguish between semantic valency options, e.g. BE WITH SOMEONE and DO SOMETHING WITH SOMEONE, that may be lumped together under conventional terminology.

something HAPPENS	[minimal frame]
something HAPPENS to someone/something	[undergoer frame]
something HAPPENS somewhere	[locus frame]

someone DOES something	[minimal frame]
someone DOES something to someone else	[patient ₁ frame]
someone DOES something to something	[patient ₂ frame]
someone DOES something to someone/something with something	[instrument frame]
someone DOES something with someone	[comitative frame]

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]

Table 2. Valency frame arrays for three semantic primes HAPPEN, DO and THINK

To recapitulate then, the Natural Semantic Metalanguage consists of a small, well-specified vocabulary from which phrases, sentences and texts can be constructed according to well-specified grammatical rules. Evidence indicates that this highly constrained vocabulary and grammar has equivalents in all or most languages of the world, which means that semantic explications framed in this metalanguage can be transposed without loss or distortion of meaning from one language to another.

Table 3 (next page) gives details of a selection of languages that have been studied in depth from an NSM point of view. Languages for which a full version of the metalanguage has been documented are specially indicated. Despite significant gaps in coverage (in particular, there are few indigenous languages of the Americas, Africa, Central and South Asia), these studies justify a high degree of confidence that the proposed semantic primes have lexical (word-like) status in all or most languages of the world. This conclusion follows from two facts. The first is the observation that most non-primitive meanings, even those which seem intuitively basic to native speakers of English, such as ‘go’ and ‘hot’, do not have equivalents even in a very small sample of other languages. A familiar language like German, for example, has no precise equivalent of ‘go’, because *gehen*, the nearest counterpart, cannot be used to speak about going somewhere by means of a vehicle; a distinct verb *fahren* is required for this. As for ‘hot’, French has no precise equivalent, because the nearest candidate *chaud* covers English ‘warm’, as well as ‘hot’. In this perspective, the fact that lexical exponents of semantic primes are attested in more than 30 diverse languages assumes greater evidentiary significance.

Language family/type	Language(s)	Sources (not necessarily exhaustive)
Austronesian	Malay†, Mbula†, Longgu, Samoan	Goddard (2001a, 2001b), Bugenhagen (2001, 2002), Hill (1994), Mosel (1994)
Indo-European	Spanish†, Danish†, French†, Polish†, Russian†	Travis (2002, 2004, 2006), Levisen (2012), Peeters (2006, 2010), Wierzbicka (2002), Gladkova (2007, 2010)
Semitic	Amharic†, Arabic†, Hebrew†	Amberber (2008), Habib (2011a, 2011b)
Finno-Ugric	Finnish†	Vanhatalo et al. (forthcoming)
Sinitic	Chinese (Mandarin)†, Cantonese	Chappell (2002), Ye (2006, 2007, 2010), Tong et al. (1997), Tien (2009), Leung (2012), Wakefield (2011)
Algonquian	†East Cree	Junker (2003, 2007, 2008), Junker and Blacksmith (2006)
Niger-Congo	Ewe	Ameka (1994, 2006, 2009)
Japonic	Japanese†	Hasada (2001, 2008), Asano-Cavanagh (2009, 2010)
Korean	Korean†	Yoon (2006, 2007)
Tai-Kadai	Lao†	Enfield (2002)
Papuan	Koromu†, Kalam, Makasai	Priestley (2002, 2008, 2102a, 2012b), Pawley (1994), Brotherson (2008)
Pama-Nyungan (Australia)	Pitjantjatjara/Yankunytjatjara, Arrernte	Goddard (1991, 1994), Bromhead (2011a), Harkins and Wilkins (1994), Harkins (2001)
Non-Pama-Nyungan (Australia)	Bunuba	Knight (2008)
Creoles	Hawaii Creole English, Roper Kriol†	Stanwood (1997, 1999), Nicholls (2009)

Table 3. Selection of languages studied in the NSM framework (full metalanguage studies marked with †)

The second reason to expect that semantic primes are robustly lexicalised across languages is that a large number of counter-claims and/or apparent counter-examples have been investigated. Most of these turn out to be based on a failure to appropriately recognize polysemy in the language concerned. [Note 4] For a recent example, see the exchange in this journal between Khanina (2008) and Goddard and Wierzbicka (2010) on WANT. Other exchanges about the supposed non-lexicalisation of primes in particular languages, include: Bohnemeyer (1998, 2003) and Goddard (2001c) on BEFORE and AFTER in Yukatek Maya; 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 (2011b) on WORDS. Except for a handful of unresolved cases, the lexical status of the primes has been upheld. [Note 5]

So far we have been talking about the deepest level of the semantic metalanguage, i.e. the level of semantic primes and their grammar. NSM semantics also makes extensive use of semantic molecules, i.e., complex semantic units, themselves decomposable into semantic primes, that function as building blocks in the meanings of other, still more complex meanings (Goddard 2010a; Wierzbicka 2009, forthcoming). Examples of semantic molecules

include ‘hand [m]’, ‘child [m]’, ‘ground [m]’, ‘water [m]’, ‘round [m]’, and ‘grow (in the ground) [m]’. Molecules are marked in explications with the notation [m]. Semantic molecules allow explications to be phrased more simply and comprehensibly, but this is not their main rationale. Using semantic molecules in explications embodies a claim about semantic dependency between concepts; for example, when we include the molecule ‘child [m]’ in the explication for *woman*, the claim is that the concept behind *woman* depends conceptually (in part) on the concept of ‘child’. The concept of semantic molecules is highly pertinent to cross-linguistic semantics, and hence to semantic fieldwork, because some semantic molecules are very likely universal, some are near-universal, and others are highly language-specific. The candidates for universal and near-universal molecules should be targets for lexical investigation in any language documentation project. A provisional partial list is given in Table 4, though it must be emphasised that research into semantic molecules is decades behind that into semantic primes. Equally, the fieldworker should be on the lookout for language-specific molecules that may hold the key to areas of cultural elaboration.

Body-parts and products	hands, mouth, eyes, head, ears, nose, face, legs, teeth, fingers, fingernails, breasts, skin, blood, poo
Social categories and family	children, men, women, mother, father, wife, husband, be born
Physical	hard, long, round, flat, sharp, heavy, thick, thin, smooth
Environmental and ambient	sky, ground, sun, fire, water, day, night
Life forms and related words	grow (in the ground), creature, bird, fish, tree, egg
Materials	wood, stone
Actions and activities	hold, make, play, kill, laugh, sing
Manner	quickly, slowly

Table 4. Provisional list of possible universal or near-universal molecules

Finally in this sketch of NSM operating principles, we want to highlight the point that reductive paraphrase provides a decision procedure for settling an analytical issue that can appear imponderably difficult, namely, how to decide on questions of polysemy. Goddard (2000: 132) put it as follows:

One assumes to begin with that there is but a single meaning, and attempts to state it in a clear and predictive fashion, in the form of a translatable reductive paraphrase. Only if persistent efforts to do this fail is polysemy posited. The next hypothesis is that there are two distinct meanings, and attempts are made to state both in a clear and predictive

fashion, and so the process goes, until the full range of application of the word can be captured within the specified range of senses. (Goddard 2000: 132)

In other words, to justify a monosemy analysis we must be able to state the posited single meaning in the form of a single coherent, non-disjunctive reductive paraphrase which is substitutable across the full range of attested uses. To justify two-way polysemy, we must be able to state the two posited meanings in the form of two coherent, non-disjunctive reductive paraphrases and show that they account for the full range of use; and so on. It must be admitted that applying these procedures is not necessarily a simple matter. Like other forms of linguistic analysis, semantic analysis requires training, skill and sustained effort over time (Goddard 2011). Some illustrations will be given in sections 3–5. Many NSM studies have dealt with lexical polysemy, either as a focus of interest in itself, or, more commonly, in connection with some broader topic of interest. For some recent studies of polysemy in English, the interested reader could consult Goddard (2007, 2011b), Wierzbicka (2010a, 2010b), and Bromhead (2011b).

3. NSM as a tool in language documentation

In this section we argue that the inventory of lexical universals discovered by NSM researchers is relevant in the field, both as a tool for lexical semantic analysis and also as a part of basic lexicogrammatical documentation.

3.1 Semantic primes and their grammar as part of basic language documentation

In our view, the inventory of semantic primes should be an item in the toolkit (and backpack) of every field linguist. Why? First and foremost, because of its *prima facie* claim, based both on theoretical and empirical work (including extant studies of over 30 languages), to represent a shared lexicosemantic core of all languages. This gives it a theoretical significance far beyond that of practical wordlists commonly used in field linguistics, such as the Comrie and Smith (1977) list shown in Table 5. This list was originally part of a *Lingua Descriptive Studies Questionnaire*, designed to gather in comparable lexical and grammatical information about many languages of the world. It has since been used or adapted for use in other projects and settings, and was included in Bower's (2008) textbook on field linguistics. The list comes in English alphabetical order. [Note 6]

all, and, animal, ashes, at, back, bad, bark, because, belly, bit, bird, bite, black, blood, bone, blow, breast, breathe, burn, child, claw, cloud, come, count, cut, day, die, dig, dirty, dog, drink, die, dull, dust, ear, earth, eat, egg, eye, fall, far, fat/grease, father, fear, feather, few, fight, fire, fish, five, float, flow, flower, fly, fog, foot, four, freeze, fruit, full, give, good, grass, green, guts, hair, hand, he, head, hear, heart, heavy, here, hit, hold/take, horn, how, hunt, husband, I, ice, if, in, kill, knee, know, lake, laugh, leaf, leftside, leg, lie (be in lying position), live, liver, long, louse, man/male, many, meat/flesh, moon, mother, mountain, mouth, name, narrow, near, neck, new, night, nose, not, old, one, other, person, play, pull, push, rain, red, right/correct, rightside, river, road, root, rope, rotten, round, rub, salt, sand, say, scratch, sea, see, seed, sew, sharp, short, sing, sit, skin, sky, sleep, small, smell, smoke, smooth, snake, snow, some, spit, split, squeeze, stab/pierce, stand, star, stick, stone, straight, suck, sun, swell, swim, tail, that, there, they, thick, thin, think, this, thou, three, throw, tie, tongue, tooth, tree, turn, two, vomit, walk, warm, wash, water, we, wet, what, when, where, white, who, wide, wife, wind, wing, wipe, with, woman, woods, worm, ye, year, yellow

Table 5. Wordlist frequently used in field linguistics (Comrie and Smith 1977; Bower 2008)

Most semantic primes are, impressionistically speaking, pretty “ordinary” meanings, so one might have expected that a basic wordlist of over 200 items devised by experienced field linguists would turn out to include all or most of them. Many are there, of course (such as ‘good’, ‘I’, ‘hear’, ‘here’, ‘if’, ‘know’ and ‘many’), and some that are not immediately obvious are there implicitly; for example, ‘what’ and ‘who’ correspond to SOMETHING and SOMEONE, ‘when’ and ‘where’ correspond to WHEN~TIME and WHERE~PLACE, ‘in’ corresponds to INSIDE. Even so, this still leaves the following 30-plus elements – more than half of the prime inventory – missing from the Comrie and Smith (1977) list:

PEOPLE, BODY, KIND, PARTS, THE SAME, FEEL, WANT, WORDS, TRUE, DO, HAPPEN, TOUCH, MOVE, THERE IS, BE (SOMEWHERE), BE (SOMEONE/SOMETHING), BE (SOMEONE’S), NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME, MOMENT, HERE, ABOVE, BELOW, SIDE, MAYBE, CAN, VERY, MORE, LIKE.

The absence of words like DO, HAPPEN, WANT, LIKE, BEFORE, AFTER, ABOVE, BELOW, MAYBE and VERY seems unfortunate, given that linguists have such a high interest in agents, undergoers, volition and control, similarity, time reference, spatial relations, epistemic certainty, and gradability, and, in our experience, it is usually not difficult to identify lexical exponents of these concepts. Exponents of KIND and PARTS can sometimes be more difficult to identify, but their omission is unfortunate in view the importance of relations of taxonomic hierarchy (based on KIND) and paronymy/meronymy (based on PARTS) throughout the nominal lexicon. In our opinion, fieldworkers would be well served by finding out words and

expressions for these expressions as early as they can. Certainly, they should be included in any basic lexical documentation.

A second observation that emerges from a comparison of standard wordlists and the NSM inventory of primes is that the items on the standard wordlists bear no particular relationship to lexicogrammar. The NSM list, on the other hand, can “work” for the field linguist as a way into the lexicogrammar of the language. It is a broad semantically-based sample of important words and constructions from different domains: events and actions, mental predicates, existence, possession, time, place, logical relationships, and quantification. As mentioned in section 1, NSM researchers have advanced a set of clear and testable hypotheses about the mini-grammar of every semantic prime. Their properties encompass grammatical phenomena usually covered under such rubrics as semantic roles and valency properties of predicates, complementation possibilities, conditionals, time and space adjuncts, quantification, and other topics.

Approaching this point from another angle, Goddard (2002: 32-33) argued that because semantic primes and their grammar constitute the irreducible semantic core of any language L: “[a]s we describe, for any language, the L-specific realisation of this core, we are describing what may be called the ‘core grammar’ of that particular language” and that in this way we are taking “a foundational step in the comprehensive grammatical description of that language.” This is not to say that every distinctive and interesting feature of a language will be apparent from how the grammar of semantic primes is realised in that language. For example, sentence-final particles are a distinctive part of Lao grammar but they are not part of the grammar of Lao semantic primes (Enfield 2002: 243); nor would the mechanics of an elaborate gender or noun class system necessarily be apparent from the “stripped down” grammar of semantic primes. The point remains, however, that getting to know the NSM lexicogrammar of any given language is likely to yield a great deal of insight into the main grammatical workings of that language. It can also be for the linguist a practical first step towards gaining a basic speaking capacity in the local language (cf. Goddard and Wierzbicka 2007).

As a practical aid to fieldworkers, we provide as an Appendix a list of 150 “canonical sentences” that can be useful for identifying semantic primes and for sampling the core lexicogrammar of a language.

3.2 NSM as an aid to communication with consultants and a medium of empowerment

Many fieldwork guides emphasise the importance of communication and meaningful engagement with language consultants. For example, Mosel (2012) says that: “The success of a fieldwork trip very much depends on how the researchers interact with their indigenous consultants. The more the consultants understand the aims and the methods of the research project, the more they will be interested in cooperation and contribute good data”. The same author calls “the possibility of training native speakers as consultants” an “important aspect of research”, while noting that most field guides treat it only in passing or ignore it altogether. Some of the papers in Austin (ed., 2004) deal with “capacity building in endangered language communities”, though chiefly in connection with training local consultants in making recordings and transcriptions, and in other aspects of basic language documentation. Collaboration between speakers and linguists is a major theme of the Grenoble and Furbee (eds. 2010) collection. Hellwig (2010) has illustrated how a linguist’s own language, such as English, German or Spanish, can “spin” his or her interpretation of the categories and meanings of a field language. Similar distorting effects, Hellwig argues, can be caused by a local *lingua franca*.

What we find lacking in all these discussions is anything about how to talk more effectively with consultants about meaning. We want to offer a very simple idea: that the metalanguage of semantic primes can be the basis for a way of talking with consultants about meanings in their own language.

The approach has been used in lexical semantic analysis by Jean Harkins (2001) in her work on emotion terminology in Arrernte (Central Australia), by Marie-Odile Junker (2003, 2007, 2008; also Junker and Blacksmith 2006) in various lexicographic and grammatical projects in East Cree, by Carol Priestley (2002, 2008, 2012a, 2012b) in work on Koromu (PNG), and in a more low-key fashion by various NSM-influenced fieldworkers in Southeast Asia and elsewhere.

Harkins (2001) worked with a group of Arrernte women on a study into Arrernte emotion words. They first identified a wide selection of related words, then grouped them into batches with similar meanings, then focussed on contrastive examination of the most common words in several of the groups. At least 20 examples were collected and/or elicited for each word, showing its use across a range of natural utterance contexts. The collective process of discussion and explanation was conducted substantially in Arrernte, with the end result that the speakers “formulated statements of meaning in their own language, without proceeding via English” (p.199). For each emotion word, they developed a “*ayeye akweke angketye*

nhenhe-ke ‘story small word this-DAT’, that is, a little story, script or mini-narrative” (p.206). Harkins observed that this process was a natural one because it corresponded to Arrernte perceptions of how emotions arise out of actor construals in particular social scenarios.

Harkins noted that these script-like Arrernte explanations showed “a surprising degree of convergence with the kinds of semantic explications produced through more formal NSM analyses”. The following is an example, for the base word *arnkelye* (which Harkins also describes as “a prolonged state of brooding resentment, fuming over some significant offence, and carrying a real danger of erupting into violence”, p204). The Arrernte consultants themselves chose to use X and Y as “false initials” to represent the hypothetical participants in the scenario “for reasons of anonymity” (Harkins 2001: 207). As well as unpacking the meaning content in a highly explicit fashion, the explication establishes that the Arrernte concept is a unitary one, i.e. there is no need to posit polysemy even though in “normal” translation *arnkelye* (and derivatives) could be rendered by several different English words in different contexts of use.

Harkins’ (2001) explication for Arrernte *arnkelye* (roughly, “resentful, brooding”)

[Note 7]

X arnkelye-irre-me:

X re awelhe-me-le itirre-ke Y-ke alakenhe:

X 3sg feel-NPP-SS think-PC Y-DAT thus

“*Y-le akenge mpwarre-ke-nge*

Y-ERG rotten make/do-PC-DS

the akurne ile-tyehnge Y renhe ikwere-nge”

1sA bad do-SBSQ Y 3sO 2SDAT-ABL

X alakenhe itirre-me-le, X akurne awlhe-me Y-ke

X thus think-NPP-SS X bad feel-NPP Y-DAT

‘X is feeling something because X thought like this about Y:

“Y did something very bad

I will do something bad to Y because of this”

when X is thinking like this, X feels something bad towards Y’

Harkins (2001) concluded her article as follows:

[This study] has demonstrated something of the practicality of the NSM approach in facilitating intercultural discussion and understanding of people's emotional life and behaviour in cultural context, and, furthermore, as a tool for stating meanings in the language of inquiry. This is a practical tool that fieldworkers and research participants can use in their own language and cultural setting – a far cry from some popular notions of semantics as an abstract, ivory-tower activity. (Harkins 2001: 211-212)

Junker's (2003, 2007, 2008) on-going work using Cree-based NSM with monolingual and bilingual consultants has involved extensive lexicography, as well as grammatical semantics. She comments on her experience as follows:

Identifying the exponents of semantic primes in a poorly described language has an undeniable heuristic value. Focusing on the subset of the language represented by the semantic primes allows for in-depth grammatical and lexical analysis, leading to discoveries that can then be tested against and extended further to other parts of the grammar. The East Cree version of NSM can then enrich lexical and grammatical analysis by providing tools to discuss lexical properties and grammatical description with speakers. (Junker 2008: 197)

We will see one specific example from the East Cree project shortly, in section 4.1.

Needless to say, field situations are different from one another in many ways. In some local traditions it may be difficult to talk about what someone means (i.e. wants to say) by using certain words, even in relation to real-world or highly contextualised examples (Foley 2003). Nevertheless, we believe that in many cultural contexts the idea of explaining in simple words is likely to have a foothold in traditional ways of speaking to very young children and/or in ways of using simplified language variants with linguistic outsiders. It is impossible to be prescriptive or definitive about how best to go about it in different situations, except to say that the first step is to identify local exponents of semantic primes, and other relatively basic meanings, and to use them oneself in talking with one's consultants or teachers. In some settings, consultants may be quite receptive to the idea of paraphrase (saying the same thing in other words) as a way of explaining meanings. They may welcome the increased engagement and involvement the paraphrase technique offers them, and be open to explicit training in its use. In other cases, it will fall to the linguist to devise successive

approximations of the meaning (based on consultants' comments and naturalistic observation) and offer these, in whole or part, to consultants for confirmation or correction.

Even if consultants are unwilling or unable to engage in collaborative discussion and/or construction of explications, it is always going to be a more practical proposition to communicate with one's consultants using shared simple words (with equivalents in both languages) than it would be to try to adapt the kind of technical linguistic jargon that many linguists fall back on when talking about meanings among themselves. At the same time, casting questions of meaning into semantic primes (with equivalents in both languages) can help safeguard the analyst against the conceptual biases of his/her own language. We return to this point in Section 5. In the meantime, we want to focus on some problems that can arise when interpreting data obtained from semantic fieldwork.

4. Polysemy, paraphrase and morphology: case studies in semantic methodology

In this section, we look at some case studies in lexical semantics from little-described indigenous languages (Dalabon, East Cree, Pirahã) with a view to illustrating several methodological difficulties, including assuming that shared morphology implies shared meaning (even when the assumed shared meaning cannot be clearly identified) and failing to recognise lexical polysemy (even when there is clear language-internal evidence for it). We argue that reductive paraphrase provides a useful conceptual and practical tool that can help overcome these problems. These particular examples have been chosen, in part, because they raise claims with far-reaching implications. [Note 8]

4.1 Cognitive verbs in Dalabon and East Cree

In the *Oxford Handbook of Linguistic Typology*, Evans (2010: 515) acknowledges that NSM “has gone deepest into the possibilities of setting up a cross-linguistically valid set of basic semantic categories in which all meanings can be stated”, but asserts that languages can have language-specific semantic primes. Citing an earlier study of his own, Evans (2007), he states that in Dalabon, a highly endangered language from Central Arnhem Land:

... *think* and *know* lack specific exponents, and both (along with ‘remember’) derive from a root *beng* whose meaning is close to English *mind*. A derivative *bengkan*, etymologically ‘keep/carry in mind’, covers both ‘know’ and ‘think’, with the exact sense coloured by aspect and context. (Evans 2010: 516)

In the following we will review this conclusion in the light of the field data in Evans (2007). Two points need to be made by way of orientation. First, *beng* is a bound root which does not occur as an independent word; Evans (2007) presents it as \sqrt{beng} and we follow the same convention here. Second, the Evans (2007) study does not contain any statement that the verb *bengkan* “covers both ‘know’ and ‘think’”. Rather, it discusses two related verbs – *bengkan* and *bengdi* – the former being close to ‘know’ and the latter being close to ‘think’. About 25 naturally occurring examples of these two verbs are provided, along with examples of other derivationally related words. We will confine ourselves to *bengkan* and *bengdi*, discussing each in turn. [Note 9]

Evans (2007) states that *bengkan* [lit. \sqrt{beng} -carry] “can refer to stably-maintained knowledge and skills ... can also be used of what might be called ‘maintained memory’ ... is best translated as ‘to have continually in one’s mind’, and can variously be used of maintained knowledge, sustained thought, or lasting memory.” He supplies various examples, noting: “In many contexts the best English translation is ‘know’, which, as a stative verb, naturally aligns with constancy of representation.” On our view – and we advance this as a plausible counter-hypothesis – the primary meaning of *bengkan* is indeed KNOW. In sentences like the following, it seems to us that the free English translation with ‘know’ is not only perfectly appropriate, but superior to ‘keep.in.mind’. In (1), for example, we believe that the Dalabon elder is making a knowledge-claim, not talking about what is continually in his mind. (We retain Evans’ original glossing and translations throughout. The original example number is given after the free translation.) [Note 10]

- (1) *Law ngey-kûn law ngah-dja-bengka-n, Mardayin Yaburdurrwa*
law 1sg-GEN law 1/3-Ass-just-keep.in.mind-PR [ceremony] [ceremony]
and whiteman law ngah-dja- bengka-n. Warhdû-kûn
1/3-As-just-keep.in.mind-PR white.person- GEN
ngah-bengka-n rowk.
1/3-Ass- keep.in.mind-PR all
 ‘I know my law, I know Mardayin, Yaburdurrwa and whiteman’s law. I know all that whitefeller stuff.’ [Example 32]

In (2), we see *bengkan* (seemingly KNOW) being contrasted with *warhwa*, apparently “ignorant” (not-know). To us, the interlinear glosses in terms of ‘have.in.mind’ seem

unnecessarily contorted. No clear reason is given for regarding the putative gloss ‘have continually in one’s mind’ as superior to ‘know’.

- (3) *Yibung biy kinikun, biy, mak bula-bengkan bula-monwo-n*
 3sg person other person NEG 3pl/3-keep.in.mind-PR 3pl/3-make-PR
kanunh morlu, bulah-warhwa-n
 DEM didgeridoo 3pl/3-not.have.in.mind-PR
 ‘Other people don’t know how to make didgeridoos, they’re ignorant.’ [Example 3]

In some examples, Evans (2007) renders *bengkan* in free translation as ‘remember’, but in these examples *bengkan* is being contrasted with *bengmukan* “forget”. Arguably, it is this contrast that creates the possibility of using ‘remember’ as a translation. The translations can be re-phrased with ‘know’ and still make perfect sense, e.g. for (4): ‘I still know his face. I’ll never forget it’; for (5) ‘I didn’t forget about my sister, I knew that she was coming’. There is therefore no apparent need to posit any distinct “remember” sense for *bengkan*. Furthermore, Evans (2007) supplies no examples of *bengkan* as ‘remember’ in context such as “I tried to remember her face but I couldn’t” or “Remember to bring your keys”. On the contrary, in such cases, the verb that is used is *bengdi*, which we claim means THINK. Overall Evans (2007) presents 17 examples of *bengkan*, and, with minor qualifications about two ambiguous cases [Note 11], KNOW seems to fit them all perfectly.

- (4) *Nunh ngah-dja-bengka-n, mak nga-bengmukm-iyen,*
 DEM 1/3-just-keep.in.mind-PR NEG 1/3-forget-F
mak nga-bengmukm-û.
 NEG 1/3-forget-PR
 ‘I still remember his face, I’ll never forget it.’ [Example 41]

- (5) *Mak nga-bengmukm-i yabok-ngan-kun ngah-dja-bengka-ng,*
 NEG 1-forget-IRR sister-1sgPoss-DAT 1/3-just-keep.in.mind-PP
kah-burlhm-inj, kah-marnu-burlhm-inj wadda-kah.
 3-appear-PP 3/1-BEN-appear-PP place-LOC
 ‘I didn’t forget about my sister, I remembered [i.e. I held/kept it in my thoughts] that she was coming, that she was coming to my place.’ [Example 42]

Coming now to *bengdi* [lit. \sqrt{beng} -stand]. Evans (2007) states: “This is used for a range of situations all characterised by conscious attention, of having a thought or experience temporarily in one’s conscious mind, or ‘at the top of one’s mind’.” Our hypothesis is that *bengdi* is polysemous, with the primary meaning THINK. A clear example is:

- (6) *Ngah-marnu-bengdi* *nah-ngan* *kah-moyh-boyenj*,
 1/3-BEN-have.in.mindPR mother-1sgPoss 3-sick-big
 ‘I’m thinking about my mother, she’s really sick, ...’ [Partial example, 26]

Evans (2007) states: “This verb is often used in its past perfective form, which focusses on the transition from one mental state (not having in mind) to another (having in mind), to mean ‘remember’, in the sense of remembered information becoming accessible again.” [Note 12] We see no reason to posit the meaning ‘remember’ as such, because a simpler translation with ‘think’ is available. [Note 13] For example, for (7), “in a while, I’ll think where it is, it’ll come back to me”.

- (7) *Bad* *kardu* *mak* *nga-bengka-n*, *manjyelek*,
 money maybe NEG 1/3-carry.in.mind-PR wait.now
ngah-dja-bengda-ngiyan, *ngah-kodj-da-ngiyan*, *wohkardu* *marruh*
 1/3-just-have.in.mind-F 1-head-stand-F or.maybe somewhere
nga-yu-nj, *kardu* *bu* *kenbo* *ngah-dja-ngalk-iyen*.
 1/3-put-PP maybe SUB later 1/3-just-find-F
 ‘Maybe I can’t think right now where the money is, but in a while I’ll remember/recall where it is, it’ll come back to me, where I put it, and then maybe I’ll find it.’ [Example 12]

Bengdi also has a second meaning, roughly ‘decide’ or ‘think to do’: “In non-transition aspects, this verb can also be used to denote situations of ‘deciding, putting one’s mind to’” (Evans 2007). Polysemy between semantic prime THINK and a complex meaning something like ‘decide’~‘to think to do something’, is known to be cross-linguistically common (Goddard 2003).

To summarise to this point, examination of the examples in Evans (2007) shows that his claim that Dalabon lacks “specific exponents” of KNOW and THINK is open to question. A

strong case can be made that the exponents are, respectively, *bengkan* KNOW and *bengdi* THINK. Evans (2007, 2010) has another axe to grind, however; namely, the idea that \sqrt{beng} is a Dalabon-specific semantic prime. According to him, even if *bengkan* and *bengdi* have meanings close to English ‘know’ and ‘think’, at a deeper level the semantics of the two languages are profoundly different because *bengkan* and *bengdi* are derived – semantically, as well as morphologically – from Dalabon-specific \sqrt{beng} . In support of this contention, Evans (2007: 92–94) adduces hypothetical Dalabon paraphrases for *bengdi* and *bengkan* (constructed without direct input from Dalabon speakers). We will review and assess these here, not only because they are important proposals in their own right, but also because the exercise is instructive in pointing to several ways in which a putative paraphrase can go wrong. The proposed Dalabon paraphrase for *bengdi* is given below. The essential parts are line 1 and line 3, but when looked at closely, both are problematical. [Note 14]

Evans’ (2007) proposed paraphrase for Dalabon *bengdi*

(*Ngəy ngah-*)*bengdi*

- | | | | | | |
|----|-------------------|-----------------------|------------------------|--------------|------------------------------|
| 1. | <i>Dubmi-yah</i> | <i>manjh-kerninjh</i> | <i>kah-burlhmû</i> | | <i>kanûm-ngan-kah</i> |
| | now-just | something | 3-Ass-come.upPRES | | ear/mind-1sgPoss-LOC |
| 2. | <i>Kardû</i> | <i>kahnûnda</i> | <i>ngah-burlhka,</i> | <i>kardû</i> | <i>kah-dja-marnû-burlhmû</i> |
| | maybe | this | 1/3-bring.upPR | maybe | 3sg-just-BEN-come.upPR |
| 3. | <i>Ngah-wonan</i> | <i>manjhkeninjh</i> | <i>(kodj-ngan-kah)</i> | | <i>kanunh-yih.</i> |
| | 1/3-sense-PRES | something | (head-1sgPoss-LOC) | | this-INSTR |

1. At this time something is happening in my *beng/kanûm*
2. (I can make this something happen, or it can just happen to me)
3. I am seeing or hearing or feeling something (in my head) because of that

Line 1 depends on the assumption that the free form *kanûm* ‘ear/mind’ is an exact semantic equivalent of the putative Dalabon-specific prime \sqrt{beng} . This amounts to claiming that speakers will or would be unable to paraphrase *kanûm*. This claim remains to be tested and until it is tested, the viability of the paraphrase as a whole remains in the balance. Line 3 depends on the Dalabon verb *wonan*, which appears in the interlinear gloss as ‘sense’ and in the translation as the three-way disjunction ‘seeing or hearing or feeling’. Obviously a great deal hinges on this *wonan*, but what exactly does it mean? Evans says (p. 93) that it normally

means ‘hear’ or, less commonly, ‘feel’ and that he has recorded one example in which it is used about ‘noticing a visual clue’, on which basis he is prepared to speculate that it may have “a range closer to the verb *sense*”. In our view, it is hardly satisfactory that the claimed definability of *bengdi* rests on the poorly understood verb *wonan*. [Note 15] A further, important, difficulty is that it is unclear how the proposed paraphrase for *bengdi* can be substituted across the word’s attested range of use. How, for example, would it work with examples like (6) (‘I’m thinking about my mother’) or (7) (‘I can’t think where the money is’)? Moreover, thinking about someone does not necessarily involve seeing, hearing, or feeling, i.e. a quasi-sensory experience. In short, the proposed explication raises as many questions as it might appear to answer.

Evans (2007: 92) also offers a problematical “possible definition” of *bengkan* (with line numbers added for convenience).

Evans’ (2007) proposed paraphrase for Dalabon *bengkan*

1. *Ngahdjayidnjan manjkerninj kanûmngankah munguyh*
2. *Kanunhyih ngahwernhbengdi*
3. *Kanunhyih, bu ngayedjare,*
4. *ngahwernhwonawonan wohkardu ngahwernhyin menmunguhdjam*

1. I keep something in my *kanûm* (=beng) for a long time
2. Because of this I can “*bengdi*” properly
3. Because of this, when I want to,
4. I can experience things properly or I can do all sorts of things properly

As before, this definition invites queries of its own. It depends not only on *kanûm* ‘ear/mind’ (line 1) and on *bengdi* (line 2), but also on a Dalabon verb (line 4) glossed as “experience” (*wonawonan*, a derivative of *wonan*). As well, the proposed paraphrase does not appear to be substitutable into attested contexts of use. How would it work with an example such as (4)? It does not seem very plausible that, from a Dalabon point of view, knowing someone’s face is a matter of “experiencing it properly” or being able to “do all sorts of things properly” (not to mention that a disjunctive definition cannot capture a unified concept).

In short, both putative Dalabon paraphrases depend on speculative or questionable assumptions, and they fall short of the basic requirement of a viable paraphrase, i.e. that it be substitutable into the word's attested contexts of use. We are left with the conclusion that *bengkan* and *bengdi* correspond to KNOW and THINK, respectively, just as appears to be case from the Dalabon examples. Even if it were possible to establish something about Dalabon-specific "sublexical semantics" involving \sqrt{beng} , the words *bengkan* and *bengdi* would remain as serviceable equivalents of the lexical universals KNOW and THINK for the purposes of cross-linguistic semantics. Moreover, if Dalabon speakers were to engage in Dalabon-internal paraphrase of other cognitive verbs in their own language, such as *bengmukmû* 'have forgotten, not be able to recall', *kodj-mayahmû* 'be confused, be lost, mixed up', and *warhwa* 'be ignorant, unaware of, not know', they would presumably need to employ *bengkan* and *bengdi* as key elements in their own language of reductive paraphrase.

If *bengkan* KNOW and *bengdi* THINK are both semantically unanalyzable, it follows that the shared morphological element, namely, the bound root \sqrt{beng} , is not independently meaning-bearing in these words, i.e. there is a mismatch between the semantics and the morphology. The qualification 'in these words' is important, for it is perfectly possible that \sqrt{beng} is independently meaningful in some of the other words in which it appears. To broaden the discussion, we will in a moment compare the Dalabon situation with a similar situation in another polysynthetic language that has been studied specifically using NSM methodology, namely, East Cree (Junker 2003, 2007, 2008).

The general issue here is of high relevance to all fieldworkers: namely, the extent to which morphological analysis can or should drive semantic analysis; or, to put it in a more cautionary formulation, the dangers of assuming that "shared morphology = shared semantics". Field linguistics is driven in many ways by morphological analysis. Morphological analysis depends heavily on developing a catalogue of lexical and grammatical morphemes and creating bodies of inter-linear glossed sentences and texts. The status of inter-linear glosses, especially for lexical items, depends greatly on the practice of the linguist. When each word receives only a single standard gloss, as frequently occurs, the gloss is effectively just an identifying label for the lexeme, but it is all too easy to fall into thinking that each gloss represents a "meaning". This would be a dangerous assumption for various reasons (cf. Mosel 2006: 50). In particular, it may discourage the linguist from thinking about polysemy.

Bound roots pose additional problems. It is difficult to discuss their meanings with consultants, which increases the analytical burden on the linguist. Bound roots, furthermore, may well have no clear or uniform meanings. A field linguist working on English, for example, would encounter prefixes such as *con-*, *de-*, and *per-*, and bound roots such as *-ceive* and *-sist*, in words like *conceive*, *deceive*, *perceive*, *consist*, *desist*, *persist*, etc., and it would be hard to know what to make of it. In approaches to semantics that feel free to posit highly abstract meanings, not testable by paraphrase, it would be possible to propose some invariant abstract meanings. The NSM approach to semantics puts a constraint on the positing of abstract meanings, while at the same time providing (via the paraphrase technique) a practical procedure for testing the extent to which semantics mirrors morphology in particular cases. Sometimes the alignment is close, sometimes partial, sometimes non-existent; and, moreover, it does not have to be consistent. Marie-Odile Junker's (2003) study of mental predicates in East Cree illustrates this well, and, in our opinion, furnishes an instructive counterpoint to Evans (2007) on Dalabon.

East Cree is an Algonquian language spoken in Northern Quebec, Canada. Like Dalabon, it is a polysynthetic and head-marking, but unlike Dalabon it has thousands of speakers. Junker has been involved, together with a team of dedicated East Cree co-workers, in documenting and developing teaching resources over many years.

East Cree has a verb element *-eyi-*, termed a classifier, that is found in a large number of cognition words: "all thinking, most knowing, all wanting, and several feeling verbs" (Junker 2003: 167). It evidently indicates, in some sense, "mental activity". In fact, the basic verb stem that means THINK is actually made up of a relative root *it-* ('thus', 'so', 'like this'), the element *-eyi-*, and a final *-ht-* or *-m-*, which normally indicates "involvement of the face or mouth". Junker (2003) and her team investigated more than 70 stems containing *-eyi-*. They looked into the meaning of the other elements in the stems and tried to find if speakers saw them as carrying any separable meaning. They identified related verbs without *-eyi-* and tried to establish the meaning differences and similarities, and they encouraged speakers to paraphrase the meanings in their own language (Junker 2003: 173-180).

Not surprisingly, Junker found that most of the complex verbs could be paraphrased by native speakers using Cree THINK verbs. Some of the complex verbs followed a transparent compositional pattern, e.g. "be.lazy-think" for 'to think someone is lazy'. In other cases the meaning of the complex verb was recognised by speakers as metaphorical, e.g. "sweep-think" for 'forgive', "turn-think" for 'change one's mind', "bad-bad-think" for 'be jealous', "down-

think” for ‘die of grief/sorrow’. As indicated by the last two examples, some of these verbs involve other elements as well as THINK, especially FEEL, but also HAPPEN, GOOD, BAD, and others.

Significantly, however, Junker (2003: 180) found that not all the *-eyi-* verbs in Cree could be semantically decomposed. In particular, there are “two verbs where speakers do recognise the initial component, but which they refuse to decompose”. The first one is the Cree equivalent of WANT. It contains a recognizable initial *nituw-*, which is found in several hunting verbs and other “verbs of exploration”. The second is the equivalent to KNOW, whose initial element *chisch-* is found in various non-*eyi-* words, such as ‘prophecy’, ‘remind’, ‘learn’ and ‘lead’. In short, despite the fact that the great majority of *-eyi-* words are semantically analyzable (to East Cree speakers), some are not; and among the semantically unanalyzable ones are the East Cree equivalents of THINK, WANT, and KNOW (each verb is shown in two variants, depending on transitivity (Junker 2003: 181; cf. Junker 2008):

iteyihtam/iteyimeu THINK

nituweyihtam/nituweyimeu WANT

chischeyihtam/chischeyi-eu KNOW

In broad terms, we see the East Cree situation as similar to that of Dalabon. The point is that a bound form (Dalabon \sqrt{beng} , East Cree *-eyi-*) can occur in numerous mental activity verbs and that some of these verbs can be analyzable, to some extent at least, in a way that follows their morphological makeup – without this necessarily implying that every single verb containing the bound form is semantically analyzable. The litmus test of whether a meaning is semantically analyzable is whether a coherent substitutable paraphrase can be devised.

As mentioned at several points, paraphrase also provides the test needed to decide whether or not a word is polysemous. We see a particularly instructive example of this in relation to the Amazonian language Pirahã.

4.2 Polysemy, paraphrase and words for ‘one’ and ‘two’ in Pirahã

Several studies (Gordon 2004; Everett 2005, 2008; Frank et al. 2008) have concluded that Pirahã lacks words for ONE and TWO, meanings that NSM researchers have long regarded as semantic primes. From a methodological angle, an interesting aspect of the debate about

Pirahã is the use of experiments in the field, first by Gordon (2004) and later by Frank et al. (2008). From an analytical point of view, the interesting thing is that although the three major players – Gordon, Everett and Frank – agree that Pirahã does not have words for ONE and TWO, they do not agree about what the would-be equivalents – namely, *hói* “one” (with falling tone) and *hoi* “two” (with rising tone) – actually do mean. We argue below that in fact none of their proposals is coherent and that none fits the attested range of use of the Pirahã words. The solution that really fits the facts, we argue, depends on recognizing that the Pirahã words are polysemous, each having two or more distinct and clearly storable meanings (Wierzbicka 2012). The exercise shows that rigorous application of the paraphrase technique can sharpen and focus semantic description.

Gordon’s (2004) article opens with the sentence “Most of the Pirahã tribe use a ‘one-two-many’ system of counting”. The words in question are *hói* “one”, *hoi* “two”, and *baagi* or *aibai* “many”. However, Gordon hastens to point out that there are complications: “the word for ‘one’ was sometimes used to denote just a small quantity such as two or three or sometimes more (...) Most of the time, in the enumeration task, *hói* referred to ‘one’, but not always”. We will state immediately our own hypothesis about the explanation for these inconsistencies, based not solely on Gordon’s material but on the full range of available evidence. It is that the word *hói* has three distinct meanings: it can mean either ONE, or LITTLE~FEW, i.e. “small quantity”, or, in still other contexts, SMALL.

Gordon’s interpretation was that *hói* has a single meaning – namely, ‘roughly one’, and that “the distinction between discrete and continuous quantification is quite fuzzy in the Pirahã language” (Gordon 2004: 497). But if *hói* can refer to “two or three or sometimes more” (2004: 497), how can it possibly mean ‘roughly one’? The gloss ‘roughly one’ just doesn’t fit with three or more items. Furthermore, testing undertaken later by Frank et al. (2008) showed that in some contexts, speakers used *hói* about as many as 6 items. We will come to this data shortly.

Everett (2005: 623) presented a different interpretation, but again without considering the possibility of polysemy: “*hói* ‘small size or amount’, *hoi* ‘somewhat larger size or amount and *báagi* ... (loosely, ‘many’)”. One of his key examples is given below (with his glosses and translation). In this context an interpretation of *hói* as vague may seem plausible enough.

- (8) *tí* *’ítí’isi* *hói* *hii* *’aba’áigio* *’oogabagai*
 I fish small predicate only want
 “I only want [one/a couple/a small] fish.”

Everett comments that, except as a joke, *hói* could not indicate one very large fish. This might seem to rule out the possibility that *hói* can have the meaning ONE “pure and simple”, as it were. The effect to which Everett refers, however, is presumably induced by the presence in the sentence of the word *'aba'áigio* ‘only’. Saying ‘I want only one fish’ would have to be jocular, in the context of a shared meal, if the fishes were very large.

In other contexts, *hói* must be interpreted as ‘one’, and an interpretation as ‘small size or amount’ would not make sense. We see one such context in an anecdote from Everett’s (2008: 195) book, in which he describes how he was eliciting words for various body parts. The conversation, originally in Pirahã, went like this:

“What is this?” I asked, pointing at my nose.

“*Xitaooi*.”

“*Xitaooi*,” I repeated, perfectly I thought.

“*Xaio, xitaopai*,” he said.

Aargh, I thought. What is that *-pai* business doing at the end of the word?

So, naively, I asked, “Why are there two words for nose?”

“There is one word, *xitaopai*,” came the exasperating answer.

“Just *xitaopai*?”

“Right, *xitaooi*,” he said.

As Everett goes on to explain, his confusion was due to the fact that in Pirahã “*-pai* at the end of body part word (...) means something like ‘my own’. So *xitaooi* means just ‘nose’ but *xitaopai* means ‘my own nose’”. What is relevant for our present purpose is that when the Pirahã informant insists that there is only one word for nose, he evidently does not mean ‘roughly one’ (Gordon’s gloss) or ‘small size or amount’ (Everett’s gloss). The only obviously sensible way to interpret him is that he means ONE.

Everett’s data also show that *hói* has a third meaning: SMALL, as in the phrase ‘a small child’, or in the sentence ‘the child is small’. As noted by Gordon (2004), when *hói* is used in the sense of ‘small’, it has its opposite in *agii* ‘big’ (whereas, we might add, when it is used in the sense of ‘little~few’, it clearly has its opposite in *aibaagi* ‘much~many’). We will not pursue this third meaning here; for discussion, see Wierzbicka (2012).

Frank et al. (2008, Abstract) state: “We show that the Pirahã have no linguistic method whatsoever for expressing exact quantity, not even ‘one’”. For some, this conclusion may

carry extra weight on account of being supported by experimental studies in the field. In our opinion, however, the experimental results are open to a very different interpretation. They actually provide strong support for the polysemy analysis that we are advocating. Frank et al. (2008) conducted two experiments, one with ascending (or increasing) quantity elicitation and one with descending (or decreasing) quantity elicitation, finding that “each of the three words [*hói*, *hoí*, and *baágiso* ‘much/many’] was used for a dramatically different range of values in the ascending and descending elicitations”. Their results are summarised in Fig.1 below.

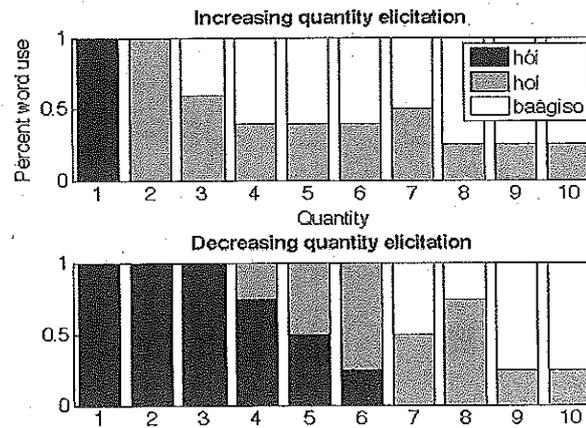


Fig. 1. Proportion of Pirahã speakers using each of the three proposed quantity words in Pirahã. Sets with different quantities were presented in increasing order and participants were asked to describe their quantity.

These results confirm the two different uses of *hói* which we interpret as ONE and LITTLE~FEW (‘small amount’), respectively. As shown in the top diagram of Figure 1, when presented with one spool and asked “how many?”, all the informants responded with *hói* and when subsequently presented with a larger number (from 2 to 10), none of them responded with *hói*. However, as shown in the lower diagram in Figure 1, when they were first presented with 10 spools, and then, successively, with smaller numbers (from 10 to 1) some informants began to respond with *hói* to six, five or four spools, and all of them responded with *hói* to three and two spools. If we assume that *hói* has two meanings – ONE and LITTLE~FEW – this makes perfect sense. Three to six spools cannot be thought of as ONE, but in comparison with ten spools, they can be thought as FEW.

This seemingly obvious interpretation was not available to Frank and colleagues because they did not consider the possibility that words can be polysemous. Looking for a unitary meaning of *hói*, they reasoned that: “these words are much more likely to be relative or comparative terms like ‘few’ or ‘fewer’ than absolute terms like ‘one’ or even pronouns (like ‘roughly one’, as suggested in Gordon 2004)” (Frank et al. 2008: 820). When assessed

carefully against the data, this interpretation ('few' or 'fewer') is no more plausible than Gordon's 'roughly one' or Everett's 'small amount'. When it is used about 4, 5 or 6 spools (going down from 10), one can see how *hói* could conceivably mean 'few' or 'fewer', but how could it possibly mean this when it is used about the **first** spool presented in the increasing elicitation? Even more tellingly, in increasing elicitation why would all the informants use it about one spool, but none of them use it about two spools?

Frank et al. (2008: 823) emphasise that the Pirahã clearly understand the concept of 'one' and that they use it when they perform one-to-one matching tasks set by the experimenters: "[they] understand that the addition or subtraction of exactly one object makes a match incorrect, even for large quantities. Thus, the Pirahã understand the concept of one (in spite of having no word for the concept)". It seems clear from the evidence, however, that the Pirahã do have a word for ONE, namely *hói* (falling tone). The fact that this word also has some other meanings, i.e. LITTLE~FEW and SMALL, may be confusing at first but it is hardly astonishing. Polysemy is rampant in natural languages, especially in high frequency words.

What then of 'two'? What is its status, or non-status, in Pirahã? Gordon (2004: 497) stated that *hoi* (rising tone) "always denoted a larger quantity than the word for 'one' (when used in the same context)", but he did not attempt an explicit gloss. Everett (2005) was more explicit, suggesting 'somewhat larger size or amount' (in comparison, that is, with *hói* (falling tone)). This, however, makes little sense in contexts where *hoi* is used to identify a quantity in absolute, rather than in relative, terms.

Frank et al.'s (2008) data in Figure 1 shows that in increasing elicitation all the Pirahã informants used *hoi* (rising tone) for two spools and that some of them also used it for sets of 3-10 spools. In decreasing elicitation "*hoi* was used for quantities between 4 and 10", though *baagiso* 'many' was strongly preferred for the larger numbers (9 and 10). In our view, the clear message is that *hoi* (rising tone) is polysemous: it can mean TWO and it can mean 'some, not many', i.e. a meaning comparable to that of the English phrase *a few* (not *few*, but *a few*). [Note 16, 17] The fact that *hói* (falling tone) is never used for sets with 7, 8, 9 or 10 objects, whereas *hoi* (rising tone) is sometimes used for all for those (in competition with *baagiso* 'many') indicates that quantity suggested by the second meaning of *hoi* is larger than that suggested by the second meaning of *hói*. This is consistent with our semantic hypothesis: when the number of spools is gradually decreased from 10 ('many'), 8 and 7 can be seen as 'some, not many', but not yet as 'a small quantity' (LITTLE~FEW); whereas 6, 5, and 4 can be regarded as either 'some, not many' or as 'a small quantity' (LITTLE~FEW).

In short, the interpretation that *hói* and *hoí* are polysemous as we have proposed (i.e. *hói* ONE or LITTLE~FEW or SMALL; *hoí* TWO or ‘some, not many’) is an interpretation that makes sense. It is a coherent and viable hypothesis, and it leads to a tighter fit with the available data than the hypotheses advanced by Gordon (2004), Everett (2005, 2008), or Frank et al. (2008).

To wrap up this discussion of Pirahã lexical semantics, we would like to focus attention on three issues. The first is the need for semantic analysts to be open to the possibility of polysemy. Gordon (2004) and Frank et al. (2008) operated on an unexamined assumption of monosemy. Everett (2005, 2008, 2102) appears wedded to monosemy as a matter of principle. In any case, this assumption prevented any clear and coherent hypothesis from emerging. The second is that both Gordon (2004) and Frank et al. (2008) placed excessive reliance on extensionalist data gathering techniques, at the expense of interpretative analysis. Frank et al.’s (2008) experimental procedures were ingenious and yielded valuable results which lent themselves to representation in diagrams. But in order to advance to an explanatory hypothesis, interpretation – conceptual analysis – is needed, and this is where they struggled. The Frank (2008) team clearly perceived the weaknesses in Gordon’s (2004) earlier account, but they were not able to come up with a coherent interpretation of their own. We put this down, in part, to the fact that they had no disciplined analytical framework (metalanguage) into which to cast their interpretation.

Our two final points of a more general nature. The fact that, as we hope to have shown, the Pirahã language does indeed have lexical equivalents for ONE and TWO does not mean that Pirahã speakers engage in “numerical cognition” as it is normally understood in European terms. The Pirahã language does not have numerals (counting words) (Goddard 2009), the Pirahã people do not practise counting or tallying in any form, and on all reports appear to be deeply uninterested in keeping track of precise numerical quantities. From the literature in anthropological linguistics, we doubt that the Pirahã are unique in these respects, but we not doubt that their language and cognition is radically different from “ours”. Radically different is not same as fundamentally different, however, as shown by the existence in Pirahã, as in other languages, of ONE and TWO, and other lexical universals.

Our final point concerns the status and importance of polysemy as a linguistic phenomenon. Some scholars outside linguistics (and even some within linguistics) are inclined to regard polysemy as a technical semantic issue, of marginal relevance to “bigger questions” such as language universals, numerical cognition, and “how the mind works”. We hope to have shown that taking polysemy seriously is crucial to gaining a clear interpretation of linguistic facts and their cognitive implications. In our view, it ought be a source of

concern that some psychologists and other cognitive scientists are rushing to erect theoretical edifices on the shaky foundations of the alleged lack of ONE and TWO in Pirahã.

We now turn to the difficulty of grappling with complex indigenous meanings.

5. Capturing indigenous semantic perspectives without Anglocentrism

Field linguistics, and the whole project of language documentation, is committed to capturing and preserving, as far as possible, indigenous systems of knowledge and cognition. In our opinion, though, linguists are usually not prepared to face up to the simple fact that indigenous perspectives cannot be faithfully portrayed using complex English-specific (often Latinate) vocabulary. To ensure that “endangered concepts” are preserved for a wider audience (as part of the common human heritage), it may be necessary for them to be described in a widely accessible language such as English, i.e. in English words, but it is not necessary for these words to be technical or sophisticated Latinate English. Trying to explicate indigenous concepts using complex and untranslatable English words necessarily imposes an Anglocentric and/or Eurocentric perspective.

5.1 Not putting words into people’s mouths

Emphasising the need for genuine cross-linguistic understanding between linguists and native speakers, Louanna Furbee (2010: 8) argues that language documentation can proceed collaboratively by creating “intermediate models ... that will likely not faithfully reflect the thinking of the linguist nor that of the consultant but will be accessible to both and translatable to both, as well as available for revision”. This can allow “power and responsibility to pass increasingly into the hands of heritage speakers without requiring the enculturation of those speakers to a linguist’s perspective, or the linguist to the native speaker’s perspective ...”. As an example of such an enterprise, Furbee uses the ongoing analysis of evidentials in the Mayan language Tojolabal. Linguists traditionally describe such meanings using complex technical expressions that are untranslatable into indigenous languages: categories such as “direct”, “indirect”, “circumstantial”, “inferential”, and descriptions such as “from a reliable source” or “speaker certifies the truth of the information”. Even the name of the linguistic category ‘evidential’ is based on a deeply Anglo cultural key word, namely, *evidence*, that lacks equivalents in most of the world’s languages, including many European ones (Wierzbicka 2010a). When Furbee’s bilingual consultants were asked to explain the meanings of the different “evidentials” in their own words, not

surprisingly they did not come up with any such technical terms, but rather, with explanations which made sense to them in their own language. Furbee emphasizes the fact that “in that collaborative activity, the team more than tripled the inventory of evidentials previously identified”. From an NSM perspective, what is even more striking is the simplicity and clarity of the language used by the Tojolabal consultants in describing their evidential meanings: glosses such as ‘thus it is’, ‘I say it’s so’, ‘it’s true, always’, ‘I think ... (it might be true, it might not)’, ‘they said it’, ‘perhaps he said it’, ‘they say’, ‘it is said’ and ‘perhaps he said it’ (Furbee’s 2010: 14). The explanatory language of the Tojolabal consultants is very close to the language of universal semantic primes (cf. Wierzbicka 1996: Ch 15).

In general we think that field linguists should strive to find ways to express their hypotheses and insights about indigenous meanings in terms that are simple and cross-translatable, even if this comes at the cost of their being lengthier and less sophisticated in tone.

We will illustrate with two further examples. It is often assumed that in the spatial domain it is a straightforward matter to move from extensional data, such as audiovisual recordings, maps, and photographs, to an authentic intensional (cognitive) representation. Evans and Sasse (2007) argue against this view, insisting that meaning is a matter of construal and hence demands a hermeneutic process. Their point is an important one and we are in full agreement with them. We would not agree, however, that it is appropriate to render an indigenous meaning (even “roughly”) using a complex English expression such as “significant geographical discontinuity”, as they do in the following passage:

[I]n Kayardild (Evans 1995), there is a suffix *-ngurrnga*, added to directional terms based on compass points, meaning roughly ‘beyond a significant geographical discontinuity’. Thus *ringurrnga* [east-*ngurrnga*] can be used for an island, emerging from the sea to the east, but also to a stand of mangroves, or sandhills, as one leaves a saltpan travelling east. Though it is possible to photograph or map particular instances of *ringurrnga*, the underlying meaning – what is construed as a significant geographical discontinuity – cannot simply be taken for granted, or immediately inferred from a couple of instances, and to really get to the bottom of it we need to probe an open-ended set of examples. (Evans and Sasse 2007: 67)

In our view it would be preferable to see the meaning expressed by the Kayardild suffix *-ngurrnga* unpacked into terms that can be expressed in Kayardild itself. It seems to us that

this can be done using a scenario-based approach, roughly as follows: someone in a place wants to be in another place, on one side (e.g. to the east or south) of the place where he or she is to begin with; this someone knows that to reach this other place, he or she first has to move through some other places; and this someone also knows that this can't be done without making special efforts (e.g. swimming, canoeing, climbing). This is not a full or ideal explanation, but it goes some way to breaking down the meaning in a fashion that could plausibly correspond to the thinking of Kayardild speakers themselves.

For another example, we will attempt to re-think two of the glosses used by Hinton (2001b) in her article about directional suffixes in the endangered American indigenous language Karuk. The two suffixes and their glosses are: *-vara* 'in through a tubular space' and *-kiv* 'out through a tubular space'. Glosses like these (descended from Bright 1957; cf. Macaulay 2004) are common in the literature on complex directional morphology, but they are peculiar and obscure even in English. What is a "tubular space"? How can such a gloss capture an indigenous Karuk meaning if (as we assume to be the case), Karuk does not have words like "tubular" and "space"? We suggest that a plausible unpacking of the intended meaning is based on a motional scenario, namely, the scenario of something moving for some time 'inside something long'. For *-vara*, the result of such movement is that the moving thing ends up inside something else; for *-kiv*, it ends up no longer being inside something else. Again, these ideas about the Karuk meanings may be incomplete or mistaken in some respects, but at least they are formulated in a way that can plausibly represent indigenous Karuk meanings.

5.2 The challenge of cultural key words

Finally, we want to draw attention to the importance of identifying and documenting the cultural key words of endangered languages. The term "key words" refers to particularly culture-rich and translation-resistant words that occupy focal points in cultural ways of thinking, acting, feeling, and speaking (Wierzbicka 1997, 2006, in press). Examples from well described languages include: English *fairness* and *mind*, Russian *duša* (roughly "soul") and *sud'ba* (roughly "fate, destiny"), Japanese *wa* (roughly "harmony, unity") and *omoiyari* (roughly, "empathy"), Malay *setia* (roughly, "loyal") and *hati* (roughly "heart, mind"), Danish *hygge* (roughly "cosy sociality") and *tryk* (roughly, "security, safety") (Wierzbicka 1997, 2011a; Goddard 2001a, 2001b; Levisen 2012). The concept of a cultural key word is a qualitative one and somewhat inexact in the sense that it is not always possible to draw a strict

line between cultural key words, other culturally important words, less important but still culture-related words, and so on. But that is not the point. The point is that every culture has some highly distinctive, even unique, cultural concepts and it is incumbent upon fieldworkers to give them due attention. Given the emphasis placed on preserving unique indigenous ways of thinking, it is unfortunate that existing linguistic fieldwork manuals pay scant attention to the importance of locating and exploring culturally important words, or to the challenges involved in doing so. (As one reviewer remarked, the explanation no doubt lies in the fact that most field manuals are focused on grammar. The same reviewer feels that “the issue of cultural keywords is more comfortably the province of anthropological linguistics than ‘field’ linguistics”.)

Cultural key words may be found in different departments of the lexicon (sometimes in unexpected places), but we would nominate as priority areas: words for cultural values and ideals, other “ethnophilosophical” terms, words for emotions and for the psychological makeup of persons (“heart”, “mind”, “spirit” and the like), and distinctive social categories. [Note 18] Such words can be of great interest to indigenous speakers themselves, who often see their language “as the carrier of traditional culture and values”, as Hinton (2001a: 16) observes in *The Green Book of Language Revitalisation*. A typical statement would be the one quoted by Warner (2001: 138) from a parent of two students in a Hawaiian immersion school, saying (in Hawaiian) that: “there is a critical need to preserve and restore the Hawaiian language – a natural treasure and resource – with its heart and spirit intact”.

We would be first to acknowledge that unravelling the semantics of a cultural keyword is a formidable challenge and it may be necessary to settle for an approximation at first. Still, there are some guidelines that can be helpful. Chief among them, in our opinion, should be the goal of producing a formulation that is phrased as simply as possible (not necessarily exclusively in semantic primes) and that relies as little as possible on disjunction. The problem with an excess of disjunctions can be illustrated with the following quotation about the Hawaiian cultural key concept *mauli*, which would evidently be highly relevant to Hawaiian “heart and spirit”. Wilson and Kamanā (2001) describe it as follows:

Some features of *mauli* are covered by the English word “culture”, but *mauli* also includes worldview, spirituality, physical movement, morality, personal relationships, and other central features of a person’s life and the life of a people. (Wilson and Kamanā 2001:61)

This statement is useful as an indication of the cultural centrality of the *mauli* concept, but the string of abstract English words (“culture”, “worldview”, “spirituality”, “morality”, etc.) is too vague and diffuse to convey any sense of a powerful and unitary culture-specific meaning. The same applies to descriptions of a more celebrated example of a cultural key concept from Oceania, namely, *mana*, which has been variously glossed by anthropologists as “power, authority, influence, prestige, efficacy” (cf. Keesing 1984). The Māori linguist Haumihiata Mason (2008: 35) has even rendered *mana* as “identity” and “integrity” in her poetic motto: *Tōku reo, tōku mana, Ki te ngaro tōku reo, ka ngaro ko au* ‘My language is my identity, my integrity, Without my language I am lost.’ The wide range of these glosses highlights the fact that they all fail to capture the uniqueness of the Māori concept.

We are not in a position to offer any suggestions about Hawaiaian *mauli*, but Wierzbicka (2014) proposes the following explication for *mana* in Māori (cf. Moorefield 2005: 76-77):

mana:

something

people can't see it, people can't know well what it is

it can be part of some things, it can be part of some people

if someone has this part, this someone can do many things not like other people

when people think about this something, they can't not feel something good

This explication is framed exclusively in semantic primes; thus, all its terms are independent of English, transposable into Māori, and can plausibly represent part of the traditional Māori worldview. In more conventional English, a similar idea could be captured by a paraphrase as follows: “a mysterious invisible power that can be a part of certain special things and certain special people”. The latter paraphrase includes some complex English words (especially ‘power’), but it is a useful approximation: more intelligible (more “focussed”) than disjunctive explanations and capable of conveying a sense of the unique conceptualisation behind the *mana* concept.

In our view, the rich NSM literature on cultural key words in a variety of languages shows that they can be successfully unpacked into explanatory paraphrases that make sense to cultural insiders and outsiders alike, and this should be the ultimate goal for every endangered language. In the meantime, we would encourage fieldworkers to make a point of noting culturally important words (candidates for key word status) and recording copious examples of their use, including consultants’ commentaries and explanations. Efforts should be made,

however preliminary, to capture aspects of the meaning in simply phrased semantic components, avoiding untranslatable Latinate vocabulary in the explanations. At a practical organisational level, we propose that the conventional dictionary and domain-based (thesaurus-like) structure for wordlists be augmented by a special section dedicated to culturally important concepts (without regard to part of speech, lexical field, etc.). Admittedly, such sections could be potentially “messy” and inconclusive, but if lexical documentation is intended to capture (among other things) unique “knowledge systems”, we cannot neglect the task of identifying the main cultural ideas embedded in the lexicon.

In summary, we see that there is much to be gained, and very little to be lost, if field linguists were to cut loose from complex academic English and start trying to capture their emerging understandings of indigenous meanings using a metalanguage based on simple translatable words.

6. Concluding remarks

As Patrick Hanks (2007: 6) has observed, “There is no metalanguage other than words themselves (in one language or another) – or at a pinch derivatives of words such as logical symbols – for expressing thoughts about words.” Or, in a pithier formulation (sometimes called Sweeney’s Law): “You gotta use words to talk”. Uniquely in contemporary semantics, the NSM approach has sought to make a virtue of necessity by basing its method of semantic representation as closely as possible on the words of natural language. Also uniquely in contemporary semantics, the NSM approach is explicitly anti-Anglocentric and anti-Eurocentric.

In this article we have argued that the NSM technique of reductive paraphrase into simple, shared word-meanings can help cut through many of the confounds and conundrums that trouble field linguistics. Notwithstanding that any utterance is embedded, in a sense, in cultural context and can therefore sustain layers and iterations of hermeneutic analysis, we believe that practical semantic and lexicographic analysis need not be an “endless quest” (Evans and Sasse 2007).

Reductive paraphrase into simple shared meanings can also help linguists, both when they are in the field and when they are back in their academic milieu, to steer between the dangers of Eurocentrism (especially Anglocentrism), on the one hand, and exoticisation, on the other. It can help bring the task of lexical semantic analysis down from the clouds of excessive abstraction, and back to a linguistic level at which consultants and linguist can collaborate meaningfully. At the same time, it can free the linguist from the straightjacket of technical

Latinate English which often stands in the way of genuine understanding of how speakers of other languages, including endangered languages, think.

Interlinear abbreviations

ABL ablative, ASS assertative, BEN benefactive (applicative), COM comitative, DAT dative, DEM demonstrative, DS different subject, EMPH emphatic, ERG ergative, F future, GEN genitive (also covers purposive uses), H higher animate (object, only with 3sg subject), INSTR instrumental, LOC locative, NEG negative, NPP non-past progressive, PC past completive, PCUST past customary, PI past imperfective, POSS possessed noun, e.g. 1sg Poss ‘noun possessed by first person singular’, PP past perfective, PR present, PRES present, PURP purposive, SBSQ subsequent, SS same subject, SEQ sequential, sg singular, SUB subordinate, 1/3, etc. 1st person singular subject acting on third person singular object, etc., 3/3 3rd person singular subject acting on third person higher animate, 12 1st person inclusive (first plus second person).

Notes

1: Encyclopedia and handbook entries on NSM include: Goddard (2001c, 2006, 2009, 2010b), Goddard and Wierzbicka (eds. 2002, 2011), Wierzbicka (2008a, 2009). Textbook treatments can be found, inter alia, in: Allan (2001), Löbner (2002), Geeraerts (2010), Murphy (2010), Goddard (2011a).

2: We will not discuss in this article so-called extensional semantics, which concerns itself with establishing the referential range or extension of linguistic expressions, and, in its typological mode, with comparing the referential ranges of comparable expressions across languages (cf. Levinson et al. 2003; Majid and Bowerman eds. 2007; Majid 2012). While we are unimpressed by extensionalist semantics as a theory (or de facto theory) of meaning (Goddard 2011a: 10-11), we have no quarrel with using stimulus tasks as one technique, among others, for obtaining useful data for (intensional) interpretation. See section 4.1.

3: See Goddard (2008) for discussion of the “thematic” grouping of semantic primes employed in Table 1, and for discussion of alternative tabular arrangements.

4: There is no reason to expect exponents of semantic primes to be any less polysemous than other words. *A priori* one would expect, and experience bears this out, that exponents of

primes would exhibit considerable polysemy simply because they are usually fairly common words, and common words are generally more polysemous than less common ones. The growing body of work identifying exponents of primes across languages has made a significant contribution to lexical typology in this regard (Goddard 2012: 716).

5: This and other literature on identification issues, e.g. the general treatments and diverse case studies in Goddard and Wierzbicka (eds. 2002) and Goddard (ed. 2008, 2011), may be relevant for fieldworkers who encounter problems in identifying exponents of semantic primes in their language of interest.

6: The Comrie and Smith (1977) list incorporates all the entries from Swadesh's (1972) 100-item "basic vocabulary" list. Given its original purpose as a set of words hypothesised to be stable over time, it is ironic that the Swadesh wordlist is sometimes recommended (e.g. Crowley 2007: 95) as a starting point for lexical elicitation in the field.

7. In this, and in all subsequent glossed material, the interlinear glosses are those of the original author. Where there are minor inconsistencies in the glossing, we have retained these in the interests of reproducing the published data verbatim.

8: The desirability of language-internal paraphrase as a discovery and decision procedure for semantic primes was emphasised in Goddard (2002). Goddard (1991) is an early example its application to a problem of polysemy in an indigenous language (Yankunytjatjara, Central Australia).

9: The omission of any mention of *bengdi* [lit. \sqrt{beng} -stand] in Evans (2010) is puzzling, because it is actually the main focus of the 2007 study (and the basis for its title: 'Standing up your mind: Remembering in Dalabon').

10: Someone might wish to maintain that the Dalabon elder speaking in (1) is not claiming to **know** Aboriginal law but is saying that it is "continually on his mind"; however, this interpretation would hardly make sense when the elder goes on to say the same about whitefella law.

10: For the sake of completeness, we note that *bengkan* as KNOW may seem not to work very well in a couple of Evans' (2007), e.g. his (42) and (43). For readers who wish to consult the original data, we provide the following comments. In example (42) the locution with *bengkan* could well be interpreted as 'my father's known (= familiar) face'. Example (43) makes sense if one assumes that after a dream the details often fade. In other words, the speaker could mean something like "I saw a face in a dream and after I woke up it was still on my mind (= in my head). I still know [*bengkan*] it, I didn't forget it".

12: The form *bengdayhm-* is a semi-suppletive past perfective of *bengdi* (Evans 2007: 72), even though formally it looks like a decausative of *bengdayhka* 'remind, cause to *bengdi*'.

13: The perfective (inceptive) use of *bengdi* here, as with corresponding uses of English *think (of)*, is not simply semantic prime THINK, but rather involves a combination of THINK and KNOW. Roughly: 'I thought (at that time) where it was' = 'I thought like this at that time: "I know where it is"'. Even so, this meaning is not the same as English *remember*, which is significantly more complex (Wierzbicka 2007c).

14: We will not consider line 2 because it is redundant and also because it is not very plausible to suppose that when a speaker says "I am *bengdi*", he or she is saying anything like 'I can make this something happen, or it can just happen to me'.

15: To further confuse the picture, the Dalabon Dictionary (Evans, Merlin and Tukumba 2004) glosses *wonan* as 'hear' and 'think'. Ponsonnet (2009) discusses a further set of expressions, including *mon-no*, which she glosses as 'intentions', 'views', 'thoughts'.

16: Pica et al. (2004) have glossed a comparable word in another Amazonian language, Mundururu, in exactly this fashion.

17: Polysemies involving TWO are well known from other languages, including English. Thus, the word *couple* has two meanings in English, one as in *a married couple* (two people married to one another) and another as in *a couple of chairs* (could be two, three, perhaps even four). In Polish, *para* can mean either 'pair' (as in 'a pair of shoes') or a 'couple' (as in 'a married couple'), but when used in the accusative (*parę*) it means 'a few'. A similar pattern of polysemy can be found in German (*Paar₁* 'pair', *Paar₂* 'couple', *Paar₃* 'a few'). In Russian,

too, the word *para* means either ‘pair’ (as in ‘a pair of shoes’) or ‘couple’ (as in ‘a married couple’), but in the phrase *paru slov*, it means ‘a few words’ (as in ‘I want to say a few words’). If such polysemies can occur in languages with highly developed systems of numerals, it is hardly surprising that they can also occur in languages like Pirahã.

18: Speakers are not always aware of cultural key words, regarding them as obvious, natural, etc., and they may assume that the nearest comparable words in another language necessarily embody the same meanings as theirs. Examples from English include *story*, *evidence*, and *experience* (Wierzbicka 2010a, 2010b).

References

- Allan, Keith. 2001. *Natural Language Semantics*. Oxford: Blackwell.
- Amberber, Megistu. 2008. Semantic primes in Amharic. In Cliff Goddard (ed.), *Cross-Linguistic Semantics*, 83–119. Amsterdam: John Benjamins.
- Ameka, Felix. 1994. Ewe. In Cliff Goddard and Anna Wierzbicka, (eds.), *Semantic and Lexical Universals: Theory and Empirical Findings*, 57–86. Amsterdam: John Benjamins.
- . 2006. ‘When I die, don’t cry’: the ethnopragmatics of “gratitude” in West African languages. In Cliff Goddard (ed.) *Ethnopragmatics: Understanding Discourse in Cultural Context*, 231–266. Berlin: Mouton de Gruyter
- . 2009. Access rituals in West African communities: An ethnopragmatic perspective. In Gunther Senft and Ellen B. Basso (eds.), *Ritual Communication*, 127–152. NY: Berg.
- Ameka, Felix K., Alan Dench, and Nicholas Evans (eds.). 2006. *Catching Language: The standing challenge of grammar writing*. Berlin: Mouton de Gruyter.
- Asano-Cavanagh, Yuko. 2009. A semantic analysis of Japanese epistemic markers: *chigainai* and *hazuda*. *Language Sciences* 31(5), 837–852.
- . 2010. Semantic analysis of evidential markers in Japanese: *rashii*, *yooda* and *sooda*. *Functions of Language* 17 (2), 153–180.
- Austin, Peter K. (ed.). 2003-2010. *Language Documentation and Description*. Vols. 1-9. London: Hans Rausing Endangered Languages Project, School of Oriental and African Studies.
- Bohnenmeyer, Jürgen. 1998. Temporal reference from a radical pragmatics point of view. Why Yucatec Maya does not need to express ‘after’ and ‘before’. *Cognitive Linguistics* 9(3), 239–282.
- . 2003. NSM without the Strong Lexicalization Hypothesis. *Theoretical Linguistics* 29(3), 223–226.
- Bowern, Claire. 2008. *Linguistic Fieldwork: A practical guide*. New York: Palgrave Macmillan.
- Bromhead, Helen. 2009. *The Reign of Truth and Faith: Epistemic Expressions in 16th and 17th Century English*. Berlin: Mouton de Gruyter.
- . 2011a. Ethnogeographical categories in English and Pitjantjatjara/Yankunytjatjara. *Language Sciences* 33(1), 58–75.
- . 2011b. The bush in Australian English. *Australian Journal of Linguistics* 31, 445–471.
- Brotherson, Anna. 2008. The ethnogeometry of Makasai (East Timor). In Cliff Goddard (ed.), *Cross-Linguistic Semantics*, 259–276. Amsterdam: John Benjamins.

- Bugenhagen, Robert D. 2001. Emotions and the nature of persons in Mbula. In Jean Harkins and Anna Wierzbicka (eds.), *Emotions in Crosslinguistic Perspective*, 69–114. Berlin: Mouton de Gruyter.
- 2002. The syntax of semantic primitives in Mangaaba-Mbula. In Cliff Goddard and Anna Wierzbicka (eds.), *Meaning and Universal Grammar – Theory and Empirical Findings*, Volume I, 1–64. Amsterdam: John Benjamins.
- Chappell, Hilary. 2002. The universal syntax of semantic primes in Mandarin Chinese. In Cliff Goddard and Anna Wierzbicka (eds.), *Meaning and Universal Grammar – Theory and Empirical Findings*, Volume I, 243–322. Amsterdam: John Benjamins.
- Comrie, Bernard and Smith, N. 1977. Lingua Descriptive Series Questionnaire. *Lingua* 42, 1–72.
- Crowley, Terry. 2007. *Field Linguistics: A beginner's guide* (edited and prepared for publication by Nick Thieberger). Oxford: Oxford University Press.
- Dixon, R. M. W. and A.Y. Aikhenvald. 2002. Word: a typological framework. In R.M.W. Dixon & A. Y. Aikhenvald, (eds.), *Word: A Cross-linguistic Typology*, 1–41. Cambridge: Cambridge University Press.
- Durst, Uwe. 1999. BAD as a semantic primitive: Evidence from Biblical Hebrew. *Pragmatics & Cognition* 7(2), 375–403.
- Enfield, N. J. 2002. Combinatoric properties of Natural Semantic Metalanguage expressions in Lao. In Goddard, Cliff and Wierzbicka, Anna (eds.) *Meaning and Universal Grammar – Theory and Empirical Findings*, Vol II, 145–256. Amsterdam: John Benjamins.
- Evans, Nicholas. 2007. Standing up your mind: Remembering in Dalabon. In Mengistu Amberber (ed.), *The Language of Memory in a Cross-Linguistic Perspective*, 67–96. Amsterdam: John Benjamins.
- 2010. Semantic typology. In Jae-Jung Sung (ed.), *The Oxford Handbook of Linguistic Typology*, 504–533. Oxford: Oxford University Press.
- Evans, Nicolas, Francesca Merlin and Maggie Tukumba. 2004. *A first dictionary of Dalabon (Ngalkbon)*. Maningrida: Maningrida Arts and Culture.
- Evans, Nicholas and Hans-Jürgen Sasse. 2007. Searching for meaning in the Library of Babel: Field semantics and problems of digital archiving. In Peter K. Austin (ed.), *Language Description and Documentation* Vol. 4, 58–99.
- Everett, D. L. 2005. Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology* 46, 621–646.
- 2008. *Don't Sleep. There are Snakes*. London: Profile Books.

- 2012. Understanding others requires adaptive thinking: Response to Wierzbicka. *Pragmatics & Cognition* 20(2), 417–428.
- Foley, William A. 2003. Genre, register and language documentation in literate and preliterate communities. In Peter K. Austin (ed.), *Language Documentation and Description*. Vol. 1, 85–98.
- Frank, Michael C., Daniel L. Everett, Evelina Fedorenko, and Edward Gibson. 2008. Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition* 108, 819–824.
- Furbee, N. Louanna. 2010. Language Documentation: Theory and Practice. In Lenore A. Grenoble and N. Louanna Furbee (eds.), *Language Documentation: Practice and Values*, 3–24. Amsterdam: John Benjamins.
- Geeraerts, Dirk. 2010. *Theories of Lexical Semantics*. Oxford: Oxford University Press.
- Gippert, Jost, Nikolaus P. Himmelmann and Ulrike Mosel (eds.). 2006. *Essentials of language documentation*. Berlin: Mouton de Gruyter.
- Gladkova, Anna. 2007. Universal and language-specific aspects of “propositional attitudes”: Russian vs. English. In Andrea C. Schalley and Drew Khlentzos (eds.), *Mental States: Volume 2: Language and Cognitive Structure*, 61–83. Amsterdam: John Benjamins.
- 2010. *Russkaja kul’turnaja semantika: emocii, cennosti, žiznennye ustanovki* [Russian cultural semantics: Emotions, values, attitudes.] Moscow: Languages of Slavic Cultures.
- Goddard, Cliff. 1991. Testing the translatability of semantic primitives into an Australian Aboriginal Language. *Anthropological Linguistics* 33(1), 31–56.
- 1994. Lexical primitives in Yankunytjatjara. In Cliff Goddard and Anna Wierzbicka (eds.), *Semantic and Lexical Universals – Theory and Empirical Findings*, 229–262. Amsterdam: John Benjamins.
- 2000. Polysemy: A problem of definition. In Yael Ravin and Claudia Leacock (eds.), *Polysemy: Theoretical and Computational Approaches*, 129–151. Oxford: Oxford University Press.
- 2001a. *Sabar, ikhlas, setia* – patient, sincere, loyal? A contrastive semantic study of some “virtues” in Malay and English. *Journal of Pragmatics* 33, 653–681.
- 2001b. *Hati*: A key word in the Malay vocabulary of emotion. In Jean Harkins and Anna Wierzbicka (eds.), *Emotions in Crosslinguistic Perspective*, 171–200. Berlin: Mouton de Gruyter.
- 2001c. Lexico-semantic universals: a critical overview. *Linguistic Typology* 5(1). 1–66.

- 2002. The search for the shared semantic core of all languages. In Cliff Goddard and Anna Wierzbicka (eds.), *Meaning and Universal Grammar – Theory and Empirical Findings*, Vol. I, 5–41. Amsterdam: John Benjamins.
- 2003. Thinking across languages and cultures: six dimensions of variation. *Cognitive Linguistics* 14(2-3), 109–140.
- 2006. Natural Semantic Metalanguage. In Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd edition, 544–551. Oxford: Elsevier..
- 2007. A “lexicographic portrait” of *forgetting*. In Mengistu Amberber (ed.), *The Language of Memory in a Cross-Linguistic Perspective*, 119–137. Amsterdam: John Benjamins.
- 2008. Towards a systematic table of semantic elements. In Cliff Goddard (ed.) *Cross-Linguistic Semantics*, 59–81. Amsterdam: John Benjamins.
- 2009. Cultural scripts. In Gunter Senft, Jan-Ola Östman and Jef Verschueren (eds.), *Culture and Language Use*, 68-80. (Handbook of Pragmatics Highlights, v2.) Amsterdam: John Benjamins.
- 2010a. Semantic molecules and semantic complexity (with special reference to “environmental” molecules). *Review of Cognitive Linguistics* 8(1), 123–155.
- 2010b. The Natural Semantic Metalanguage approach. In Bernd Heine and Heiko Narrog (eds.) *The Oxford Handbook of Linguistic Analysis*, 459–484. Oxford: Oxford University Press.
- 2011a. *Semantic Analysis*. [Revised 2nd edition]. Oxford: Oxford University Press.
- 2011b. The lexical semantics of ‘language’ (with special reference to ‘words’). *Language Sciences* 33(1), 40–57.
- 2012. Semantic primes, semantic molecules, semantic templates: Key concepts in the NSM approach to lexical typology. *Linguistics* 50(3), 711–743.
- Goddard, Cliff (ed.). 2008. *Cross-Linguistic Semantics*. Amsterdam: John Benjamins.
- 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.
- Goddard, Cliff and Anna Wierzbicka. 2007. Semantic primes and cultural scripts in language teaching and intercultural communication. In Farzad Sharifian and Gary Palmer (eds.), *Applied cultural linguistics: Implications for second language learning and intercultural communication*, 105–124. Amsterdam: John Benjamins.

- 2010. ‘Want’ is a semantic and conceptual universal (response to Khanina). *Studies in Language* 34(1), 108–123.
- 2011. Semantics and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science (WIREs Cognitive Science)* 2(2), 125–135.
- 2014. *Words and Meanings: Lexical Semantics across Domains, Languages, and Cultures*. Oxford: Oxford University Press.
- Gordon, Peter. 2004. Numerical cognition without words: Evidence from Amazonia. *Science*, 306, 496–499.
- Grenoble, Lenore A. and N. Louanna Furbee (eds.) 2010. *Language Documentation. Practice and Values*. Amsterdam: John Benjamins.
- Habib, Sandy. 2011a. Contrastive lexical conceptual analysis of folk religious concepts in English, Arabic and Hebrew: NSM approach. PhD thesis, University of New England.
- 2011b. Angels can cross cultural boundaries. *RASK, International Journal of Language and Communication* 34, 49–75.
- Hanks, Patrick W. 2007. General Introduction. In Patrick W. Hanks (ed.), *Lexicology: Critical Concepts in Linguistics*, Vol I, 1–23. Abingdon: Routledge.
- Harkins, Jean. 2001. Talking about anger in Central Australia. In Jean Harkins and Anna Wierzbicka (eds.), *Emotions in Crosslinguistic Perspective*, 201–220. Berlin: Mouton de Gruyter.
- Harkins, Jean and David P. Wilkins 1994. Mparntwe Arrernte and the search for lexical universals. In Cliff Goddard and Anna Wierzbicka (eds.), *Semantic and Lexical Universals – Theory and Empirical Findings*, 285–310. Amsterdam: John Benjamins.
- Hasada, Rie. 2001. Meanings of Japanese sound-symbolic emotion words. In Jean Harkins and Anna Wierzbicka (eds.), *Emotions in Crosslinguistic Perspective*, 221–258 Berlin: Mouton de Gruyter.
- 2008. Two virtuous emotions in Japanese: *Nasake/joo* and *jihhi*. In Cliff Goddard (ed.) *Cross-Linguistic Linguistics*, 331–347. Amsterdam: John Benjamins.
- Hellwig, Birgit. 2010. Meaning and translation in linguistic fieldwork. *Studies in Language* 34(4), 802–831.
- Hill, Deborah. 1994. Longgu. In Cliff Goddard and Anna Wierzbicka (eds.), *Semantic and Lexical Universals – Theory and Empirical Findings*, 310–330. Amsterdam: John Benjamins.
- Hinton, Leanne. 2001a. Language Revitalisation: An Overview. In Leanne Hinton and Kenneth Hale (eds.) *The Green Book of Language Revitalization in Practice*, 3–18.

- Boston: Academic Press.
- 2001b. The Karuk Language. In Leanne Hinton and Kenneth Hale (eds.) *The Green Book of Language Revitalization in Practice*, 191–194. Boston: Academic Press.
- Junker, Marie-Odile. 2003. A Native American view of the “mind” as seen in the lexicon of cognition in East Cree. *Cognitive Linguistics* 14(2-3), 167–194.
- 2007. The language of memory in East Cree. In Mengistu Amberber (ed.), *The Language of Memory in a Crosslinguistic Perspective*, 235–261. Amsterdam: John Benjamins.
- 2008. Semantic primes and their grammar in a polysynthetic language: East Cree. In Cliff Goddard (ed.), *Cross-Linguistic Linguistics*, 163–204. Amsterdam: John Benjamins.
- Junker, Marie-Odile and Blacksmith, Louise. 2006. Are there emotional universals? Evidence from the Native American language East Cree. *Culture & Psychology* 12(3), 275–303.
- Keesing, R. 1984. Rethinking mana. *Journal of Anthropological Research* 40, 137–156.
- Knight, Emily. 2008. Hyperpolysemy in Bunuba, a polysynthetic language of the Kimberley, Western Australia. In Cliff Goddard (ed.), *Cross-linguistic semantics*, 205–223. Amsterdam: John Benjamins.
- Leung, Helen Hue Lam. 2012. The semantics of the Cantonese utterance particle ‘laa1’. In M. Ponsonnet, L. Dao & M. Bowler (eds.), *Proceedings of the 42nd Australian Linguistic Society Conference – 2011*, Australian National University, Canberra ACT, 2-4 December 2011 (pp. 245-280).
- Levisen, Carsten. 2012. *Cultural Semantics and Social Cognition. A case study of the Danish universe of meaning*. Berlin: Mouton de Gruyter.
- Levinson, Stephen C. and Sergio Meira. 2003. ‘Natural’ concepts in the spatial topological domain — adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language* 79(3), 485–516.
- Löbner, Sebastian. 2002. *Understanding Semantics*. London: Arnold Publications
- Macaulay, Monica. 2004. On the Karuk directional suffixes. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 30(2).
[elanguage.net/journals/bls/article/download/839/727]
- Majid, Asifa. 2012. A guide to stimulus-based elicitation for semantic categories. In Nicholas Thieberger (ed.), *The Oxford Handbook of Linguistic Fieldwork*, 54–71. Oxford: Oxford University Press.
- Majid, Asifa and Melissa Bowerman (eds.). 2007. Cutting and breaking events across languages. [Special issue] *Cognitive Linguistics* 18(2).

- Mason, Te Haunihiata. 2008. The Incorporation of *Mātauranga Māori* or Māori Knowledge into *Te Mātāpuna*, the First Monolingual Māori Dictionary for Adults. In Rob Amery and J. Nash (eds.). *Warra wiltaniappendi Strengthening languages: Proceedings of the Inaugural Indigenous Languages Conference (ILC) 2007*. Adelaide: University of Adelaide SA. 35–39.
- Matthewson, Lisa. 2004. On the methodology of semantic fieldwork. *International Journal of American Linguistics* 70, 369–415.
- Moorfield, John C. 2005. *Te aka. Maori-English, English-Maori dictionary and index*. Auckland: Pearson/Longman.
- Mosel, Ulrike. 1994. Samoan. In Cliff Goddard and Anna Wierzbicka (eds.), *Semantic and Lexical Universals – Theory and Empirical Findings*, 331–360. Amsterdam: John Benjamins.
- 2006. Grammaticography: The art and craft of writing grammars. In Felix Ameka, Alan Dench and Nicholas Evans (eds.), *Catching Language. The standing challenge of grammar writing*, 41–68. Berlin: Mouton de Gruyter.
- 2012. Morphosyntactic analysis in the field: a guide to the guides. In Nicholas Thieberger (ed.) *The Oxford Handbook of Linguistic Fieldwork*, pp.72–89. Oxford: Oxford University Press.
- Murphy, M. Lynne. 2010. *Lexical Meaning*. Cambridge: Cambridge University Press.
- Myhill, John. 1996. Is BAD a semantic primitive? Evidence from Biblical Hebrew. *Lexicology* 2(2), 99–126.
- Nicholls, Sophie. 2009. Referring Expressions and Referential Practice in Roper Kriol (Northern Territory, Australia). PhD Thesis, University of New England. [<https://e-publications.une.edu.au/vital/access/manager/Repository/une:9244>]
- Pawley, Andrew. 1994. Kalam exponents of lexical and semantic primitives. In Cliff Goddard and Anna Wierzbicka (eds.), *Semantic and Lexical Universals: Theory and Empirical Findings*, 87–422. Amsterdam: John Benjamins.
- Peeters, Bert (ed.). 2006. *Semantic Primes and Universal Grammar: Empirical Findings from the Romance Languages*. Amsterdam: John Benjamins.
- Peeters, Bert. 2010. La métalangue sémantique naturelle: acquis et défis. In J. François (ed.), *Grandes voies et chemins de traverse de la sémantique cognitive*, 75-101. Leuven: Peeters.
- Pica, P., C. Lemer, V. Izard & S. Dehaene. 2004. Exact and approximate arithmetic in an Amazonian indigene group. *Science* 306. 499–503.

- Ponsonnet, Maïa. 2009. Aspects of the semantics of intellectual subjectivity in Dalabon (south-western Arnhem Land). *Australian Aboriginal Studies* 2009/1, 16–28.
- Priestley, Carol. 2002. Insides and emotion in Koromu. *Pragmatics & Cognition* 10(1/2), 243–270.
- 2008. The semantics of “inalienable possession” in Koromu (PNG). In Cliff Goddard (ed.), *Cross-Linguistic Semantics*, 277–300. Amsterdam: John Benjamins.
- 2012a. Koromu temporal expressions: Semantic and cultural perspectives. In Luna Filipović and Kasia Jaszczolt (eds.), *Space and Time in Languages and Cultures: Language, Culture and Cognition*, 143–165. Amsterdam/Philadelphia: John Benjamins.
- 2012b. The expression of potential event modality in the Papuan language of Koromu. In M. Ponsonnet, L. Dao & M. Bowler (eds.), *Proceedings of the 42nd Australian Linguistic Society Conference – 2011*, Australian National University, Canberra ACT, 2-4 December 2011 (pp. 389-422).
- Sakel, Jeanette and Everett, Daniel L. 2012. *Linguistic Fieldwork: A Student Guide*. Cambridge: Cambridge University Press.
- Shi-xu. 2000. To feel or not to feel, that is the question. *Culture & Psychology* 6(3), 375–383.
- Stanwood, Ryo E. 1997. The primitive syntax of mental predicates in Hawaii Creole English: A text-based study. *Language Sciences* 19(3), 209–217.
- 1999. On the Adequacy of Hawai’i Creole English. PhD dissertation. University of Hawai’i.
- Thieberger, Nicholas (ed.) 2012. *The Oxford Handbook of Linguistic Fieldwork*. Oxford: Oxford University Press.
- Tien, Adrian. 2009. Semantic prime HAPPEN in Mandarin Chinese: In search of a viable exponent. *Pragmatics & Cognition* 17(2), 356–382.
- Tong, Malindy, Yell, Michael and Goddard, Cliff. 1997. Semantic primitives of time and space in Hong Kong Cantonese. *Language Sciences* 19(3), 245–261.
- Travis, Catherine E. 2002. La Metalengua Semántica Natural: The Natural Semantic Metalanguage of Spanish. In Cliff Goddard and Anna Wierzbicka (eds.), *Meaning and Universal Grammar – Theory and Empirical Findings*, Volume I, 173–242. Amsterdam: John Benjamins.
- 2004. The ethnopragmatics of the diminutive in conversational Colombian Spanish. *Intercultural Pragmatics* 1(2), 249–274.

- 2006. The communicative realization of *confianza* and *calor humano* in Colombian Spanish. In Cliff Goddard (ed.) *Ethnopragmatics: Understanding Discourse in Cultural Context*, 199–230. Berlin: Mouton de Gruyter.
- Vanhatalo, Ulla, Anna Idström and Heli Tissari. Forthcoming. The foundations of the Finnish based Natural Semantic Metalanguage.
- Wakefield, John C. 2011. *The English Equivalents of Cantonese Sentence-Final Particles: A Contrastive Analysis*. (PhD Thesis). The Hong Kong Polytechnic University.
- Warner, Sam L. No'eau. 2001. The movement to revitalize Hawaiian language and culture. In Leanne Hinton & Kenneth Hale (eds.), *The green book of language revitalization in practice*, 133–44. Boston: Academic Press.
- Wierzbicka, Anna. 1996. *Semantics: Primes and Universals*. New York: Oxford University Press.
- 1997. *Understanding cultures through their key words: English, Russian, Polish, German, Japanese*. New York: Oxford University Press.
- 2002. Semantic primes and universal grammar in Polish. In Cliff Goddard and Anna Wierzbicka (eds.), *Meaning and Universal Grammar – Theory and Empirical Findings*. Vol. II, 65–144. Amsterdam: John Benjamins.
- 2007. Is “remember” a universal human concept? “Memory: and culture. In Mengistu Amberber (ed.), *The Language of Memory in a Crosslinguistic Perspective*, 13–39. Amsterdam: John Benjamins.
- 2006. *English: Meaning and culture*. Oxford: Oxford University Press.
- 2008. Case in NSM: A reanalysis of the Polish Dative. In Andrej Malchukov and Andrew Spencer (eds.), *The Oxford Handbook of Case*, 151–169. Oxford: Oxford University Press.
- 2009. The theory of the mental lexicon. In Sebastian Kempgen, Peter Kosta, Tilman Berger and Karl Gutschmidt (eds), *Die slavischen Sprachen/The Slavic Languages: Eine internationales Handbuch zu ihrer Struktur, ihrer Geschichte und ihrer Erforschung/An international handbook of their structure, their history and their investigation*, 848–863. Berlin: Mouton de Gruyter.
- 2010a. *Experience, Evidence, and Sense: The hidden cultural legacy of English*. New York: Oxford University Press.
- 2010b. ‘Story’ – An English cultural keyword and a key interpretive tool of Anglo culture. *Narrative Inquiry* 20(1), 53–181.

- 2012. Understanding others requires shared concepts. *Pragmatics & Cognition* 20(2), 356-379.
- 2014. *Imprisoned in English: The Hazards of English as a Default Language*. New York: Oxford University Press.
- in press. Two levels of verbal communication: universal and culture-specific. In Louis de Saussure and Andrea Rocci (eds.), *The Handbooks of Communication Science, Volume 3: Verbal Communication*. Berlin: Mouton de Gruyter.
- Forthcoming. Back to 'mother' and 'father': Overcoming the Eurocentrism of kinship studies through eight lexical universals.
- Wilson, William H. and Kaijanoe Kamanā. 2001. "Mai Loko Mai O ka 'I'ini: Proceeding from a dream": The 'Aha Pūnana Leo Connection in Hawaiian language revitalization. In *The Green Book of Language Revitalization in Practice*, 147–178. Boston: Academic Press.
- Ye, Zhengdao. 2006. Why the 'inscrutable' Chinese face? Emotionality and facial expression in Chinese. In Cliff Goddard (ed.) *Ethnopragmatics: Understanding Discourse in Cultural Context*, 127–169. Berlin: Mouton De Gruyter.
- 2007. '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*, 127–169. Amsterdam: Benjamins.
- 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.
- Yoon, Kyung-Joo. 2006. *Constructing a Korean Natural Semantic Metalanguage*. Seoul: Hankook Publishing Co.
- 2007. Mental states reflected in cognitive lexemes related to memory: A case in Korean. In Andrea C. Schalley and Drew Khlentzos (eds.), *Mental States: Volume 2: Language and Cognitive Structure*, 85–108. Amsterdam: John Benjamins.

APPENDIX: 150 CANONICAL SENTENCES FOR IDENTIFYING SEMANTIC PRIMES AND THE CORE LEXICOGRAMMAR OF ANY LANGUAGE

Cliff Goddard and Anna Wierzbicka [February 2014]

This list can be used as an aid to identifying exponents of semantic primes and for eliciting a semantically-based sample of the core lexicogrammar of any language. It is not exhaustive of all combinatorial possibilities posited for semantic primes. The list is divided into sections that correspond roughly to the conventional groupings used for semantic primes, but there is a great deal of overlap between the sections. Many primes occur in several different sections and some very common items, e.g. ‘big’, ‘small’, ‘not’, do not have separate sections devoted to them. Some sentences are composed exclusively in semantic primes, while others include complex vocabulary items which may need to be adjusted to suit the cultural context.

The current “canonical sentences” list is the latest in a line that goes back to the *Semantic and Lexical Universals* project (Goddard and Wierzbicka eds., 1994). It is not regarded as final and suggestions are welcome for its improvement. We have tried to supply sentences that are likely to be fairly “ordinary” in most cultural settings. They can be used in a variety of ways. For example, one can use them in direct elicitation, embed them in mini-narratives or scenarios as stimuli or prompts for informants, or seek to locate matching or comparable examples in natural texts.

SOMEONE~WHO, SOMETHING~THING~WHAT, PEOPLE, BODY

Many other examples occur throughout. Note that sometimes expressions like ‘one person’ may be used instead of SOMEONE.

1. Who did it?
2. Someone took (ate, broke, etc.) it. I don’t know who.
3. What happened?
4. I did something bad.
5. What is this?
6. This thing is big/small.
7. They are good people.
8. Many people don’t know this.
9. Sometimes bad things happen to (good) people.
10. The body of a man is not like the body of a woman.
11. The stomach (liver, etc.) is part of the body.

PARTS, KIND

12. This thing (knife, axe, etc.) has two parts.

13. This thing has many small parts.
14. Two kinds of people live here.
15. There are two/many kinds of yam (bat, bird, etc.).
16. [of two kinds of fish, fruit, etc.] This kind of fish is good to eat. This (other) kind is not good to eat.

THIS, THE SAME, OTHER~ELSE

Many other examples occur throughout.

17. This is good/bad.
18. This is a big/small basket (knife, bowl, etc.).
19. (I saw) these two people; all these people.
20. I saw the same person (or: the same people) the day after.
21. It happened at the same place/time, not at another place/time.
22. She did/said the same (the same thing).
23. Someone else did it.

ONE, TWO, SOME, ALL, MANY~MUCH, LITTLE~FEW

There are many examples of ONE, TWO, and MUCH~MANY in other sections.

24. How many children does she have?
25. She has one child (two/many children).
26. How much water is there (in that well, bowl, etc.)?
27. Many (or: not many) people live here.
28. (There were many people there.) After that, one/two (of them) left.
29. Some of these yams (lemons, etc.) are good, some are rotten.
30. In some places there are lots of fish (rabbits, birds, etc.).
31. There is much (a lot of) water here.
32. There is very little water here.
33. Very few people live there now.
34. Many of them went to the east, some of them went to the west.
35. All his sons (daughters, etc.) are now married.
36. The fire burnt everything (e.g. in the house, on the island).
37. (From the top of the mountain) I could see everything (below me).

GOOD, BAD

BIG and SMALL are not listed separately but appear in many examples throughout. There are plenty of additional examples of GOOD and BAD in other sections too.

38. Good people don't do things like this.
39. This was good/bad for me.
40. It is good/bad if someone (or: people) does something like this.
41. I know this person/place well.
42. Some people can do this well, some people can't.

THINK, KNOW, WANT, DON'T WANT

Note that the word for 'know (someone)' will often be different from the word for 'know (something)'.

43. What are you thinking about?
44. I thought about it for a long time.
45. This person thinks something bad about me.
46. I thought like this: “ – – ”
47. Think well about it first (before you do it).
48. Where is he now? I don't know. Maybe Mary knows.
49. He knows much (many things) about things like this.
50. I know (or: don't know) that person.
51. I want (don't want) you to do it.
52. I want (don't want) this to happen.

FEEL, SEE, HEAR

There are several examples of SEE in other sections.

53. I don't feel well today.
54. I often feel like this.
55. At that time (then), I didn't feel anything.
56. I saw something (a dog, bird, etc.) there.
57. (After that), I saw something else.
58. She was singing – I heard it.
59. I hear something (over there in the bushes).
60. I couldn't hear it.

SAY, WORDS, TRUE

There are several examples of SAY in other sections.

61. 'This is good', I said.
62. What did he say?
63. He said something to her, but she didn't hear it.
64. If I do this, people can say bad things about me.
65. She said the same thing, but in other (i.e. different) words.
66. In language X, the word for 'good' is palya.
67. If you say one more word,
68. That's not true.

DO, HAPPEN, MOVE, TOUCH

There are many examples of DO and HAPPEN in other sections.

69. What did you do then (after that)?
70. Those people did something good for me.
71. What do people do with things like this (a spoon, needle, etc.)?
72. People do this with a knife (axe, stick, etc.).
73. Something good/bad happened to this person (to me).
74. Something bad happened in that place.
75. (Look!) Something is moving over there.
76. Don't move! (e.g. while putting on a bandaid or removing headlice)
77. I was very scared. I couldn't move.
78. Something was touching my hand. (I could feel it.)
79. Don't touch it! (It's hot.)

THERE IS, BE (SOMEONE'S)

80. There is someone in the garden.
81. There is/are no [e.g. water/spiders] here.
82. There are no ghosts (unicorns, etc.).
83. There are many kinds of nuts (yams, bats, etc.).
84. This knife (dog, etc.) is John's.
85. Whose is it?.
86. Don't touch it. It's not yours. It's someone's else's.

BE (SOMEWHERE), BE (SOMEONE/SOMETHING)

There are several examples of BE (SOMEWHERE) in other sections.

87. I want to be with my mother (grandfather, father, etc.).
88. This is something big/small.
89. I don't know what this is.
90. I don't know who this is.

LIVE, DIE

There are many examples of LIVE in other sections.

91. [In story of olden days] These people lived for a long time.
92. Turtles (horses, whales, etc.) live for a long time.
93. At that time she was living with her mother.
94. Fish live in the sea.
95. All people die.
96. Jesus (Buddha, etc.) died a long time ago.

TIME~WHEN, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME, IN ONE MOMENT

There are many examples of TIME~WHEN and A LONG TIME in other sections. Note that BEFORE and AFTER may have adverbial, rather than adpositional, syntax.

97. When did you do it? I did it when everyone was asleep.
98. I don't know when it happened.
99. At the same time, something happened to me.
100. Sometimes dogs bark at night, sometimes they don't (bark).
101. This happened two times (twice) / many times (often).
102. This dog always barks at night.
103. They live in X now. They lived in Y before (this).
104. It happened beforehand/afterwards.
105. John was born before/after Harry.
106. This happened a long time ago.
107. She slept for a long time/for a short time.
108. She thought about it for some time.
109. It happened some time ago (i.e. some time before).
110. He (only) stayed there for a short time, but she stayed there for a long time.
111. It happened in one moment.

PLACE~WHERE, HERE, NEAR, FAR, ABOVE, BELOW, ON (THIS) SIDE, INSIDE,

ON

112. Where is he now?
113. It didn't happen here. It happened somewhere else (in another place).
114. Is it far from here?
115. Was Fred standing (sitting, etc.) near John?
116. The house (camp, etc.) is near the river.
117. It is far above the ground.
118. This thing is above/below this other thing.
119. The head is above other parts of the body.
120. The feet are below other parts of the body.
121. John is now [standing/sitting] on this side (of me).
122. Mary was [standing/sitting] on my right-hand side, Sally was on my left-hand side.
123. They live on the other side of the river (mountain, etc.)
124. There were people standing on both sides of the road.
125. There is an insect inside this [e.g. walnut].
126. This cave is very big — lots of people could live inside it.
127. There are many small things on the ground in this place.

BECAUSE, IF, MAYBE, CAN

NOT is not listed separately but appears in numerous examples throughout.

128. There was a lot of noise (shouting, etc). Because of this, I couldn't sleep.
129. Why are you crying? Because he hit me.
130. If you do this, people can say something bad about you.
131. Maybe he will come tomorrow, maybe he won't (come tomorrow).
132. I can't do it now, maybe my brother (friend, etc.) can.
133. I can't do it now, but I can do it later.
134. I can't not do this.
135. Bad things can happen to everyone/good people.
136. This thing can move.

VERY, MORE

There are several examples of VERY in other sections.

137. It is very good/bad/big/small.
138. This place [e.g. Gundaroo] is very far from here.
139. I very much want to do this.
140. More! (e.g. a child crying for food)
141. (I don't know much about it.) I want to know more.
142. I want to see more.
143. She doesn't live here anymore.
144. I want one more.

LIKE

There are several examples of LIKE in other sections.

145. This person is not like other people.
146. This is like lilac (or: any plant), but it is not lilac.

147. At a time like this; in a place like this.
148. He did it (sang, danced, etc.) like this: ...
149. It happened like this: ...
150. She sang like this: —— [to demonstrate loudness, pitch, voice quality, etc.]

