# Lexical Semantics

Introductory article

*Chris Barker*, University of California, San Diego, California, USA

*Word meaning mediates between conceptualization and language: simply put, words name concepts. Studying which concepts can have names – and how the fine-grained structure of those concepts interacts with the linguistic structures that contain them – reveals something important about the nature of language and cognition.*

## WHAT CAN WORDS MEAN?

To a first approximation, lexemes are words, so lexical semantics is the study of word meaning. The main reason why word-level semantics is especially interesting from a cognitive point of view is that words are names for individual concepts. Thus lexical semantics is the study of those concepts that have names. The question 'What can words mean?', then, amounts to the question 'What concepts can have names?'

There are many more or less familiar concepts that can be expressed by language but for which there is no corresponding word. There is no single word in English that specifically names the smell of a peach, or the region of soft skin on the underside of the forearm, though presumably there could be. Furthermore, it is common for one language to choose to lexicalize a slightly different set of concepts from another. American speakers do not have a noun that is exactly equivalent to the British *toff*, an 'upper-class person', nor does every language have a verb equivalent to the American *bean*, 'to hit on the head with a baseball'.

The situation is quite different for other concepts, however. To adapt a famous example from philosophy, there is no word in any language that refers specifically to objects that were green before 1 January 2000 but blue afterwards. Of course, we can artificially agree to assign this concept to a made-up word; in fact, let's call it *grue*. But doing so doesn't make *grue* a legitimate word, let alone a genuine one. Could there be such a word? That is,

is naming the grue concept truly unnatural, or merely unlikely? (*See* **Meaning**)

Identifying systematic patterns governing the meanings of related words can provide convincing answers to such questions. Furthermore, the insights go in two directions simultaneously: on the one hand, natural classes of words that resemble each other syntactically (for instance, the class of nouns) have similar kinds of meaning. This places bounds on the set of possible word meanings. On the other hand, natural classes of words that resemble each other semantically (for instance, the class of words whose meanings involve the notion of causation) behave similarly from a syntactic point of view. Thus studying how the fine-grained conceptual structure of a word's meaning interacts with the syntactic structures that contain it provides a unique and revealing window into the nature of language, conceptualization, and cognition.

## POLYSEMY

Lexemes are linguistic items whose meaning cannot be fully predicted based on the meaning of their parts. Idiomatic expressions by definition, then, are lexemes: the expression *kicked the bucket* (as in the idiomatic interpretation of 'My goldfish kicked the bucket last night') means essentially the same thing as the word *died*. (Although it is convenient to draw examples from English in this article, it gives an unfortunate English-centric perspective to the discussion. Those readers familiar with agglutinative languages such as Turkish or West Greenlandic should consider whether 'morpheme' might be a more accurate term than 'word' in the general case.) Other articles in this encyclopedia discuss cases in which lexemes do not correspond exactly to words, but this article will adopt the simplifying assumption that lexemes are words. (*See* **Construction Grammar**; **Lexicon**)

But how many distinct words are there? We can feel comfortable deciding that homonyms (words that sound the same but have completely different meanings) should be counted as different lexical items: *bat* the flying mammal versus *bat* the wooden stick used in baseball certainly ought to be considered as two different words. It is more difficult to be confident of such judgments, however, when the two meanings seem to be closely related. Is *wave* in reference to a beach a different concept than *wave* in 'He experienced a wave of anger'? What about verbal uses such as 'The infant waved' or 'The policeman waved us through'? Words that have multiple related but distinct meanings are 'polysemous'. Identifying and dealing with polysemy is a hotly debated and difficult problem in lexical semantics.

Some polysemy seems to be at least partly systematic. Pustejovsky proposes that language users rely on *qualia*, which are privileged aspects in the way they conceptualize objects. The four main qualia are the object's constituent parts, its form (including shape), its intended function, and its origin. Thus a fast car is fast with respect to performing its intended function of transporting people, but fast food is fast with respect to its origin. We can recognize the polysemy of *fast* as indeterminacy in selection of the most relevant qualia, without needing to postulate two lexical entries for *fast*, let alone two distinct kinds of fastness.

In addition, qualia provide a way to understand how to coerce a conventional meaning into a variety of extended meanings. In 'The ham sandwich at table 7 is getting restless' (spoken by one waiter addressing another in a restaurant), *the ham sandwich* refers not to the material parts of the sandwich, but to the person involved in specifying the sandwich's intended function (i.e. to be eaten by someone).

## LEXICAL RELATIONS (MEANING IN RELATION TO OTHER WORDS)

There are two main modes for exploring word meaning: in relation to other words (this section), and in relation to the world (see below). The traditional method used in dictionaries is to define a word in terms of other words. Ultimately, this strategy is circular, since we must then define the words we use in the definition, and in their definitions, until finally we must either run out of words or reuse one of the words we are trying to define.

One strategy is to try to find a small set of *semantic primes* or basic constituent elements: Wierzbicka identifies on the order of 50 or so concepts (such as *good*, *bad*, *before*, *after*, *I*, *you*, *part*, *kind* …) that allegedly suffice to express the meaning of all words (in any language). Whether this research program succeeds or not has important implications for the nature of linguistic conceptualization.

In any case, speakers clearly have intuitions about meaning relations among words. The most familiar relations are synonymy and antinomy. Two words are synonyms if they mean the same thing, for example *filbert* and *hazelnut*, *board* and *plank*, etc. Two words are antonyms if they mean opposite things: *black* and *white*, *rise* and *fall*, *ascent* and *descent*, etc. One reason to think that it is necessary to recognize antinomy as an indispensable component of grammatical descriptions is because at most one member of each antonymous pair is allowed to occur with measure phrases: *tall* is the opposite of *short*, and we can say 'Bill is 6 feet tall', but not '*Tom is 5 feet short'. (Asterisks denote a deviant construction.)

Other major semantic lexical relations include hyponymy (names for subclasses: *terrier* is a hyponym of *dog*, since a terrier is a type of dog), and meronymy (names for parts: *finger* is a meronym of *hand*).

Words whose meanings are sufficiently similar in some respect are often said to constitute a 'semantic field', though this term is rarely if ever given a precise definition. Terms such as *red*, *blue*, *green*, etc., are members of a semantic field having to do with color.

Miller and associates have developed a lexical database called WordNet, a kind of multidimensional thesaurus, in which these types of lexical relations are explicitly encoded. Thus WordNet is an attempt to model the way in which a speaker conceptualizes one kind of word meaning.

## DENOTATION (MEANING IN RELATION TO THE WORLD)

Defining words in terms of other words or concepts will take us only so far. Language talks about the world, and ultimately, in order to know what a word such as *horse* or *despair* means, it is necessary to know something about the world. One way of getting at the connection between meaning and the world is to consider 'truth conditions': what must the world be like in order for a given sentence to be true? In general, words participate in most of the kinds of truth-conditional meaning that sentences do, plus one or two more of their own.

'Entailment' is the most concrete aspect of meaning: one sentence entails another just when the

truth of the first sentence is sufficient to guarantee the truth of the second. For instance, the sentence 'I have four children' entails the sentence 'I have at least two children'. We can extend this notion to words by assuming that one word entails another in cases where substituting one word for the other in a sentence produces an entailment relation. Thus *assassinate* entails *kill*, since 'Jones assassinated the President' entails 'Jones killed the President'. Similarly, *whisper* entails *spoke*, *devour* entails *ate*, and so on.

'Presuppositions' are what must already be assumed to be true in order for a use of a sentence to be appropriate in the first place. For instance, *and* and *but* mean almost the same thing, and differ only in the presence of a presupposition associated with *but*. If I assert that 'Tom is rich and kind', it means exactly the same thing as 'Tom is rich but kind' as far as what Tom is like; the only difference is that the use of the sentence containing *but* presupposes that it is unlikely for someone who is rich also to be kind.

'Selectional restrictions' are a kind of presupposition associated exclusively with words. The verb *sleep* presupposes that its subject is capable of sleeping; that is, it selects a restricted range of possible subjects. One reason why the sentence '*Green ideas sleep furiously' is deviant, then, is because ideas are not capable of sleeping, in violation of the selectional restriction of *sleep*.

'Implicature' is weaker than entailment. If you ask me how a student is doing in a course, and I reply 'She's doing fine', I imply that she is not doing great (otherwise, I would presumably have said so). Yet the sentence 'She's doing fine' does not entail the sentence 'She's not doing great', since it can be true that a student is doing fine work even if she is actually doing spectacular work. The implicature arises in this case through contrast with other words that could have been used. Interestingly, such lexical implicatures favor the formation of lexical gaps: because *fine* implicates *not good*, there is no word that means exactly *not good* (for instance, *mediocre* entails both *not good* and *not bad*). Similarly, because *some* implicates *not all*, there is no single word (perhaps not in any language) that means exactly *not all*.

'Connotation' is the part of the meaning of a word that adds a rhetorical spin to what is said. More specifically, connotation signals the attitude of the speaker towards the object or event described. If I describe someone as *garrulous* rather than *talkative*, the described behavior may be exactly the same, but *garrulous* conveys in addition

the information that I disapprove of or dislike the behavior in question.

'Vagueness' is perhaps the quintessential problem for scientific theories of word meaning. Imagine a person with a full head of hair; clearly, this person is not bald. Now imagine pulling out one hair at a time. Eventually, if this process is continued long enough, the person will qualify as bald. But at what point precisely does the switch from not-bald to bald occur? It's impossible to say for sure. But if we can't answer this simple question, then how can we possibly claim to truly understand the meaning of the word *bald*? A little reflection will reveal that virtually every word that refers to objects or events in the real world contains vagueness: exactly how tuneless must a vocal performance get before the claim 'She is singing' is no longer true? Is that short sofa really a chair? The pervasiveness and intractability of vagueness gives rise to profound philosophical issues. (*See* **Vagueness**)

## KINDS OF WORDS, KINDS OF MEANING

One of the most important principles in lexical semantics (part of the heritage of Richard Montague) is that words in the same syntactic class (e.g. noun, verb, adjective, adverb, determiner, preposition, etc.) have similar meanings. That is, the meaning of a determiner like *every* differs from the meaning of a verb like *buy* in a way that is qualitatively different from the difference in meaning between two verbs like *buy* and *sell*.

The overarching division in word classes distinguishes function words from content words. Function words (determiners, prepositions, conjunctions, etc.) are shorter, higher-frequency words that serve as syntactic glue to combine words into sentences. Content words (nouns, verbs, adjectives, adverbs) carry most of the descriptive payload of a sentence. In the sentence 'He asked me to give her a haircut', the content words are *asked*, *give*, and *haircut*.

Semantically, function words are much closer to being able to be expressed as a mathematical or logical function. We can write a rule that gives a good approximation of the contribution of *and* to truth conditions in the following way: 'A sentence of the form [S1 and S2] will be true just in the case where S1 is true and S2 is also true.' Thus the sentence 'It was raining and it was dark' will be true just in the case where the sentence 'It was raining' is true and the sentence 'It was dark' is also true.

Among function words, the meaning of determiners such as *every*, *no*, *most*, *few*, *some*, etc. (but crucially excluding noun phrase modifiers such as *only*) have been studied in exquisite detail, especially from the point of view of their logical properties. One of the most celebrated results is the principle of 'conservativity': given any determiner D, noun N, and verb phrase VP, sentences of the form [D N VP] are guaranteed to be semantically equivalent to sentences of the form [D N be N *and* VP]. For instance, if the sentence 'Most cars are red' is true, then the sentence 'Most cars are cars and are red' is also true, and vice versa; similarly, 'Some cars are red' is equivalent to 'Some cars are cars and are red', and so on. It is conjectured that every determiner in every language obeys conservativity. These equivalences may seem so obvious that it is hard to appreciate just how significant this generalization is; it may help, therefore, to imagine a new determiner *antievery* as having a meaning such that a sentence 'Antievery car is red' means the same thing as 'Everything that is not a car is red'. The claim is that there could be no such determiner in any natural language; that is, *antievery* (or any determiner that violated conservativity) is an impossible thing for a word to mean.

There is a class of lexemes whose behavior is especially fascinating from a cognitive point of view. In English this class includes expressions such *ever*, *anymore*, *budge an inch*, and many others. These items can only be used in sentences that exhibit certain properties related to the presence of negation. For instance, it is possible to say 'She doesn't ever go to the movies anymore', but it is impossible to say '*She ever goes to the movies anymore'. It is the presence of the negation in the first sentence that licenses the words in question, and it is because of this affinity for negation that they are called 'negative polarity items'. Not only are NPIs interesting in their own right, they provide clues about the internal structure of the meaning of other words. Consider the following contrast: 'I doubt that she ever goes to the movies anymore' is fine, but '*I think that she ever goes to the movies anymore' is deviant. Because negative polarity items are licensed by *doubt* but not by *think*, we can conclude that the meaning of *doubt* contains negation hidden within it, so that *doubt* means roughly *think that not* (i.e. *think that it is not the case that*).

## EVENTS

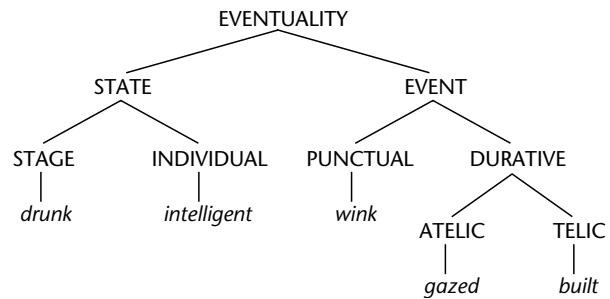The main linguistically important divisions among event types are as shown in Figure 1. Stage-level



**Figure 1.**

adjectives such as *drunk* correspond to temporary stages of an individual rather than properties that are (conceived of as being) permanent properties of individuals, such as *intelligent*. Telic activities differ from atelic ones in having a natural end point. Crucially, these semantic distinctions correspond to systematic differences in syntactic behavior. For instance, assume that the temporal use of the preposition *for* requires the absence of an end point, but *in* requires a definite end point. If *gaze* is naturally atelic but *built* is telic, this explains why it is possible to say 'He gazed out to sea for an hour' but not '*He gazed out to sea in an hour'; and conversely, why it is impossible to say '*He built the raft for an hour' at the same time that 'He built the raft in an hour' is perfectly fine.

Although many lexical predicates are naturally atelic or telic, the telicity of the sentence in which they occur can be determined by other elements in the sentence. For instance, some verbs (*drink*, *write*, *draw*, *erase*, *pour*, etc.) have meanings that establish a systematic correspondence between the physical subparts of one of their arguments and the described event. Consider the verb *mow*: the subparts of an event of mowing the lawn correspond to the subparts of the lawn. If the event of mowing the lawn is half over, then roughly half the lawn will have been mown. This contrasts with a verb like *interview*: if an interview with the President is half over, that does not mean that half of the President has been interviewed. Because of the systematic correspondence between the internal structure of the noun phrase (NP) denotation and the internal structure of the described event, semantic properties of this special type of verbal argument-NP (which is called an 'incremental theme') can project to determine semantic properties of the described event. In particular, the telicity of sentences containing incremental theme verbs depends on whether the incremental-theme argument is conceived of as quantized (*a beer* is quantized,

compared with just *beer*): 'John drank a beer in/*for an hour' versus 'John drank beer for/*in an hour'.

# THEMATIC ROLES

## Decomposition

Sentences have internal structure in which the basic unit is the word. For instance, the sentence 'The girl hit the boy' has for its coarse-grained structure [[the girl] [hit [the boy]]]. Words have internal structure as well, in which the basic unit is the morpheme. For instance, *undeniable* = [un [deni [able]]]. Just as the meaning of a sentence must be expressed as a combination of the meaning of its parts and their syntactic configuration, we expect the meaning of a word (in the normal case) to be a combination of the meanings of its parts, and indeed *undeniable* means roughly *not able to be denied*. The internal structure of words can be quite complex, as in agglutinative languages like Turkish, in which there is no principled limit to the number of morphemes in a word. (*See* **Morphology**; **Syntax**)

Yet in general, word structure is qualitatively less complex than syntactic structure in a variety of measurable ways. If the structure of word meaning mirrors the way in which words are built up from morphemes, then we might also expect that word meaning will be simpler than sentence meaning in some respects. However, as discussed in the next section, many scholars believe that the meaning of words has internal structure that does not correspond to any detectable morphological structure.

## Argument Structure and Conceptual Structure

One of the main ways in which words in the same main syntactic class differ in their syntax and in their meaning is in terms of 'argument structure'. *Fall*, *hit*, and *give* are all verbs, but differ in the number and arrangement of noun-phrase arguments they occur with: intransitive verbs have one argument ('John fell'); transitive verbs have two arguments ('John hit the ball'), and ditransitive verbs have three arguments ('John gave the book to Mary'). Thus it is necessary to annotate each word with its argument structure, and the meaning of the verb will refer to elements in the argument structure.

'Conceptual structure' defines the connection between argument structure and meaning. According to Jackendoff, for example, the transitive verb *drink*

corresponds to the following (simplified) lexical entry:

| | |
|---|---|
| Word: | *drink*, verb (transitive) |
| Argument structure: | $\langle NP_i, NP_j \rangle$ |
| Conceptual structure: | CAUSE ($NP_i$, GO ($NP_j$, TO (IN (MOUTH-OF ($NP_i$))))) |

In the sentence 'John drank the soda', then, $NP_i$ = *John*, $NP_j$ = *the soda*, and the sentence as a whole means 'John caused the soda to go into John's mouth'.

Note that the conceptual structure is considerably more complex than the syntactic structure in which the verb participates: the concept named by *drink* has been decomposed into five separate predicates (CAUSE, GO, TO, IN, and MOUTH-OF), and there are three arguments at the conceptual level (John, the soda, and John's mouth) instead of two, as at the overt syntactic level.

Levin and others show how separating conceptual structure from argument structure provides a framework for explaining a type of polysemy that gives rise to various syntactic 'alternations': cases in which a single word can be used with a variety of argument-structure configurations. The basic idea is that when a verb like *break* occurs in a transitive sentence such as 'The girl broke the window', it has a conceptual structure that explicitly involves causation – perhaps (CAUSE (THE-GIRL, BECOME (BROKEN, THE-WINDOW))). When the same verb occurs with a single argument, as in 'The window broke', it takes for its conceptual meaning a proper subpart of the transitive structure – (BECOME (BROKEN, THE-WINDOW)). Crucially, the substructure omits mention of the causing entity.

A particularly interesting kind of alternation is known as an 'implicit argument'. For instance, when comparing the meaning of 'John ate something' versus 'John ate', the omitted argument in the shorter version is intuitively still present conceptually. In this case the two senses of *eat* differ only in argument structure but not in conceptual structure.

## Thematic Roles and Linking

It is not an accident given the meaning of the word *drink* that the entity that causes the drinking event is expressed as the subject of the sentence: in 'John drank the soda', we know that John is consuming the soda, and not vice versa. More generally, to some degree at least, the meaning of verbs clearly

interacts with the syntactic form of sentences containing those verbs. The various theories developed to characterize the connection between verb meaning and syntactic structure are called 'linking theories', and in many linking theories the connection depends on 'thematic roles'. A verb like *kiss*, for instance, describes events that have two main conceptual participants, which we can call the kisser (the entity that does the kissing) and the kissee (the entity that gets kissed). When *kiss* occurs as the main verb in a sentence of English, the kisser is always syntactically realized as the subject and the kissee is realized as the direct object:

| Syntactic structure: | [The girl] | kissed | [the boy] |
|---|---|---|---|
| Syntactic function: | Subject | | Direct object |
| Verb-specific roles: | kisser | | kissee |
| Thematic roles: | Agent | | Patient |

Thus the linking rules for English must guarantee that the verb-specific kisser role always comes to be associated with the grammatical subject of the sentence. (Passive sentences such as 'The boy was kissed by the girl' are a systematic variation on the basic pattern in which the normal basic linking is deliberately reversed for discourse purposes.)

Building on ideas of Davidson that emphasize the importance of events in the semantics of verbs, Parsons accounts for linking regularities by providing thematic roles as primitive notions, and stipulating that when a verb assigns a noun phrase the role of Agent, that NP must appear in subject position. The sentence 'Brutus stabbed Caesar with a knife' would be rendered as [STABBING($e$) & AGENT (BRUTUS, $e$) & PATIENT (CAESAR, $e$) & INSTRUMENT (KNIFE, $e$)], which asserts that $e$ is a stabbing event, Brutus is the agent of the stabbing event, Caesar is the patient, and the knife is the instrument.

Other linking theories attempt to derive thematic roles from independently motivated aspects of meaning. In a theory that recognizes conceptual structure, we can hypothesize that for any verb that contains as part of its meaning the conceptual primitive CAUSE, the first conceptual argument of the CAUSE predicate will be an agent. Put another way, the hypothesis is that the syntactic subject is the participant that the speaker conceptualizes as taking the more active causal role in the event.

Both the neo-Davidsonian approach and the conceptual-structure approach rely on defining the meaning of a word in terms of other more basic concepts. Dowty approaches the issue from the point of view of the relation between word meaning and the world. In particular, he suggests that it suffices to consider the entailments of the verb in question. The verb *kiss* entails that the kisser participant must have lips, that those lips must come in contact with the kissee participant, and the kisser must intentionally cause the contact to occur. In contrast, it is not an entailment that the kissee has lips, since it is perfectly possible to kiss a clam, or a rock, nor is there an entailment that the kissee intentionally causes the kissing event to occur (imagine kissing a sleeping child). On entailment-based theories of thematic roles, if a verb entails that a participant intends for the event described by the verb to come about, or causes it to come about, then that participant will (be more likely to) be expressed as the subject rather than as the direct object.

All theories that recognize the existence of thematic roles provide at least two basic thematic roles: Agent and Patient (the term 'Patient' is often collapsed with the term 'Theme', whence the cover term 'thematic roles'). Agents tend to be subjects; Patients, which are conceived of as undergoing an action (alternatively, are entailed to undergo an action, or which are incremental themes) make good direct objects. Among the other thematic roles that play a prominent part in many linking theories are Experiencer, Source, Goal, and Instrument.

Linking theories make strong predictions about possible and impossible words. Assume we are presented with a verb *glarf* that is allegedly a previously unknown transitive verb of English. We are told the following concerning its meaning: it describes situations in which one participant comes into violent contact with the foot of another participant. Furthermore, we know that it is the participant that possesses the foot who initiates or causes the touching event to occur. Now we encounter the following sentence: 'The ball glarfed the player', which is used as a description of an event in which a football player kicked a football for a goal. (In other words, *glarf* seems to be the verb *kick* but with subject and direct object reversed.) At this point, we can conclude that *glarf* is not a legitimate word of English (and in fact, is not a legitimate basic verb in any human language that recognizes the grammatical relations of subject and object), because it associates the volitional participant with the direct object and the more passive undergoer participant with the subject, contrary to the predictions of linking theories.

In fact, thematic roles can provide insight even into the syntactic behavior of verbs that have only one argument. The 'unaccusative hypothesis', due largely to Perlmutter, was a breakthrough in the use of lexical semantics to explain syntactic patterns. The hypothesis claims that intransitive verbs whose sole argument is semantically an Agent (predicates like *sing*, *smile*, or *walk*) behave differently cross-linguistically from so-called 'unaccusative' verbs, whose one argument is semantically a Patient (e.g. *fall*, *bleed*, or *die* – crucially, these activities are conceived of as not being done intentionally). For instance, unaccusative verbs in Italian generally take *essere* ('to be') as a past auxiliary rather than *avere* ('to have').

Like the correspondence between thematic roles and argument linking, the unaccusativity pattern is systematic enough to be quite compelling. However, as with virtually every attempt to derive syntactic behavior directly from semantic distinctions, examination of a wider range of data has revealed a considerable number of complications and exceptions. Nevertheless, there can be no doubt that meaning and form are deeply and inextricably connected, and that lexical semantics is the place at which they come into contact with one another. Therefore understanding lexical semantics is critical to understanding the yin and yang of language and conceptualization.

## Further Reading

Dowty D (1991) Thematic proto-roles and argument selection. *Language* **67**(3): 547–619.

Fellbaum C (ed.) (1998) *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Grimshaw J (1990) *Argument Structure*. Cambridge, MA: MIT Press.

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

Kamp H and Partee B (1995) Prototype theory and compositionality. *Cognition* **57**: 129–191.

Levin B (1993) *English Verb Classes and Alternations. A Preliminary Investigation*. Chicago, IL: University of Chicago Press.

Levin B and Pinker S (eds) (1992) *Lexical and Conceptual Semantics*. Oxford, UK: Blackwell.

Parsons T (1990) *Events in the Semantics of English: A Study in Subatomic Semantics*. Cambridge, MA: MIT Press.

Pustejovsky J (1995) *The Generative Lexicon*. Cambridge, MA: MIT Press.

Wierzbicka A (1996) *Semantics, Primes and Universals*. Oxford, UK: Oxford University Press.