# 1. Ontology and the Lexicon

Graeme Hirst

Department of Computer Science, University of Toronto,
Toronto M5S 3G4, Ontario, Canada
e-mail: gh@cs.toronto.edu

## 1.1 Lexicons and lexical knowledge

### 1.1.1 Lexicons

A **lexicon** is a list of words in a language—a **vocabulary**—along with some knowledge of how each word is used. A lexicon may be general or domain-specific; we might have, for example, a lexicon of several thousand common words of English or German, or a lexicon of the technical terms of dentistry in some language. The words that are of interest are usually **open-class** or **content** words, such as nouns, verbs, and adjectives, rather than **closed-class** or **grammatical function** words, such as articles, pronouns, and prepositions, whose behaviour is more tightly bound to the grammar of the language. A lexicon may also include multi-word expressions such as fixed phrases (*by and large*), phrasal verbs (*tear apart*), and other common expressions (*merry Christmas!; teach ⟨someone⟩'s grandmother to suck eggs; Elvis has left the building*).

Each word or phrase in a lexicon is described in a **lexical entry**; exactly what is included in each entry depends on the purpose of the particular lexicon. The details that are given (to be discussed further in sections 1.2.1 and 1.3.2 below) may include any of its properties of spelling or sound, grammatical behaviour, meaning, or use, and the nature of its relationships with other words. A lexical entry is therefore a potentially large record specifying many aspects of the linguistic behaviour and meaning of a word.

Hence a lexicon can be viewed as an index that maps from the written form of a word to information about that word. This is not a one-to-one correspondence, however. Words that occur in more than one syntactic category will usually have a separate entry for each category; for example, *flap* would have one entry as a noun and another as a verb. Separate entries are usually also appropriate for each of the senses of a **homonym**—a word that has more than one unrelated sense even within a single syntactic category; for example, the noun *pen* would have distinct entries for the senses writing instrument, animal enclosure, and swan. **Polysemy**—related or overlapping senses—is a more-complex situation; sometimes the senses may be discrete enough that we can treat them as distinct: for example, *window* as both opening in wall and glass pane in opening in wall (*fall through the window; break the window*). But this is not always so; the word *open*, for example, has many overlapping senses concerning unfolding, expanding, revealing, moving to an open position, making openings in, and so on, and separating them into discrete senses, as the writers of dictionary definitions try to do, is not possible (see also sections 1.2.3 and 1.3.1 below).

On the other hand, morphological variants of a word, such as plurals of nouns and inflected forms of verbs might scarcely warrant their own complete lexical entry. Rather, the entry for such forms might be little more than a pointer to that for the base form of the word. For example, the entries for *takes, taking, took,* and *taken* might just note that they are inflected forms of the base-form verb *take*, and point to that entry for other details; and conversely, the entry for *take* will point to the inflected forms. Similarly, *flaps* will be connected both to the noun *flap* as its plural and to the verb *flap* as its third-person singular. The sharing of information between entries is discussed further in section 1.2.2 below.

A lexicon may be just a simple list of entries, or a more-complex structure may be imposed upon it. For example, a lexicon may be organized hierarchically, with default inheritance of linguistic properties (see section 1.2.2 below). However, the structures that will be of primary interest in this chapter are semantic, rather than morphological or syntactic; they will be discussed in section 1.3.2 below.

### 1.1.2 Computational lexicons

An ordinary dictionary is an example of a lexicon. However, a dictionary is intended for use by humans, and its style and format are unsuitable for computational use in a text or natural language processing system without substantial revision. A particular problem is the dictionary's explications of the senses of each word in the form of definitions that are themselves written in natural language; computational applications that use word meanings usually require a more-formal representation of the knowledge. Nonetheless, a dictionary in a machine-readable format can serve as the basis for a computational lexicon, as in the ACQUILEX project (Briscoe, de Paiva, and Copestake 1993)—and it can also serve as the basis for a semantic hierarchy (see section 1.5.2 below). (An alternative or complementary source of lexical information is inference from the usage observed in text corpora; see, e.g., Boguraev and Pustejovsky (1996).)

Perhaps the best-known and most widely used computational lexicon of English is WordNet (Fellbaum 1998). The primary emphasis of WordNet is on semantic relationships between words; it contains little syntactic and morphological data and no phonetic data. The basic lexical entry in WordNet is the **synset** (for "synonym set"), which groups together identical word senses. For example, the synonymous nouns *boarder, lodger,* and *roomer* are grouped together in a synset. WordNet includes an extensive network of relationships between synsets; this will be discussed in detail in section 1.3.2. Following the success of WordNet for English, wordnets with a similar (but not necessarily identical) structure have been (or are being) developed for a number of other languages, including several European languages (some as part of the EuroWordNet project (Vossen 1998)), Hindi, Tamil, and Basque (see *www.globalwordnet.org*).

Some other important general-purpose lexicons include CELEX (Baayen, Piepenbrock, and van Rijn 1993), which is a set of large, detailed lexicons of Dutch, German, and English, and the PAROLE project (*www.hltcentral.org/projects/PAROLE*) and its successor SIMPLE (Lenci et al. 2000), which are large, rich lexicons for 12 European languages.

Two important sources for obtaining lexicons are these:

**ELDA:** The Evaluations and Language resources Distribution Agency (*www.elda.fr*) distributes many European-language general-purpose and domain-specific lexicons, both monolingual and multilingual, including PAROLE and EuroWordNet.

**LDC:** The Linguistic Data Consortium (*www.ldc.upenn.edu*), although primarily a distributor of corpora, offers CELEX and several other lexicons.

In addition, English WordNet is available free of charge from the project's Web page (*www.cogsci.princeton.edu/∼wn*).

## 1.2 Lexical entries

### 1.2.1 What's in a lexical entry?

Any detail of the linguistic behaviour or use of a word may be included in its lexical entry: its phonetics (including pronunciations, syllabification, and stress pattern), written forms (including hyphenation points), morphology (including inflections and other affixation), syntactic and combinatory behaviour, constraints on its use, its relative frequency, and, of course, all aspects of its meaning. For our purposes in this chapter, the word's semantic properties, including relationships between the meaning of the word and those of other words, are the most important, and we will look at them in detail in section 1.3.2 below.

Thus, as mentioned earlier, a lexical entry is potentially quite a large record. For example, the CELEX lexicons of English, Dutch, and German (Baayen, Piepenbrock, and van Rijn 1993) are represented as databases whose records have 950 fields. And in an **explanatory combinatorial dictionary** (ECD) (e.g., Mel'čuk 1984, Mel'čuk and Zholkovsky 1988), which attempts to explicate literally every aspect of the knowledge that a speaker needs to have in order to use a word correctly, lexical entries can run to many pages. For example, Steele's (1990) ECD-style entry for eight senses of *hope* (noun and verb) is 28 book-sized pages long, much of which is devoted to the combinatory properties of the word, such as that the noun *hope* permits *flicker of* to denote a small amount (whereas *expectation*, in contrast, does not).

Many linguistic applications will require only a subset of the information that may be found in the lexical entries of large, broad-coverage lexicons. Because of their emphasis on detailed knowledge about the linguistic behaviour of words, these large, complex lexicons are sometimes referred to as **lexical knowledge bases**, or **LKBs**. Some researchers distinguish LKBs from lexicons by regarding LKBs as the larger and more-abstract source from which instances of lexicons for particular applications may be generated. In the present chapter, we will not need to make this distinction, and will just use the term *lexicon*.

### 1.2.2 Inheritance of linguistic properties

Generally speaking, the behaviour of words with respect to many non-semantic lexical properties in any given language tends to be regular: words that are phonetically, morphologically, or syntactically similar to one another usually exhibit similar phonetic, morphological, or syntactic behaviour. For example, in English most verbs form their past tense with either *-ed* or *-d*, and even most of those that don't do so fall into a few small categories of behaviour; and quite separately, verbs also cluster into a number of categories by their **alternation** behaviour (see section 1.4.3 below).

It is therefore possible to categorize and subcategorize words by their behaviour—that is, build an ontology of lexical behaviour—and use these categories to construct a lexicon in which each word, by default, inherits the properties of the categories and subcategories of which it is a member. Of course, idiosyncratic properties (such as many of the combinatory properties listed in an ECD) will still have to be specified in each word's entry. Inheritance of properties facilitates both economy and consistency in a large lexicon. A hierarchical representation of lexical knowledge with property inheritance is really just a special case of this style or method of knowledge representation. Accordingly, the inheritance of properties in the lexicon and the design of formal languages for the representation of lexical knowledge have been areas of considerable study (e.g., Briscoe, de Paiva, and Copestake 1992, Gazdar and Daelemans 1992; for an overview, see Daelemans, De Smedt, and Gazdar 1992; for the DATR language for lexical knowledge representation, see Evans and Gazdar 1996).

It should be clear that a hierarchical representation of similarities in lexical behaviour is quite distinct from any such representation of the **meaning** of words; knowing that *boy* and *girl* both take *-s* to make their plural form whereas *child* does not tells us nothing about the relationship between the meanings of those words. Relationships between meanings, and the hierarchies or other structures that they might form, are a separate matter entirely; they will be discussed in section 1.3.2.

### 1.2.3 Generating elements of the lexicon

Even with inheritance of properties, compiling a lexicon is a large task. But it can be eased by recognizing that because of the many regularities in the ways that natural languages generate derived words and senses, many of the entries in a lexicon can be automatically predicted.

For example, at the level of inflection and affixation, from the existence of the English word *read*, we can predict that (among others) *reading, reader, unreadable,* and *antireadability* are also words that should be in the lexicon, and in three out of these four cases we'd be right. Viegas et al. (1996) present a system of **lexical rules** that propose candidate words by inflection and affixation (an average of about 25 from each base form), automatically generating lexical entries for them; a lexicographer must winnow the proposals. In their Spanish lexicon, about 80% of the entries were created this way. But a lexicon can never anticipate all the nonce words and neologisms that are easily created from combinations of existing words in languages such as German and Dutch; additional word-recognition procedures will always be needed.

At the level of word sense, there are also regularities in the polysemy of words. For example, the senses of the word *book* include both its sense as a physical object and its sense as information-content: *The book fell on the floor; The book was exciting*. (A problem for natural language processing, which need not concern us here, is that both senses may be used at once: *The exciting book fell on the floor*.) In fact, the same polysemy can be seen with any word denoting an information-containing object, and if a new one comes along, the polysemy applies automatically: *The DVD fell on the floor; The DVD was exciting*. There are many such regularities of polysemy; they have been codified in Pustejovsky's (1995) theory of the **generative lexicon**. Thus it is possible to write rules that generate new lexical entries reflecting these regularities; if we add an entry for *DVD* to the lexicon as an information-containing object, then the other sense may be generated automatically (Buitelaar 1998). (*A fortiori*, the theory of the generative lexicon says that a purely enumerative lexicon—one that is just a list of pre-written entries—can never be complete, because the generative rules always permit new and creative uses of words.)

## 1.3 Word senses and the relationships between them

Most of the issues in the relationship between lexicons and ontologies pertain to the nature of the word senses in the lexicon and to relationships between those senses—that is, to the semantic structure of the lexicon.

### 1.3.1 Word senses

By definition, a **word sense**, or the "meaning" of a word, is a semantic object—a **concept** or **conceptual structure** of some kind, though exactly what kind is a matter of considerable debate, with a large literature on the topic. Among other possibilities, a word sense may be regarded as a purely **mental object**; or as a structure of some kind of **primitive units of meaning**; or as the **set of all the things in the world that the sense may denote**; or as a **prototype** that other objects resemble to a greater or lesser degree; or as an **intension** or **description** or **identification procedure**—possibly in terms of necessary and sufficient conditions—of all the things that the sense may denote.

Word senses tend to be fuzzy objects with indistinct boundaries, as we have seen already with the example of *open* in section 1.1.1 above. Whether or not a person may be called *slim*, for example, is, to some degree, a subjective judgement of the user of the word. To a first approximation, a word sense seems to be something like a category of objects in the world; so the word *slim* might be taken to denote exactly the category of slim objects, with its fuzziness and its subjectivity coming from the fuzziness and subjectivity of the category in the world, given all the problems that are inherent in categorization (**Ref to the chapter in the book that talks about**

**this**; see also Lakoff 1987). Indeed, some critics have suggested that word senses are *derived, created,* or *modulated* in each context of use, and can't just be specified in a lexicon (Ruhl 1989, Kilgarriff 1997).

Nonetheless, one position that could be taken is that a word sense *is* a category. This is particularly appealing in simple practical applications, where the deeper philosophical problems of meaning may be finessed or ignored. The problems are pushed to another level, that of the ontology; given some ontology, each word sense is represented simply as a pointer to some concept or category within the ontology. In some technical domains this may be entirely appropriate (see section 1.5.1 below). But sometimes this move may in fact make matters worse: all the problems of categorization remain, and the additional requirement is placed on the ontology of mirroring some natural language or languages, which is by no means straightforward (see section 1.4 below); nonetheless, an ontology may act as an interpretation of the word senses in a lexicon (see section 1.5.4 below).

In addition to the **denotative** elements of meaning that refer to the world, word senses also have **connotation**, which may be used to express the user's attitude: a speaker who chooses the word *sozzled* instead of *drunk* is exhibiting informality, whereas one who chooses *inebriated* is being formal; a speaker who describes a person as *slim* or *slender* is implying that the person's relative narrowness is attractive to the speaker, whereas the choice of *skinny* for the same person would imply unattractiveness.

### 1.3.2 Lexical relationships

Regardless of exactly how one conceives of word senses, because they pertain in some manner to categories in the world itself, **lexical relationships** between word senses mirror, perhaps imperfectly, certain relationships that hold between the categories themselves. The nature of lexical relationships and the degree to which they may be taken as ontological relationships are the topics of most of the rest of this chapter. In the space available, we can do no more than introduce the main ideas of lexical relationships; for detailed treatments, see Cruse (1986), Evens (1988), and Green, Bean, and Myaeng (2002).

The "classical" lexical relationships pertain to identity of meaning, inclusion of meaning, part–whole relationships, and opposite meanings. **Identity of meaning** is **synonymy**: Two or more words are synonyms (with respect to one sense of each) if one may substitute for another in a text without changing the meaning of the text. This test may be construed more or less strictly; words may be synonyms in one context but not another; often, putative synonyms will vary in connotation or linguistic style (as in the *drunk* and *slim* examples in section 1.3.1 above), and this might or might not be considered significant. More usually, "synonyms" are actually merely near-synonyms (see section 1.4.1 below).

The primary **inclusion** relations are **hyponymy** and its inverse **hypernymy** (also known as **hyperonymy**) (Cruse 1986, 2002). For example, *noise* is a hyponym of *sound* because any noise is also a sound; conversely, *sound* is a hypernym of *noise*.

Sometimes names such as *is-a* and *a-kind-of* are used for hyponymy and *subsumption* for hypernymy; because these names are also used for ontological categories, we avoid using them here for lexical relationships. The inclusion relationship between verbs is sometimes known as *troponymy*, emphasizing the point that verb inclusion tends to be a matter of “manner”; *to murmur* is *to talk* in a certain manner (Fellbaum 2002). Inclusion relationships are transitive, and thus form a **semantic hierarchy**, or multiple hierarchies, among word senses; words without hyponyms are leaves and words without hypernyms are roots. (The structures are more usually networks than trees, but we shall use the word *hierarchy* to emphasize the inheritance aspect of the structures.)

The **part–whole** relationships **meronymy** and **holonymy** also form hierarchies. Although they may be glossed roughly as *has-part* and *part-of*, we again avoid these ontologically biased terms. The notion of part–whole is overloaded; for example, the relationship between *wheel* and *bicycle* is not the same as that of *professor* and *faculty* or *tree* and *forest*; the first relationship is that of functional component, the second is group membership, and the third is element of a collection. For analysis of part–whole relationships, see Cruse (1986), Iris, Litowitz, and Evens (1988), or Pribbenow (2002).

Words that are opposites, generally speaking, share most elements of their meaning, except for being positioned at the two extremes of one particular dimension. Thus *hot* and *cold* are opposites—**antonyms**, in fact—but *telephone* and *Abelian group* are not, even though they have no properties in common (that is, they are “opposite” in every feature or dimension). Cruse (1986) distinguishes several different lexical relations of oppositeness, including **antonymy** of gradable adjectives, **complementarity** of mutually exclusive alternatives (*alive–dead*), and directional opposites (*forwards–backwards*).

These “classical” lexical relationships are the ones that are included in the WordNet lexicon. Synonymy is represented, as mentioned earlier, by means of synsets: if two words have identical senses, they are members of the same synset. Synsets are then connected to one another by pointers representing inclusion, part–whole, and opposite relations, thereby creating hierarchies.

In addition to the “classical” lexical relationships, there are many others, which may be broadly thought of as **associative** or **typicality** relations. For example, the relationship between *dog* and *bark* is that the former is a frequent and typical agent of the latter. Other examples of this kind of relationship include typical instrumentality (*nail–hammer*), cause (*leak–drip*), and location (*doctor–hospital*).

Synonymy, inclusion, and associative relations are often the basis of the structure of a **thesaurus**. While general-purpose thesauri, such as *Roget’s*, leave the relationships implicit, others, especially those used in the classification of technical documents, will make them explicit with labels such as *equivalent term*, *broader term*, *narrower term*, and *related term*.

## 1.4 Lexicons are not (really) ontologies

The obvious parallel between the hypernymy relation in a lexicon and the subsumption relation in an ontology suggests that lexicons are very similar to ontologies. It even suggests that perhaps a lexicon, together with the lexical relations defined on it, *is* an ontology (or is a kind of ontology in the ontology of ontologies). In this view, we identify word senses with ontological categories and lexical relations with ontological relations. The motivation for this identification is clear from the preceding discussion (section 1.3.2).

Nonetheless, a lexicon, especially one that is not specific to a technical domain (see section 1.5.1 below), is not a very good ontology. An ontology, after all, is a set of categories of objects or ideas in the world, along with certain relationships among them; it is not a linguistic object. A lexicon, on the other hand, depends, by definition, on a natural language and the word senses in it. These give, at best, an ersatz ontology, as the following sections will show.

### 1.4.1 Overlapping word senses and near-synonymy

It is usually assumed in a genus–differentia ontology that subcategories of a category are mutually exclusive. For example, if the category domesticated-mammal subsumes the categories dog and cat, among others, then dog ∩ cat is empty: nothing is both a dog and a cat. This is not always so for the hyponymy relation in lexicons, however; rather, two words with a common hypernym will often overlap in sense—that is, they will be **near-synonyms**.

Consider, for example, the English words *error* and *mistake*, and some words that denote kinds of mistakes or errors: *blunder, slip, lapse, faux pas, bull, howler,* and *boner*. How can we arrange these in a hierarchy? First we need to know the precise meaning of each and what distinguishes one from another. Fortunately, lexicographers take on such tasks, and the data for this group of words is given in *Webster's New Dictionary of Synonyms* (Gove 1973); an excerpt appears in Fig. 1.1; it lists both denotative and connotative distinctions, but here we need consider only the former. At first, we can see some structure: *faux pas* is said to be a hyponym of *mistake*; *bull, howler,* and *boner* are apparently true synonyms—they map to the same word sense, which is a hyponym of *blunder*. However, careful consideration of the data shows that a strict hierarchy is not possible. Neither *error* nor *mistake* is the more-general term; rather, they overlap. Neither is a hypernym of the other, and both, really, are hypernyms of the more-specific terms. Similarly, *slip* and *lapse* overlap, differing only in small components of their meaning. And a *faux pas*, as a mistake in etiquette, is not really a type of mistake or error distinct from the others; a faux pas could also be a lapse, a blunder, or a howler.

This example is in no way unusual. On the contrary, this kind of cluster of near-synonyms is very common, as can be seen in *Webster's New Dictionary of Synonyms* and similar dictionaries in English and other languages. Moreover, the differences between the members of the near-synonym clusters for the same broad concepts are different in different languages. The members of the clusters of near-synonyms

**Error** implies a straying from a proper course and suggests guilt as may lie in failure to take proper advantage of a guide . . .
**Mistake** implies misconception, misunderstanding, a wrong but not always blameworthy judgment, or inadvertence; it expresses less severe criticism than *error*.
**Blunder** is harsher than *mistake* or *error*; it commonly implies ignorance or stupidity, sometimes blameworthiness.
**Slip** carries a stronger implication of inadvertence or accident than *mistake*, and often, in addition, connotes triviality.
**Lapse**, though sometimes used interchangeably with *slip*, stresses forgetfulness, weakness, or inattention more than accident; thus, one says a *lapse* of memory or a *slip* of the pen, but not vice versa.
**Faux pas** is most frequently applied to a mistake in etiquette.
**Bull, howler,** and **boner** are rather informal terms applicable to blunders that typically have an amusing aspect.

**Fig. 1.1.** An entry (abridged) from *Webster's New Dictionary of Synonyms* (Gove 1973).

relating to errors and mistakes in English, French, German, and Japanese, for example, do not line up neatly with one another or translate directly (Edmonds and Hirst 2002); one cannot use these word senses to build an ontology of errors.

These observations have led to the proposal (Edmonds and Hirst 2000, 2002) that a fine-grained hierarchy is inappropriate as a model for the relationship between the senses of near-synonyms in a lexicon for any practical use in tasks such as machine translation and other applications involving fine-grained use of word senses. Rather, what is required is a very coarse-grained conceptual hierarchy that represents word meaning at only a very coarse-grained level, so that whole clusters of near-synonyms are mapped to a single node: their **core meaning**. Members of a cluster are then distinguished from one another by explicit differentiation of any of the **peripheral concepts** that are involved in the fine-grained aspects of their denotation (and connotation). In the example above, *blunder* might be distinguished on a dimension of severity, while *faux pas* would be distinguished by the domain in which the mistake is made.

### 1.4.2 Gaps in the lexicon

A lexicon, by definition, will omit any reference to ontological categories that are not **lexicalized** in the language—categories that would require a (possibly long) multi-word description in order to be referred to in the language. That is, the words in a lexicon, even if they may be taken to represent categories, are merely a subset of the categories that would be present in an ontology covering the same domain. In fact, every language exhibits **lexical gaps** relative to other languages; that is, it simply lacks any word corresponding to a category that is lexicalized in some other language or languages. For example, Dutch has no words corresponding to the English words *container* or *coy*; Spanish has no word corresponding to the English verb *to stab* 'to injure by puncturing with a sharp weapon'; English has no single word for the German *gemütlich* or for the French *bavure* 'embarrassing bureaucratic error'. On the face of it, this seems to argue for deriving a language-independent ontology from the union of the lexicons of many languages (as attempted by Emele et al. (1992)); but this is not quite feasible.

Quite apart from lexical gaps in one language relative to another, there are many categories that are not lexicalized in *any* language. After all, it is clear that the number of categories in the world far exceeds the number of word senses in a language, and while different languages present different inventories of senses, as we have just argued, it nonetheless remains true that, by and large, all will cover more or less the same "conceptual territory", namely the concepts most salient or important to daily life, and these will be much the same across different languages, especially different languages of similar cultures. As the world changes, new concepts will arise and may be lexicalized, either as a new sense for an existing word (such as *browser* 'software tool for viewing the World Wide Web'), as a fixed phrase (*road rage*), or as a completely new word (*demutualization* 'conversion of a mutual life insurance company to a company with shareholders', *proteomics*, *DVD*). That large areas remain unlexicalized is clear from the popularity of games and pastimes such as Sniglets ("words that don't appear in the dictionary but should") (Hall 1987) and Wanted Words (Farrow 2000), which derive part of their humour from the identification of established concepts that had not previously been articulated and yet are immediately recognized as such when it is pointed out.

But even where natural languages "cover the same territory", each different language will often present a different and mutually incompatible set of word senses, as each language lexicalizes somewhat different categorizations or perspectives of the world. It is rare for words that are **translation equivalents** to be completely identical in sense; more usually, they are merely cross-lingual near-synonyms (see section 1.4.1 above).

An area of special ontological interest in which the vocabularies of natural languages tend to be particularly sparse is the upper ontology (see Borgo et al., this volume). Obviously, all natural languages need to be able to talk about the upper levels of the ontology. Hence, one might have thought that at this level we would find natural languages to be in essential agreement about how the world is categorized, simply because the distinctions seem to be so fundamental and so basic to our biologically based, and therefore presumably universal, cognitive processes and perception of the world. But natural languages instead prefer to concentrate the richest and most commonly used parts of their vocabulary in roughly the middle of the hierarchy, an area that has come to be known as the **basic-level categories**; categories in this area maximize both informativeness and distinctiveness (Murphy and Lassaline 1997). A standard example: one is more likely to choose the word *dog* for *X* in the context *Be careful not to trip over the X* than *entity, living thing, animal, mammal*, or *Beddlington terrier*, even though the alternatives are equally ontologically correct. Certainly, all languages have words similar to the English *thing, substance*, and *process*; but these words tend to be vague terms and, even here, vary conceptually from one language to another. That this is so is clear from the difficulty of devising a clear, agreed-on top-level ontology, a project that has exercised many people for many years. That is, we have found that we cannot build a satisfactory top-level ontology merely by looking at the relevant vocabulary of one or even several natural languages; see, for example, the extensive criticisms by Gangemi, Guarino, and Oltramari (2001) of the top level of WordNet as an ontology. From this, we can conclude that the upper levels of the lexical hierarchy are a poor ontology.

### 1.4.3 Linguistic categorizations that are not ontological

And yet, even though natural languages omit many distinctions that we would plausibly want in an ontology, they also make semantic distinctions—that is, distinctions that are seemingly based on the real-world properties of objects—that we probably wouldn't want to include in an ontology. An example of this is semantic categorizations that are required for "correct" word choice within the language and yet are seemingly arbitrary or unmotivated from an ontological point of view. For example, English requires the division of vehicles into categories according to whether or not the vehicle can be viewed as a container (e.g., *bus* and *canoe* versus *bicycle*), and, if a container, whether or not one can stand up in it (e.g., *bus* versus *car* and *canoe*). The verb *board* may only be used with vehicles that are containers that one can stand in; the verb *ride in* rather than *ride on* may only be used with vehicles that are containers that one cannot stand in.

Often, the linguistic categorization is not even a reliable reflection of the world. For example, many languages distinguish in their syntax between objects that are discrete and those that are not: **countable** and **mass nouns**. This is indeed an important distinction for many ontologies; but one should not look in the lexicon to find the ontological data, for in practice, the actual linguistic categorization is rather arbitrary and not a very accurate or consistent reflection of discreteness and non-discreteness in the world. For example, in English, *spaghetti* is a mass noun, but *noodle* is countable; the English word *furniture* is a mass noun, but the French *meuble* and German *Möbel* are countable.

A particularly important area in which languages make semantic distinctions that are nonetheless ontologically arbitrary is in the behaviour of verbs in their **diathesis alternations**—that is, alternations in the optionality and syntactic realization of the verb's arguments, sometimes with accompanying changes in meaning (Levin 1993). Consider, for example, the English verb *spray*:

(1) Nadia sprayed water on the plants.
(2) Nadia sprayed the plants with water.
(3) Water sprayed on the plants.
(4) ∗The plants sprayed with water.

(The '∗' on (4) denotes syntactic ill-formedness.) These examples (from Levin 1993) show that *spray* permits the **locative alternation** (examples 1 and 2), with either the medium or the target of the spraying (*water* or *the plants*) being realized as the syntactic object of the verb, and the second case (example 2) carrying the additional implication that the entire surface of the target was affected; moreover, the agent of spraying (*Nadia*) is optional (the **causative alternation**) in the first case (example 3) but not the second (example 4).

In view of the many different possible syntactic arrangements of the arguments of a verb, and the many different possible combinations of requirement, prohibition, and optionality for each argument in each position, a large number of different kinds of alternations are possible. However, if we classify verbs by the syntactic alternations that they may and may not undergo, as Levin (1993) has for many verbs of English, we see a semantic coherence to the classes. For example, many verbs that denote the covering of a surface behave in the same manner as *spray*, including *daub, splash,* and *sprinkle*. Nonetheless, the semantic regularities in alternation behaviour often seem ontologically unmotivated, and even arbitrary. For example, verbs of destruction that include in their meaning the resulting state of an entity (*smash, crush, shatter*), fall into a completely different behaviour class from verbs that do not (*destroy, demolish, wreck*) (Levin 1993, p. 239).

Even what is perhaps the most basic and seemingly ontological distinction made by languages, the distinction between nouns, verbs, and other syntactic categories, is not as ontologically well-founded as it seems. From the viewpoint of **object-dominant** languages (Talmy 2000a) such as English (and the majority of other languages), we are used to the idea that nouns denote physical and abstract objects and events (*elephant, Abelian group, running, lunch*) and verbs denote actions, processes, and states (*run, disembark, glow*). But even within European languages, we find that occasionally what is construed as an action or state in one language is not in another; a commonly cited example is the English verb *like* translating to an adverb, a quality of an action, in German: *Nadia likes to sing: Nadia singt gern*. But there are **action-dominant** languages in which even physical objects are referred to with verbs:

> For example, in a situation in which English might say *There's a rope lying on the ground*, Atsugewi [a language of Northern California] might use the single polysynthetic verb form *ẃoswalak·a* … [This can] be glossed as 'a-flexible-linear-object-is-located on-the-ground because-of-gravity-acting-on-it'. But to suggest its nounless flavor, the Atsugewi form can perhaps be fancifully rendered in English as: "it gravitically-linearizes-aground". In this example, then, Atsugewi refers to two physical entities, a ropelike object and the ground underfoot, without any nouns. (Talmy 2000a, p. 46)

### 1.4.4 Language, cognition, and the world

All the discussion above on the distinction between lexicon and ontology is really nothing more than a few examples of issues and problems that arise in discussions of the relationship between language, cognition, and our view of the world. This is, of course, a Big Question on which there is an enormous literature, and we cannot possibly do more than just allude to it here in order to put the preceding discussion into perspective. Issues include the degree of mutual causal influence between one's view of the world, one's culture, one's thought, one's language, and the structure of cognitive processes. The **Sapir-Whorf hypothesis** or **principle of linguistic relativity**, in its strongest form, states that language determines thought:

> We dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organized by our minds—and this means largely by the linguistic systems in our minds. We cut nature up, organize it into concepts, and ascribe significances as we do, largely because we are parties to an agreement to organize it in this way—an agreement that holds throughout our speech community and is codified in the patterns of our language. The agreement is, of course, an implicit and unstated one, *but its terms are absolutely obligatory*; we cannot talk at all except by subscribing to the organization and classification of data which the agreement decrees. (Whorf 1940/1972)

> No two languages are ever sufficiently similar to be considered as representing the same social reality. The worlds in which different societies live are distinct worlds, not merely the same world with different labels attached. (Sapir 1929/1964)

These quotations imply a pessimistic outlook for the enterprise of practical, language-independent ontology (or even of translation between two languages, which as a distinct position is often associated with Quine (1960)); but conversely, they imply a bright future for ontologies that are strongly based on a language, although such ontologies would have to be limited to use within that language community. But taken literally, linguistic relativity is certainly not tenable; clearly, we can have thoughts for which we have no words. The position is more usually advocated in a weaker form, in which language strongly influences worldview but does not wholly determine it. Even this is not broadly accepted; a recent critic, for example, is Pinker (1994), who states bluntly, "There is no scientific evidence that languages dramatically shape their speakers' ways of thinking" (p. 58). Nonetheless, we need to watch out for the un-dramatic shaping.

From a practical standpoint in ontology creation, however, while an overly language-dependent or lexicon-dependent ontology might be avoided for all the reasons discussed above, there is still much in the nature of natural languages that can help the creation of ontologies: it might be a good strategy to adopt or adapt the worldview of a language into one's ontology, or to merge the views of two different languages. For example, languages offer a rich analysis in their views of the structure of events and of space that can serve as the basis for ontologies; see, for example, the work of Talmy (2000b), in analyzing and cataloguing these different kinds of views.

## 1.5 Lexically based ontologies and ontologically based lexicons

Despite all the discussion in the previous section, it is possible that a lexicon with a semantic hierarchy might serve as the basis for a useful ontology, and an ontology may serve as a grounding for a lexicon. This may be so in particular in technical domains, in which vocabulary and ontology are more closely tied than in more-general domains. But it may also be the case for more-general vocabularies when language dependence and relative ontological simplicity are not problematic or are even desirable—for example if the ontology is to be used primarily in general-purpose, domain-independent text-processing applications in the language in question and hence inferences from the semantic properties of words have special prominence over domain-dependent or application-dependent inferences. In particular, Dahlgren (1995) has argued for the need to base an ontology for intelligent text processing on the linguistic distinctions and the word senses of the language in question.

### 1.5.1 Technical domains

In highly technical domains, it is usual for the correspondence between the vocabulary and the ontology of the domain to be closer than in the case of everyday words and concepts. This is because it is in the nature of technical or scientific work to try to identify and organize the concepts of the domain clearly and precisely and to name them unambiguously (and preferably with minimal synonymy). In some fields of study, there is a recognized authority that maintains and publishes a categorization and its associated nomenclature. For example, in psychiatry, the *Diagnostic and Statistical Manual* of the American Psychiatric Association (2000) has this role. In botanical systematics, so vital is unambiguous communication and so enormous is the pool of researchers creating new names that a complex system of rules (Greuter et al. 2000) guides the naming of genera, species, and other taxa, and the revision of names in the light of new knowledge.

Obviously, the construction of explicit, definitive ontologies, or even explicit, definitive vocabularies, does not occur in all technical domains. Nor is there always general consensus in technical domains on the nature of the concepts of the domain or uniformity in the use of its nomenclature. On the contrary, technical terms may exhibit the same vagueness, polysemy, and near-synonymy that we see exhibited in the general vocabulary. For example, in the domain of ontologies in information systems, the terms *ontology*, *concept*, and *category* are all quite imprecise, as may be seen throughout this volume; nonetheless, they are technical terms: the latter two are used in a more precise way than the same words are in everyday speech.

However, in technical domains where explicit vocabularies exist (including glossaries, lexicons and dictionaries of technical terms, and so on, whether backed by an authority or not), an ontology exists at least implicitly, as we will see in section 1.5.2 below. And where an explicit ontology exists, an explicit vocabulary certainly does; indeed, it is often said that the construction of any domain-specific ontology implies the parallel construction of a vocabulary for it; e.g., Gruber (1993, p. 909):

"Pragmatically, a common ontology defines the vocabulary with which queries and assertions are exchanged among agents".

An example of a technical ontology with a parallel vocabulary is the Unified Medical Language System (UMLS) (e.g., Lindberg, Humphreys, and McCray 1993; *www.nlm.nih.gov/research/umls*). The concepts in the Metathesaurus component of the UMLS, along with their additional interpretation in the Semantic Net component, constitute an ontology. Each concept is annotated with a set of terms (in English and other languages) that can be used to denote it; this creates a parallel vocabulary. Additional linguistic information about many of the terms in the vocabulary is given in the separate Specialist Lexicon component. See Hahn (this volume) for more details of the UMLS.

### 1.5.2 Developing a lexically based ontology

It has long been observed that a dictionary implicitly contains an ontology, or at least a semantic hierarchy, in the genus terms in its definitions. For example, if *automobile* is defined as *a self-propelled passenger vehicle that usually has four wheels and an internal-combustion engine*, then it is implied that *automobile* is a hyponym of *vehicle* and even that automobile IS-A vehicle; semantic or ontological part–whole relations are also implied.

Experiments on automatically extracting an ontology or semantic hierarchy from a machine-readable dictionary were first carried out in the late 1970s. Amsler (1981), for example, derived a "tangled hierarchy" from *The Merriam-Webster Pocket Dictionary*; Chodorow, Byrd, and Heidorn (1985) extracted hierarchies from *Webster's Seventh New Collegiate Dictionary*. The task requires parsing the definitions and disambiguating the terms used (Byrd et al. 1987); for example *vehicle* has many senses, including *a play, role, or piece of music used to display the special talents of one performer or company*, but this is not the sense that is used in the definition of *automobile*. In the analysis of the definition, it is also necessary to recognize the semantically significant patterns that are used, and to not be misled by so-called "empty heads": apparent genus terms that in fact are not, such as *member* in the definition of *hand* as *a member of a ship's crew* (Markowitz, Ahlswede, and Evens 1986, Alshawi 1987). Perhaps the largest project of this type was MindNet (Richardson, Dolan, and Vanderwende 1998).

Often, the literature on these projects equivocates on whether the resulting hierarchies or networks should be thought of as purely linguistic objects—after all, they are built from words and word senses—or whether they have an ontological status outside language. If the source dictionary is that of a technical domain, the claim for ontological status is stronger. The claim is also strengthened if new, non–lexically derived nodes are added to the structure. For example, in The Wordtree, a complex, strictly binary ontology of transitive actions by Burger (1984), the nodes of the tree were based on the vocabulary of English (for example, to sweettalk is to flatter and coax), but names were manually coined for nodes where English fell short (to goodbadman is to reverse and spiritualize; to gorilla is to strongarm and deprive).

### 1.5.3 Finding covert categories

One way that a hierarchy derived from a machine-readable dictionary might become more ontological is by the addition of categories that are unlexicalized in the language upon which it is based. Sometimes, these categories are implicitly reified by the presence of other words in the vocabulary, and, following Cruse (1986), they are therefore often referred to as **covert categories**. For example, there is no single English word for things that can be worn on the body (including clothes, jewellery, spectacles, shoes, and headwear), but the category nonetheless exists "covertly" as the set of things that can substitute for $X$ in the sentence *Nadia was wearing (an) X*. It is thus reified through the existence of the word *wear* as the category of things that can meaningfully serve as the object of this verb.

Barrière and Popowich (2000) showed that these covert categories (or some of them, at least) can be identified and added as supplementary categories to a lexically derived semantic hierarchy (such as those described in section 1.5.2 above). Their method relies on the definitions in a children's dictionary, in which the language of the definitions is simple and, unlike a regular dictionary, often emphasizes the purpose or use of the definiendum over its genus and differentia; for example, *a boat carries people and things on the water*. The central idea of Barrière and Popowich's method is to find frequently recurring patterns in the definitions that could signal the reification of a covert category. The first step is to interpret the definitions into a conceptual-graph representation (see Sowa, this volume). Then, a graph-matching algorithm looks in the conceptual-graph representations for subgraph patterns whose frequency exceeds an experimentally determined threshold. For example, one frequent subgraph is

[X]←(agent)←[carry]→(object)→(person),

which could be glossed as 'things that carry people'. This pattern occurs in the definitions of many words, including *boat*, *train*, *camel*, and *donkey*. It thus represents a covert category that can be named and added to a semantic hierarchy as a new hypernym (or subsumer, now) of the nodes that were derived from these words, in addition to any other hypernym that they already had. The name for the covert category may be derived from the subgraph, such as carry-object-person-agent for the example above. The hierarchy thus becomes more than just lexical relations, although less than a complete ontology; nonetheless, the new nodes could be helpful in text processing. The accuracy of the method is limited by the degree to which polysemy can be resolved; for example, in the category of *things that people play*, it finds, among others, *music*, *baseball*, and *outside*, representing different senses of *play*. Thus the output of the method must be regarded only as suggestions that require validation by a human.

Although Barrière and Popowich present their method as being for general-purpose, domain-independent hierarchies and they rely on a particular and very simple kind of dictionary, their method might also be useful in technical domains to help ensure completeness of an ontology derived from a lexicon by searching for unlexicalized concepts.

### 1.5.4 Ontologies for lexicons

As mentioned in section 1.3.1, most theories of what a word sense is relate it in some way to the world. Thus, an ontology, as a non-linguistic object that more-directly represents the world, may provide an interpretation or grounding of word senses. A simple, albeit limited, way to do this is to map between word senses and elements of or structures in the ontology. Of course, this will work only to the extent that the ontology can capture the full essence of the meanings. We noted in section 1.5.1 above that the UMLS grounds its Metathesaurus this way.

In machine translation and other multilingual applications, a mapping like this could act as an interlingua, enabling the words in one language to be interpreted in another. However, greater independence from any particular language is required; at the very least, the ontology should not favour, say, Japanese over English if it is to be used in translation between those two languages. In the twelve-language SIMPLE lexicon (Lenci et al. 2000), a hand-crafted upper ontology of semantic types serves as an anchor for lexical entries in all the languages (Lenci 2001). The semantic types are organized into four **qualia roles**, following the tenets of generative lexicon theory (see section 1.2.3 above).

Hovy and Nirenburg (1992) have argued that complete language-independence is not possible in an ontologically based interlingua for machine translation, but some degree of **language-neutrality** with respect to the relevant languages can nonetheless be achieved; and as the number of languages involved is increased, language-independence can be asymptotically approached. Hovy and Nirenburg present a procedure for merging a set of language-dependent ontologies, one at a time, to create an ontology that is neutral with respect to each. Near-synonyms across languages (section 1.4.1 above) are just one challenge for this approach. (See also Hovy (1998) and Noy (this volume).)

## 1.6 Conclusion

In this chapter, we have discussed the relationship between lexicons, which are linguistic objects, and ontologies, which are not. The relationship is muddied by the difficult and vexed relationship between language, thought, and the world: insofar as word-meanings are objects in the world, they may participate in ontologies for non-linguistic purposes, but they are inherently limited by their linguistic heritage; but non-linguistic ontologies may be equally limited when adapted to applications such as text and language processing.

## Acknowledgements

The preparation of this chapter was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. I am grateful to Eduard Hovy, Jane Morris, and Nadia Talent for helpful discussions and examples.

## References


1.1 American Psychiatric Association (2000): *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR* (4th edition, text revision). American Psychiatric Association, Washington, DC.

1.2 Amsler, Robert A. (1981): A taxonomy for English nouns and verbs. *Proceedings of the 19th Annual Meeting of the Association for Computational Linguistics*, Stanford, 133–138.
*www.aclweb.org/anthology/P81-1030*

1.3 Alshawi, Hiyan (1987): Processing dictionary definitions with phrasal pattern hierarchies. *Computational Linguistics*, 13(3/4), 195–202.
*www.aclweb.org/anthology/J87-3001*

1.4 Baayen, Harald R., Piepenbrock, Richard, and van Rijn, H. (1993): *The CELEX Lexical Database. Dutch, English, German*. CD-ROM, Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

1.5 Barrière, Caroline and Popowich, Fred (2000): Expanding the type hierarchy with non-lexical concepts. In: Hamilton, Howard J. (editor), *Advances in Artificial Intelligence* (Proceedings of the 13th Biennial Conference of the Canadian Society for Computational Studies of Intelligence, Montreal, May 2000), Lecture Notes in Artificial Intelligence 1822, pages 53–68. Springer-Verlag, Berlin.
*link.springer.de/link/service/series/0558/tocs/t1822.htm*

1.6 Boguraev, Branimir and Pustejovsky, James (eds.) (1996): *Corpus Processing for Lexical Acquisition*. The MIT Press, Cambridge, MA.

1.7 Borgo, Stefano; Gangemi, Aldo; Guarino, Nicola; Masolo, Claudio; Oltramari, Alessandro; Schneider, Luc (2003): Foundational and top-level ontologies. This volume, XXX–XXX.

1.8 Briscoe, Ted; de Paiva, Valeria; and Copestake, Ann (editors) (1993): *Inheritance, Defaults, and the Lexicon*. Cambridge University Press.

1.9 Buitelaar, Paul (1998): CORELEX: An ontology of systematic polysemous classes. In Guarino, Nicola (ed.), *Formal Ontology in Information Systems*, IOS Press, 221–235.

1.10 Burger, Henry G. (1984): *The Wordtree*. The Wordtree, Merriam, KS.

1.11 Byrd, Roy J.; Calzolari, Nicoletta; Chodorow, Martin S.; Klavans, Judith L.; Neff, Mary S.; and Rizk, Omneya A. (1987): Tools and methods for computational lexicography. *Computational Linguistics*, 13(3/4), 219–240.
*www.aclweb.org/anthology/J87-3003*

1.12 Chodorow, Martin S.; Byrd, Roy J.; and Heidorn, George E. (1985): Extracting semantic hierarchies from a large on-line dictionary. *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*, Chicago, 299–304.
*www.aclweb.org/anthology/P85-1034*

1.13 Cruse, D. Alan (1986): *Lexical Semantics.* Cambridge University Press.

1.14 Cruse, D. Alan (2002): Hyponymy and its varieties. In Green, Bean, and Myaeng (2002), 3–21.

1.15 Daelemans, Walter; De Smedt, Koenraad; and Gazdar, Gerald (1992): Inheritance in natural language processing. *Computational Linguistics*, 18(2), 205–218
*www.aclweb.org/anthology/J92-2004*

1.16 Dahlgren, Kathleen (1995): A linguistic ontology. *International Journal of Human–Computer Studies*, 43(5/6), 809–818.

1.17 Edmonds, Philip and Hirst, Graeme (2000): Reconciling fine-grained lexical knowledge and coarse-grained ontologies in the representation of near-synonyms. *Proceedings of the Workshop on Semantic Approximation, Granularity, and Vagueness*, Breckenridge, Colorado.
*http://www.cs.toronto.edu/compling/Publications*

1.18 Edmonds, Philip and Hirst, Graeme (2002): Near-synonymy and lexical choice. *Computational Linguistics*, 28(2), 105–144.
*http://www.cs.toronto.edu/compling/Publications*
1.19 Emele, Martin; Heid, Ulrich; Momma, Stefan; and Zajac, Rémi (1992): Interactions between linguistic constraints: Procedural vs. declarative approaches. *Machine Translation*, 7(1/2), 61–98.
1.20 Evans, Roger and Gazdar, Gerald (1996): DATR: A language for lexical knowledge representation. *Computational Linguistics*, 22(2), 167–216.
1.21 Evens, Martha Walton (ed.) (1988): *Relational Models of the Lexicon*. Cambridge University Press.
1.22 Farrow, Jane (2000): *Wanted Words: From amalgamots to undercarments.* Stoddart.
1.23 Fellbaum, Christiane (1998): *WordNet: An electronic lexical database*. The MIT Press, Cambridge, Mass.
1.24 Fellbaum, Christiane (2002): On the semantics of troponymy. In Green, Bean, and Myaeng (2002), 23–34.
1.25 Gangemi, Aldo; Guarino, Nicola; and Oltramari, Alessandro (2001): Conceptual analysis of lexical taxonomies: The case of WordNet top-level. In Welty, Chris and Smith, Barry (eds.), *Formal Ontology in Information Systems: Collected papers from the Second International Conference*, ACM Press, 285–296.
1.26 Gazdar, Gerald and Daelemans, Walter (1992): Special issues on Inheritance. *Computational Linguistics*, 18(2) and 18(3).
*acl.ldc.upenn.edu/J/J92*
1.27 Gove, Philip B., (editor) (1973): *Webster's New Dictionary of Synonyms*. G. & C. Merriam Company, Springfield, MA.
1.28 Green, Rebecca; Bean, Carol A.; and Myaeng, Sung Hyon (eds.) (2002): *The Semantics of Relationships: An interdisciplinary perspective*. Kluwer Academic Publishers, Dordrecht.
1.29 Greuter, W.; McNeill, J.; Barrie, F.R.; Burdet, H.M.; Demoulin, V.; Filgueiras, T.S.; Nicolson, D.H.; Silva, P.C.; Skog, J.E.; Trehane, P.; Turland, N.J.; and Hawksworth, D.L. (editors) (2000): *International Code of Botanical Nomenclature (Saint Louis Code)*. Koeltz Scientific Books, Königstein.
1.30 Gruber, Thomas R. (1993): Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human–Computer Studies*, 43(5/6), 907–928.
1.31 Hahn, Udo (2003): Ontology reuse from UMLS. This volume, XXX–XXX.
1.32 Hall, Rich (1984): *Sniglets (Snig'lit): Any word that doesn't appear in the dictionary, but should.* Collier Books.
1.33 Hovy, Eduard (1998): Combining and standardizing large-scale, practical ontologies for machine translation and other uses. *Proceedings of the 1st International Conference on Language Resources and Evaluation (LREC)*, Granada, Spain.
*www.isi.edu/natural-language/people/hovy/publications.html*
1.34 Hovy, Eduard and Nirenburg, Sergei (1992): Approximating an interlingua in a principled way. *Proceedings of the DARPA Speech and Natural Language Workshop*, Hawthorne, NY.
*www.isi.edu/natural-language/people/hovy/publications.html*
1.35 Iris, Madelyn Anne; Litowitz, Bonnie E.; and Evens, Martha (1988): Problems of the part–whole relation. In Evens (1988), 261–288.
1.36 Kilgarriff, Adam (1997): "I don't believe in word senses." *Computers and the Humanities*, 31(2), 91–113.
*www.kluweronline.com/issn/0010-4817/*
1.37 Lakoff, George (1987): *Women, Fire, and Dangerous things: What categories reveal about the mind*. The University of Chicago Press, Chicago.

1.38 Lenci, Alessandro (2001): Building an ontology for the lexicon: Semantic types and word meaning. In Jensen, Per Anker and Skadhauge, Peter (eds.), *Ontology-based Interpretation of Noun Phrases: Proceedings of the First International OntoQuery Workshop*, University of Southern Denmark, 103–120.
*www.ontoquery.dk/publications/*
1.39 Lenci, Alessandro et al. (2000). SIMPLE: A general framework for the development of multilingual lexicons. *International Journal of Lexicography*, 13(4), 249–263.
1.40 Levin, Beth (1993): *English Verb Classes and Alternations: A preliminary investigation*. The University of Chicago Press, Chicago.
1.41 Lindberg, Donald A.B.; Humphreys, Betsy L.; and McCray, Alexa T. (1993): The Unified Medical Language System. *Methods of Information in Medicine*, 32(4), 281–289.
1.42 Markowitz, Judith; Ahlswede, Thomas; and Evens, Martha (1986): Semantically significant patterns in dictionary definitions. *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, New York, 112–119.
*www.aclweb.org/anthology/P86-1018*
1.43 Mel'čuk, Igor (1984): *Dictionnaire explicatif et combinatoire du français contemporain*. Les Presses de l'Université de Montréal.
1.44 Mel'čuk, Igor and Zholkovsky, Alexander (1988): The explanatory combinatorial dictionary. In Evens 1988, 41–74.
1.45 *The Merriam-Webster Pocket Dictionary.* G.&C. Merriam Company, Springfield, MA.
1.46 Murphy, Gregory L. and Lassaline, Mary E. (1997): Hierarchical structure in concepts and the basic level of categorization. In Lamberts, Koen and Shanks, David (eds.), *Knowledge, Concepts, and Categories*, The MIT Press, Cambridge, MA, 93–131.
1.47 Noy, Natalya F. (2003): Tools for mapping and merging ontologies. This volume, XXX–XXX.
1.48 Pinker, Steven (1994): *The Language Instinct*. William Morrow and Company, New York.
1.49 Pribbenow, Simone (2002): Meronymic relationships: From classical merology to complex part–whole relationships. In Green, Bean, and Myaeng (2002), 35–50.
1.50 Pustejovsky, James (1995): *The Generative Lexicon*. The MIT Press, Cambridge, MA.
1.51 Quine, Willard Van Orman (1960): *Word and Object*. The MIT Press, Cambridge, MA.
1.52 Richardson, Stephen D.; Dolan, William B.; and Vanderwende, Lucy (1998): MindNet: Acquiring and structuring semantic information from text. *Proceedings, 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics (COLING-98)*, Montreal, 1098–1104.
*www.aclweb.org/anthology/P98-2180*
1.53 Roget, Peter Mark. *Roget's Thesaurus*. Many editions and variant titles.
1.54 Ruhl, Charles (1989): *On Monosemy: A study in linguistic semantics*. State University of New York Press, Albany, NY.
1.55 Sapir, Edward (1929/1964): The status of linguistics as a science. *Language*, 5, 207–214. Reprinted in Mandelbaum, David G. (ed.), *Culture, Language, and Personality: Selected essays of Edward Sapir*, University of California Press, Berkeley.
1.56 Sowa, John F. (2003): Conceptual graphs. This volume, XXX–XXX.
1.57 Steele, James (1990): The vocable *hope*: A family of lexical entries for an explanatory combinatorial dictionary of English. In Steele, James (ed.), *Meaning–Text Theory: Linguistics, lexicography, and implications*, University of Ottawa Press, 131–158.
1.58 Talmy, Leonard (2000a): The relation of grammar to cognition. In Talmy 2000b, I-21–96.
1.59 Talmy, Leonard (2000b): *Toward a Cognitive Semantics*, two volumes. The MIT Press, Cambridge, MA.
1.60 Viegas, Evelyn; Onyshkevych, Boyan; Raskin, Victor: Nirenburg, Sergei (1996): From *submit* to *submitted* via *submission*: On lexical rules in large-scale lexicon acquisition. *Proceedings, 34th Annual Meeting of the Association for Computational Linguistics,*

Santa Cruz, 32–39.
*www.aclweb.org/anthology/P96-1005*

1.61 Vossen, Piek (editor) (1998): Special issue on EuroWordNet. *Computers and the Humanities*, 32(2–3), 73–251. Reprinted as a separate volume: *EuroWordNet: A multilingual database with lexical semantic networks.* Kluwer Academic Publishers, Dordrecht. *www.kluweronline.com/issn/0010-4817/*

1.62 *Webster's Seventh New Collegiate Dictionary* (1963): G.&C. Merriam Company, Springfield, MA.

1.63 Whorf, Benjamin Lee (1940/1972): Science and linguistics. *Technology Review*, 42(6), 227–31, 247–8. Reprinted in Carroll, John B. (ed.), *Language, Thought and Reality: Selected writings of Benjamin Lee Whorf*, The MIT Press, Cambridge, MA.