# IN SO MANY WORDS

# KNOWLEDGE AS A LEXICAL PHENOMENON

Willem Meijs & Piek Vossen
Amsterdam Dictionary Research Group
English Department, Amsterdam University
Spuistraat 210
1012 VT Amsterdam
The Netherlands
LINKS@ALF.LET.UVA.NL

## Abstract

Lexical knowledge is knowledge that can be expressed in words. Circular though this may seem, we think it provides a perfectly reasonable point of departure, for, in line with a long-standing philosophical tradition it posits communicability as the most characteristic aspect of lexical knowledge. Knowledge representation systems should be designed so as to fit lexical data rather than the other way round. A broad view of the possible scope of lexical semantics would thus be one which tries to chart out the systematic, generalizable aspects of word meanings, and of the relations between words, drawing on readily accessible sources of lexical knowledge, such as machine readable dictionaries, encyclopedias, and representative corpora, coupled with the kind of analytic apparatus that is needed to fruitfully explore such sources, for instance custom-built parsers to cope with dictionary definitions (Vossen 1990), statistical programs to deal with the distributional properties of lexical items in large corpora (Church & Hanks 1990) etc. At the same time this kind of massive data-acquisition should be made sensitive to the borders between perceptual experience, lexical knowledge and expert knowledge.

## 1 Introduction

Many psychologists are quite content to declare that intelligence is what you measure by means of an intelligence test. In a somewhat similar way one could say that lexical knowledge is the kind of knowledge that can be expressed in words. An advantage of this position is that you can turn it round: knowledge that cannot be expressed in words is not lexical knowledge. A further advantage is that it automatically relates lexical knowledge to something that lends itself to intersubjective inspection: linguistic utterances. And yet another advantage is that it tells us how and where to look for typical expressions of lexical knowledge within the welter of linguistic utterances. In ever widening concentric circles, we can inspect: general dictionaries as sources of the lexical knowledge that is associated with the 'common core' vocabulary; encyclopedias for additional knowledge associated with many of the items contained in the dictionaries and for lexically expressed information about all kinds of facts and events; specialized dictionaries for more expert knowledge in specific domains; manuals, treatises, newspaper articles, speeches, broadcasts and so forth for yet more details on anything and everything that humans find worth-while communicating about.

In this paper we want to defend this point of view by making clear that all knowledge is necessarily dependent on representation. According to Dik (1986, 1989) this representation takes the form of either perceptual images or verbal structures (section 2). In section 3, we want to demonstrate that knowledge encoded in verbal structures is reducible to grammatical elements and lexical items (grammatical knowledge and lexical knowledge respectively). The insights gained from our work with machine readable dictionaries (MRDs) the Longman Dictionary of Contemporary English, henceforth LDOCE, and the Van Dale dictionary of contemporary Dutch "Groot Woordenboek van het Hedendaags Nederlands", henceforth Van Dale, lead to the conclusion that lexical knowledge in the end is perceptual (in the form of primitive perceptual images) and grammatical (in the form of grammatical primitives). This makes all linguistically expressed knowledge ultimatily reducible to these primitives. Since only linguistic knowledge is communicable, part of all knowledge is thus private (at least the knowledge corresponding to perceptual images). The structure of lexical knowledge as it emerges from our work, suggests a view on lexical meaning as not defining the object a word stands for but as representing a typical (not necessarily unique) conceptualization of that object (section 4). This hypothesis about lexical knowledge leads to the conclusion (section 5) that there are differences between lexical knowledge and expert-domain knowledge, in that the objective of the former is the use of a word as a symbol for a conceptualization of an object, and the objective of the latter is to have knowledge on (all and any imaginable) properties of the object itself.

## 2 The representation of knowledge

Knowledge, of any kind, has to be stored in memory. What is memorizable can be distinguished in perceptual information and conceptual knowledge. What we perceive is of course influenced by the concepts we have and, the other way round, concepts are partially based on our perception. Still there is a clear difference between perceptual information and conceptual representation in that the former is tied to the capacities of the sensory-apparatus and the latter to cognitive limitations to process information. In order to reduce the cognitive load of continuously processing and storing an enormous amount of data, the incoming information from the sensory apparatus is categorized in terms of more general concepts which as much as possible abstract from irrelevant details. By storing clusters of perceived features in terms of their conceptual categories all redundant information can be 'forgotten'. This process of categorization has been described by Rosch (1977) in terms of two contrastive principles:

i. a concept aims at capturing as many features as possible shared by its instantiations; in other words to be as specific as possible.
ii. a concept aims at capturing as many instances as possible in order to reduce the number of needed categories; in other words to be as general as possible.

Ideal categories are in balance with regard to both principles, and hence provide the most efficient unit to store information. The cognitive system will therefore probably represent incoming information as much as possible in terms of these ideal categories. These conceptual categories thus constitute the building blocks of accumulating knowledge shaping, and being shaped, by the ungoing processing of information.

Any abstraction from actual experience, however, implies that there is some kind of representation system in which the concept is expressed. Without representation knowledge cannot exist because there is no way to generalize over the individual experiences. The category that represents all prototypical instances has to be 'named' somehow to be able to attach information to it that applies to all these instances. This dependence of knowledge or 'symbolization' has important consequences. As far as the above perceptual features are concerned there is evidence

that they are at least stored in terms of so-called prototype images (this also is intended to hold for perceptions as "smell", "emotions", "touch", "taste", "hearing", "movement"). These are a kind of mental pictures which can 'be seen or experienced' through 'the mind's eye' and from which even inferences can be drawn (e.g. mental rotation tasks of three-dimensional objects having a two-dimensional representation, Shepard and Metzler 1977). Not all knowledge, however, deals with perceptual features and this non-perceptual knowledge is therefore not bound to the perceptual apparatus of people. Whatever perceptual features are stored by some prototypical image or mental picture of a 'bird', many of the corresponding concepts, for instance the fact that it is an 'animal' and therefore shares the feature 'living' with all other 'animals', cannot be captured by a purely perceptual image. Having a prototypical image of an 'animal' probably does not even make sense (although there may be a stereotypical best example of it, probably some kind of mammal), because instances of perceived 'animals' are more likely categorized in terms of more specific (and therefore more informative) categories. Non-perceptual facts, generalizing over individual cases, have to be represented by some other conceptual language than perceptual images (other examples of non-perceptual knowledge are e.g. "social status", "kinship relations", "skills", "capacities"). Dik (1986, 1989, see also Weigand 1989, Meijs 1989) argues that any representation language for non-perceptual knowledge, in addition to perceptual images, must be verbal for the following reasons:

i. If there is a common conceptual language for all mankind, why are the actual natural languages so different from each other?
ii. How can we explain the great difficulties in translating if human beings are supposed to have a common universal conceptual language?
iii. Why should speakers of natural languages have two systems to represent their information, a universal conceptual language and a natural language, when almost any conceivable content can be expressed in natural language?
iv. In practice, any actual description of conceptual structures uses words of some specific natural language. If there is a language independent representation system then what are its symbols?

Although the symbols of such a knowledge representation are tied to the actual content words of languages, Dik claims that the representation structure in which the content words are captured is more abstract than the actual linguistic expressions of any one language. Such a structure should represent all the functional communicative and compositional aspects of linguistic expressions while at the same time it should abstract away from all language-dependent details of grammatical expression that are irrelevant from a conceptual point of view. For instance, the fact that the expression of information has a passive or active form is irrelevant because it does not affect the relation between entities designated by the expression. This point is nicely illustrated by the following two definitions in one and the same dictionary:

| | **Entry Word** | **Dictionary Definition** |
|---|---|---|
| a. | bray 2 0 n. | the sound that a donkey makes |
| b. | hee-haw 0 1 n. | the sound made by a donkey |

(Examples from LDOCE, 1978)

Similarly, once the semantic interpretation of function words such as determiners, relative pronouns, prepositions, and of inflection information has been established in terms of argument positions, functions and operators in the grammar, these elements can be omitted from the underlying structure. The following, for instance, is a Functional Grammar representation for both definitions:

a. [ d, s (x1) : $\text{sound}_N$ (x1) : { $\text{make}_V$ [ i, s (x2) : $\text{donkey}_N$ (x2) ]$_{\text{Ag Subj}}$ [ R (x1) ] $_{\text{Go Obj}}$ } (x1) ]
b. [ d, s (x1) : $\text{sound}_N$ (x1) : { $\text{make}_V$ [ i, s (x2) : $\text{donkey}_N$ (x2) ]$_{\text{Ag Obj}}$ [ (x1) ] $_{\text{Go Subj}}$ } (x1) ]

The structure as a whole designates a term with an entity reference index "x1". The "d" and "s" before the first occurrence of "x1" stand for term operators expressing definiteness and singularity respectively. Of "x1" two properties are stated: "donkey" and a complex property between braces. The complex property consists of an involvement of "x1" (expressed by the co-referentiality) in an event designated by the main predicate "make". The "x1" fills the second argument slot of both structures having the semantic function "Go" (Goal), whereas "x2" (classified as a "donkey") fills the first argument position of both structures ("i" and "s" are again term operators, "i" standing for indefinite), having the semantic function "Ag" (Agent). The only differences between the two structures are the assigment of the syntactic functions subject and object ("Subj" and "Obj") and the relative pronoun marker "R". These differences lead to the two different surface structures, but they do not affect the semantic interpretation of the structures, which is the same for both.

Note that this view of knowledge allows concepts to be represented by complex structures containing several word senses. Given the fact that the repertory of expressions in natural language is theoretically unlimited, there is also no theoretical limitation on the possibility of representing some concept in the underlying structure.

## 3 The cyclic nature of linguistically expressed knowledge

The advantage of taking a underlying linguistic-conceptual structure, which is claimed to be universal in Dik's theory of Functional Grammar (1989), as the format of representation is that it brings knowledge within the scope of so-called linguistic knowledge: i.e. knowledge is taken to be stored in verbal form. Any knowledge, even expert knowledge or specialized symbolic systems such as the 'language' of mathematics or logic, has to have some interface with natural language to make it communicable. To make the same point Weigand (1989) refers to the work of Gardner (1987), who has studied knowledge representation in the legal domain. She admits that despite the rather sharply defined technical terms, many items, at some point, have to be grounded in ordinary language. A law on 'dogs' may define the term "own" but it does not define the word "dog" (the so-called open texture problem; Hart 1961). The only solution is to relate expert or domain knowledge to general natural language. Once the tools for analyzing natural language in terms of its semantic content are available it is possible to squeeze knowledge out of any linguistic utterance regardless of whether it is uttered as expert knowledge, broadcasts, speeches, newspaper articles, treatises, or in manuals. For any utterance it will hold that its meaning can be explained in terms of the underlying linguistic-conceptual structure and the meaning of the content words that it contains:

| | | |
|---|---|---|
| Meaning (expression) | = | Meaning (underlying linguistic-conceptual structure)<br>+<br>Meaning (words) |

This means that expressions of expert knowledge in some natural language are subject to the same rules and restrictions imposed by the grammar on expressions of the language as a whole. If we look at the following example of the entry "water" in the Encyclopaedia Brittanica (1977) the language used to describe the knowledge is ordinary English, e.g.:

water 19:633, a familiar substance composed of the chemical elements hydrogen and oxygen and existing in vapour, liquid, and several solid forms. Water is essential to terrestrial life, participating

in virtually every process that occurs in plant and animal organisms. Although the molecules of water are simple in structure, the physical and chemical properties of the compound are extraordinarily complicated. ... etc....

Many words are familiar and common English words used in their non-expert meanings. The expert words used, such as "hydrogen" and "oxygen", can be looked up again, mostly in the same resource:

**hydrogen** (from Greek hydro- and genes, "water former"), a colourless, odourless, tasteless, flammable, gaseous substance, the simplest member of the family of chemical elements; ...

Schematically, all knowledge is thus necessarily related to linguistic knowledge in the following way:

| | | |
|---|---|---|
| Meaning (Expert word) | = | Meaning (Expert expression) |
| Meaning (Expert expression) | = | Meaning (underlying linguistic-conceptual structure)<br>+<br>Meaning (Expert words)<br>+<br>Meaning (Ordinary words) |

The meaning of the structure is defined in the grammar. This definition takes care of very basic notions and presuppositions such as 'entity', 'event', 'state', 'change of state', 'causality', 'control', 'location', 'time', 'semantic roles', 'quantification', 'countability', etc. The rest of the content is based on the meaning of the ordinary words to be found in the lexicon as lexical knowledge. Thus, to profit from the wealthy electronic resource of linguistically expressed knowledge that is accumulating every day, it will be necessary to get a hold on the semantics of linguistic utterances; that is to develop a grammar and provide a lexicon. After that everything deserving of the description 'knowledge' (except perceptual or picturized knowledge) is fair game for lexical semantic hunters.

In view of the essential character of knowledge as outlined above, it is not surprising that lexical knowledge is stored in the form of expressions in natural language as well in dictionaries (although some dictionaries additionally also use pictures). The same concentric arrangement as described above can thus be applied to the lexical knowledge contained in dictionaries:

| | | |
|---|---|---|
| Meaning (Ordinary word) | = | Meaning (Dictionary Expression) |
| Meaning (Dictionary Expression) | = | Meaning (underlying linguistic-conceptual structure)<br>+<br>Meaning (Ordinary words) |

The combined explanations and illustrations can thus be thought of as constituting a huge relational network or grid which must by necessity have a non-trivial correspondence with the knowledge that is being expressed. Furthermore most explanations follow a canonical format, so the relational network is also structured to a fairly high degree. The study of machine-readable dictionaries in the past decade or so has made it possible to trace the relational network inherent in dictionary-definitions in great detail. While these kind of studies have shown that the inherent network-organization is naturally far from perfect, they have also brought out that the overall contours are nevertheless clearly recognizable. When, for instance, this concentric principle is systematically applied to the genus words of dictionary definitions the words of a language thus

form hierarchical chains like the following examples from LDOCE:

| Entry Word | Sense | Dictionary definition |
|---|---|---|
| blue-bell | 1 | any of various blue bell-shaped *flowers* |
| flower | 1 | the part of a *plant*, often beautiful and coloured, that produces seeds or fruit |
| | 2 | a *plant* that is grown for the beauty of this part |
| plant | 1 | a living *thing* that has leaves and roots, and grows usu. in earth, esp. the kind smaller than trees |
| thing | 1 | any material *object* |
| object | 1 | a *thing* |

The entry words are decomposed via the genus terms in ever more general classes until, necessarily, a circularity occurs. Such chains, which have been generated for LDOCE nouns and verbs and Van Dale nouns, tied together form a semantic classification of the vocabulary of a language in terms of a small set of circularily defined words. At least as far as nouns are concerned the circular top of this hierarchy can be described in terms of a semantic typology (Vossen fc.). This means that all nouns finally end up in a few typological primitives (collectives, individuals, masses), defined in the grammar. This is of course not surprising in a dictionary like LDOCE (the machine-readable version of which we studied in the LINKS-project), which uses a restricted controlled vocabulary and starts from the stated aim of explaining the words which the user-learner may not know in terms of words which he does know. However, our investigation of the Van Dale Dictionary of Contemporary Dutch in the context of the ESPRIT-project 'ACQUILEX' has shown that the inherent organization of this dictionary shows a broadly similar pattern, and we have reason to assume that basically the same goes for any good general dictionary (cf eg Amsler 1980).

The relevance of these chains is not so much to get at the typological status of words but to be able to get at the properties or knowledge specified in the differentiae of the more general words in these chains, so that they can be inherited for the more specific words. The fact that a "blue-bell" is "animate", for instance, is inherited via "flower" from the properties described in the definition of "plant". At the highest level in these hierarchies the definitions become rather void and no new properties are added. At this point lexical knowledge has been dissolved in all the properties specified by the differentiae along the way, and the semantic primitive at the end. If we look at the differentiae then they can basically be divided into stated properties and involvements in events. The latter facts can be represented in the underlying linguistic-conceptual structure as the fillers of particular slots of the predicate designating that event (in the above examples second argument of "produce", and "grow", first argument of "have"). These slots can have various functions or roles, all of which are defined in terms of the semantics of the grammar (see section 2 for an example of the underlying structure). If, on the other hand, we trace down the static properties that are expressed in the differentiae by looking up their definitions in LDOCE then it appears that they often designate perceptual features:

| **Properties:** | | |
|---|---|---|
| *blue* | 1 | having the colour blue |
| colour | 1 | the quality which allows the eyes to see the difference between (for example) a red flower and a blue flower when both are the same size and shape |
| | 2 | red, blue, green, black, brown, yellow, white, etc. |
| *bell-shaped* | | not an entry in LDOCE |
| *bell* n. | 1 | a round hollow metal vessel, which makes a ringing sound when struck |

| | | |
|---|---|---|
| round | 1 | circular |
| | 2 | shaped like a ball |
| circular | 1 | round |
| hollow | 1 | having an empty space inside |
| *shape* n. | 1 | the appearance or form of something that is seen |
| form | 1 | shape; appearance |
| appearance | 2 | that which can be seen |
| | | |
| *live* v. | 1 | to be alive |
| alive | 1 | having life |
| life | 1 | the active force that makes those forms of matter (animals and plants) that grow through feeding and produce new young forms like themselves, different from all other matter (stones, machines, objects, etc.) |
| | | |
| *material* adj. | 1 | of or concerning matter or substance |
| substance | 1 | a material |
| matter | 1 | the material which makes up the world and everything in space which can be seen or touched, as opposed to thought or mind |
| material n. | 1 | anything from which something is or may be made |

The oddity of defining perceptual knowledge in terms of words is apparent from obvious circularities such as "blue = the colour blue", "colour = red, blue, green, black, brown, yellow, white, etc", "round = circular = round" (sic!). These properties at least justify a way of representing knowledge separate from the lexical means. The verbal representation of this knowledge can only be seen as a pointer to some non-verbal perceptual image. In the end all conceptual knowledge is thus reducible to relations between words (contained in a underlying linguistic-conceptual structure) which can in their turn be related to grammatical primitives and perceptual images. Figure 1 represents the elaboration of knowledge from elementary primitive concepts defined in the grammar and perceptual images, to lexical primitives, to all lexical items, to layman knowledge upto expert knowledge:

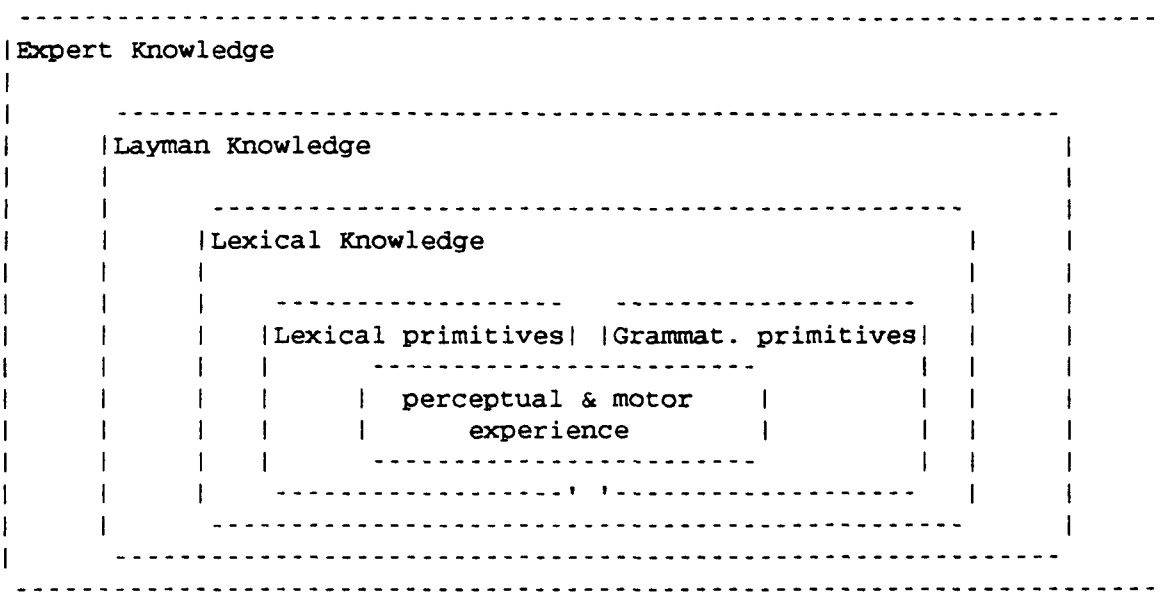


Figure 1. Layers of knowledge

The lexical primitives that end up in sensory-motor experiences represented by perceptual images and grammatical primitives, can be used to describe the higher order concepts designated by the other words of the lexicon. The vocabulary as a whole together with the grammar provides the equipment to represent any conceptual content within layman knowledge. Along the same lines, although more strict and systematic, expert knowledge is in its turn anchored in basic layman concepts.

Given the fact that only lexical knowledge is communicable, we have to conclude that some knowledge is private knowledge. This seems to be very reasonable since, although we can talk quite happily about our visual, auditory and tactile (etc.) experiences, it is ultimately impossible to actually communicate the essence of this kind of sensory-motor experience by lexical means. Instead, that requires exchanging the actual experience in some way or other (pointing to a colour, playing a tune, giving a blow, etc.). When it comes to the crunch, there is a soft centre at the heart of any kind of lexical definition. That soft centre is 'private meaning'. However tight we weave the web of words, we can never be certain that the concept one language-user associates with a particular word is exactly the same as that associated with it by another language-user. Dictionary definitions, of any kind, can thus only be approximative, never exhaustive.

## 4 The structure of lexical knowledge

Since all knowledge, in the end, reduces to relations between words, structural constraints on these relations entailed by the lexicon are relevant to any kind of knowledge. Each layer of knowledge elaborates on what is given by the inner area. Given the hypothesis that languages can express anything, there is no restriction in expressing any piece of knowledge so that the anchoring in lexical knowledge only leaves room for a very weak version of the Sapir & Whorf hypothesis (that language may influence thought and therefore knowlege). Despite that, we have evidence that the synchronic results of the diachronic process of knowledge acquisition and lexicalization illustrate that there tendencies have played a role in the development of knowledge between lexical and

expert knowledge. In a series of projects at the Univerisity of Amsterdam (ASCOT, LINKS) the syntactic and semantic information in LDOCE (and more recently in the framework of the Esprit project Acquilex, also the information contained in the Van Dale dictionaries of Contemporary Dutch) has been analysed and parsed (Akkerman e.a. 1988, Vossen 1990), and subsequently stored in a database, so that it is now possible to have systematic access to this knowledge. From studying the classificational structures (see previous section) that arise from it on a large scale and cross-linguistically (Vossen 1991, Vossen fc.) several conclusions on the structure of lexical knowledge can be drawn. The main point to be made is that the way in which words are classified can vary, as illustrated in figure 2, in which some of the relations at the top of the noun-hierarchy from LDOCE are given:

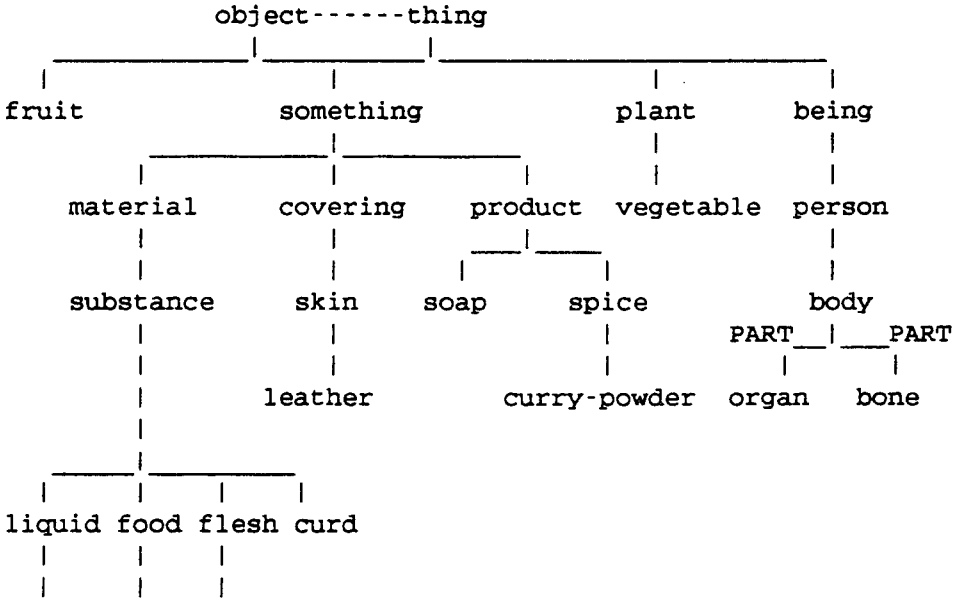


Figure 2. Hierarchical relations in LDOCE

Not all 'substances' are also found below "substance": "soap", "spice" and "curry-powder" are classified as 'products' instead. A similar thing holds for "skin" and "flesh" which are primarily not defined as 'bodypart' but as "covering" and "substance" respectively. As for 'foods' what is 'eatable' is also stored as "fruit", "vegetable", "liquid", "flesh", or directly as "substance" , which do not directly relate to the node "food". These differences in classifying words via their definition can be mainly traced back to two different modes of defining in dictionaries. Either the involvement of the designated object in some event is stressed, resulting in classes such as "food", "covering", "product" (in most cases of concrete non-animate objects its function), or a constitutional view is taken, as is the case with genus terms such as "substance", "liquid", "PART-OF body", "plant" ,"fruit", etc..

If we compare the classifications in LDOCE with their equivalents (equivalent according to the bilingual Dutch-English dictionary of Van Dale) in the hierarchy based on the Van Dale definitions, then these differences become even more striking. Many English words that are defined as some "substance" in LDOCE (e.g. "chocolate") have equivalent counterparts in Dutch (e.g. "chocolade") which are primarily defined in terms of their function ("genotmiddel" best translatable as "means of enjoying by ingesting"). The word "middel" (equivalent to English "means"), functions as a major node in the Dutch taxonomy, below which there are major compounds such as "vervoermiddel" ("means of transport"), "hulpmiddel" ("means of help" or "instrument"), "medicijnmiddel" ("means of curing" or "medicine"), "voedingsmiddel" ("means of

nutrition" or "food"). In English the word "means" has only been used for a very small set of words such as "mnemonics", "access", "approach" (the total number of word-senses related to it is 31, whereas thousands of senses are related to the Dutch "middel"), and the kind of compounds which naturally lead to "middel" in Dutch do not exist in English. Instead of that English has different words that do not so ostensibly lead to a functional classification. Figure 3 illustrates the difference between the two hierarchies:

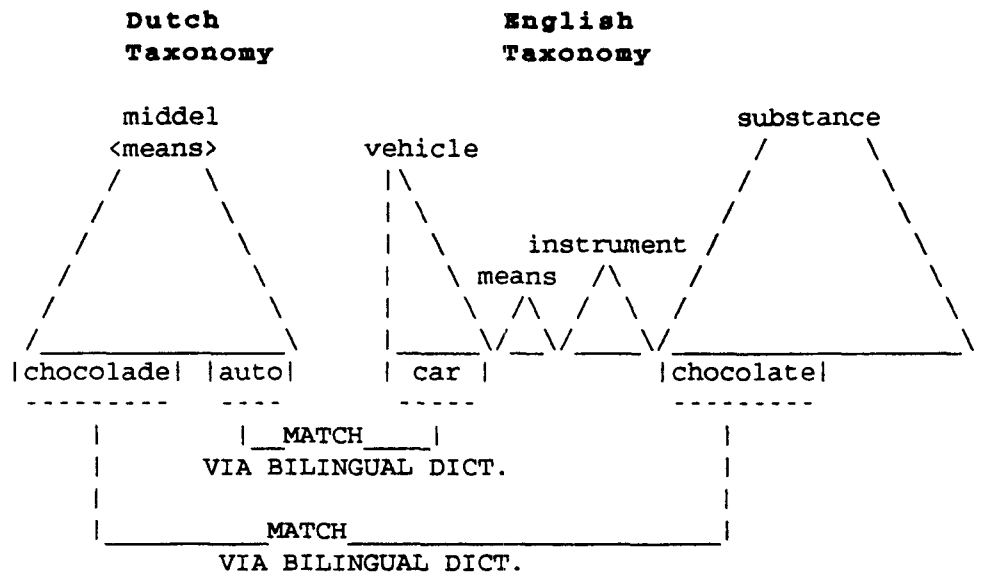


Figure 3. Hierarchical differences between LDOCE and Van Dale

Although both languages are culturally and historically closely related, large parts of the Dutch and English hierarchies differ considerably because the lexicographer's equipment to describe the concepts - i.e. the vocabularies of the two languages and hence the conceptualizations enforced by them - differ in crucial respects. The Dutch lexicographers have probably been triggered by the "middel" compounds to relate them to the functional node "middel". English lexicographers, on the other hand, were faced with words (such as "vehicle", "substance", "medicine", "food") that do not evoke "means" as a word for classification.

There are two ways of dealing with this variation in defining. Either the lexicographers have been inconsistent and one way of defining should be preferred over the other (although this may be more difficult to formulate in some cases than in others), or the work of lexicographers, in principle, is to be taken seriously and the fact that these words are differently classified is meaningful (in practice, examples of both interpretations can be found). The latter view is most relevant for our discussion because it implies that lexical meaning is not always the same as equivalence according to a bilingual dictionary. Although Dutch "chocolade" and English "chocolate" are equivalent translation pairs, they still have different lexical meanings. For translation purposes, substitution of words in all contexts is a sufficient criterion. Words, however, can have the same referential meaning but still stand for different concepts (Frege's Sinn und Bedeutung). In this view the lexical meaning as described in dictionary definitions would be taken to stand for a particular conceptualization of a possible set of referents, whereas that same set of referents can also be conceptualized in many other ways. In this sense dictionary definitions do not 'define' the object of reference, that is give the list of properties that uniquely discriminate it from all other objects, but merely tell you that using this word means that you have to look at the

object from this perspective (conceptualize it in such and such way). Within such a view of lexical meaning, variation in classification as in the follwing examples makes sense:

| Entry Word | Dictionary Definition |
|---|---|
| armour | strong protective *covering* on fighting vehicles,... |
| blanket | a thick, esp. woollen *covering*... |
| carpet | heavy woven often woollen *material* for covering floors or stairs |
| daub | (a) soft sticky *material* for covering surfaces like walls |

(Examples from LDOCE, 1978)

When using the word "blanket" or "armour" to refer to some object you are told to look at it primarily as a "covering" and secondarily as some kind of "material", whereas "carpet" and "daub", as defined here, stress the constitution of the object instead of its function "covering".

## 5 Different objectives of lexical and expert-domain knowledge

The view on general lexical knowledge developed above makes it to a certain degree different from other types of knowledge. Expert knowledge, and to some extent also part of the layman knowledge deals with knowledge about the object, whereas in case of general lexical knowledge not knowledge about the object but knowing that a particular word stands for a particular conceptualization (that can be used to refer to a corresponding object) is the goal. This perfectly fits in with the lexicographer's tradition of describing the semantic properties of words, in which the 'use of words' has always prevailed and not the knowledge about the object. A word as such is no more than a vehicle of information within a particular context. This does not necessarily imply that the 'meaning of a word' (the information it carries over) equals all the knowledge people have on the object or class of objects that is designated by that word, as is obvious (at least since Galileo) from the following examples:

| Entry Word | Dictionary Definition |
|---|---|
| sunrise | the time when the sun is seen to appear after the night |
| sunset | the time when the sun is seen to disappear as night begins |

(Examples from LDOCE, 1978)

| | |
|---|---|
| sunrise | the apparent rising of the sun above the horizon |
| sunset | the apparent descent of the sun below the horizon |

(Examples from Webster, 1974)

| | |
|---|---|
| zonsopgang | opgaan van de zon |
| (sunrise) | (rising-above of the sun) |
| zonsondergang | het ondergaan van de zon |
| (sunset) | (the descending-below of the sun) |

(Examples from Van Dale, 1984)

The words "sunrise" and "sunset" will probably be used in collocations that trigger the conceptualization associated with the words and not the knowledge we have about "stars", "planets" and their "orbits". Exactly these collocations form the material for lexicographers to extract the lexical meanings of words.

Furthermore, words not only give different conceptualizations of the same object but can

also function as a vehicle of attitudinal and social information, as in the following synonymous variants found in LDOCE on the basis of similar definitions:

| Entry Word | Dictionary Definition | |
|---|---|---|
| bobby | *infml BrE* a policeman | |
| bull | *sl, esp. AmE* a policeman | |
| copper | *infml* a policeman | |
| flatfoot | *sl* a policeman | |
| peeler | *BrE old sl* a policeman | |
| pig | *sl* policeman | |
| cop | *infml* policeman | (Examples from LDOCE, 1978) |

All these words designate the same set of potential referents, but they all carry a different attitudinal or social signal along expressed by the labels "infml" (=informal), "sl" (=slang), "AmE" (=American English), "BrE" ( Brittish English), "old" (=obselete), which tell us in what kind of context the words can be found (or used) but which has little to do with the 'knowledge we have of policemen'.

As a consequence of these different objectives of lexical and expert-domain knowledge, the connection between the different layers of knowledge, as represented in figure 1, is not allways that smooth. For some domains in which the process of lexicalization has clearly been triggered by the straightforward acquisition of knowledge through science and education, such as animal and plant names, there may be a rather regular connection between lexical, layman and expert knowledge in which more and more detailed information is added. As such part of the expert knowledge has become general lexical knowledge as well over time (perhaps a process in parallel to that of "Gesunkenes Kulturgut"). However in many domains, and even in animal and plant names (e.g. compare the functional "watchdog", "sheepdog", "pet" with the different species of animal in terms of their constitution), other conceptualizations of objects cross-classify with e.g. expert knowledge. Lexical knowledge thus gives you a wider range of views than expert knowledge. Something in the external world can be looked upon from all kinds of perspectives correlating with different words. For instance, while in Dutch we can use either "theewater" (literally "tea water") or "koffiewater" (literally "coffee water") to refer to the water that we are putting on to boil depending on its intended use, it would still be $H_2O$ to the expert. Although the expert's knowledge of "water" is more specific, the linguistic equipment to refer to "water" illustrates a greater variation in possible perspectives. In this sense, expert knowledge can be seen as more detailled and systematic subtype of lexical knowledge from a more restricted perspective. Where we expect that expert knowledge is exhaustive and defining, lexical knowledge thus is typical and associative but hardly ever exhaustive. Only those properties are described which typically mark the conceptualization that is associated with it.

References:


Akkerman E., H. Voogt-van Zutphen, and W. Meijs,
"A computerized lexicon for word-level tagging", Ascot report No 2. In: J. Aarts and W. Meijs (eds) Language and Computers. No 1, Rodopi, Amsterdam, 1988.

Amsler R.A.,
"The structure of the Merriam-Webster Pocket Dictionary", Ph.D. Thesis Texas University, 1980.

Berlin B.,
"Speculations on the growth of etnobotanical nomenclature", In: Language in Society, I, 1972, pp. 51-86.

Copestake A.,
"An approach to building the hierarchical element of a lexical knowledge base from a machine readable dictionary", Computer Laboratory, University of Cambridge, 1990.

Church K.W., and P. Hanks,
"Word association norms, mutual information, and lexicography", Computational Linguistics,16/1, 1990, pp. 22-29.

Cruse D.A.,
"Lexical Semantics", Cambridge, CUP, 1986.

Davidson D.,
"The method of truth in metaphysics", In K.Baynes, J. Bohman & Th. McCarthy (eds), After Philosophy: End or Transformation?, Cambridge Mass.: MIT Press, 1987.

Dik S.C.,
"Stepwise Lexical Decomposition", Lisse: Peter de Ridder Press, 1978 [available through Working Papers in FG].

Dik S.C.,
"Linguistically Motivated Knowledge Representation", Working Papers in Functional Grammar, No. 9. Amsterdam, 1986.

Dik S.C.,
"The theory of functional grammar. Part I: The structure of the clause", Functional grammar series 9, Foris, Dordrecht, 1989.

Dummett M.,
"Can analytic philosophy be systematic?", In K. Baynes, J. Bohman & Th. McCarthy (eds), After Philosophy: End or Transformation?, Cambridge Mass.: MIT Press, 1987.

"Encyclopaedia Britannica", Micropaedia , 15th edition,1977.

Frawley W.,
"Discussion: in defense of the dictionary: a response to Haiman", In: Lingua 55, 1981, pp.53-61.

Frege G.,
"Ueber Sinn and Bedeutung", In: G. Potzi Funktion, Begriff, Bedeutung, Gottingen, 1962, pp. 40-65.

Gardner, A. von der Lieth,
"An artificial intelligence approach to legal reasoning", MIT Press, cambridge, MA, 1977.

Grice H.P.,
"Logic and conversation", In: P. Cole & J.L. Morgan (eds) Syntax and Semantics 3. Speech Acts, New York/London, 1975, pp. 41-58.

Haiman J.,
"Dictionaries and Encyclopedias", In: Lingua 50, 1980, pp. 329-357.

Haiman J.
"Discussion: Dictionaries and Encyclopedias again", In: Lingua 56, 1982, pp. 353-355.

Hart H.L.A,
"The concept of law", Claredon Press, Oxford, 1961.
Proctor P. (ed),
"The Longman dictionary of contemporary English", London, Longman, 1978.
Martin W., and G.A.J. Tops,
"Groot Woordenboek Engels-Nederlands, VanDale Lexicografie", Utrecht, 1984.
Martin W., and G.A.J. Tops,
"Groot Woordenboek Nederlands-Engels", VanDale Lexicografie, Utrecht, 1984.
Medin D.L., and E.E. Smith,
"Concepts and concept formation", In: Annual Review Psychology 35, 1984, pp.113-138.
Meijs W.J.,
"Spreading the word: knowledge-activation in a functional perspective", In J.H. Connolly & S.C. Dik (eds),Functional Grammar and the Computer, Dordrecht, Foris, 1989.
Rosch E.,
"Classification of real world objects: origins and representation in cognition", In: P.N. Johnson-Laird and P.C. Wason, Thinking: Readings in cognitive science. Cambridge, 1977, pp. 212-222.
Rosch E.,
"Linguistic Relativity", In: P.N. Johnson-Laird and P.C. Wason, Thinking: Readings in cognitive science. Cambridge, 1977, pp. 501-521.
Shepard R.N., and J. Metzler,
"Mental rotation of three-dimensional objects", In: P.N. Johnson-Laird and P.C. Wason, Thinking: Readings in cognitive science. Cambridge, 1977, pp. 532-536.
Sterkenburg J. van, and W.J.J. Pijnenburg,
"Groot woordenboek van hedendaags Nederlands", Van Dale Lexicografie, Utrecht, 1984.
Vossen P.,
" A Parser-grammar for the Meaning Descriptions of the Longman Dictionary of Contemporary English", Technical Report NWO, project no. 300-169-007, University of Amsterdam, 1990.
Vossen P.,
"The end of the chain: where does stepwise lexical decomposition lead us eventually?", In: Proceedings of the 4th Functional Grammar Conference, June 1990, Kopenhagen, Denmark, fc., [also available as Acquilex Working Paper No. 010, July 1990 , Esprit BRA-3030, University of Amsterdam].
Vossen P.
"Comparing noun-taxonomies cross-linguistically", Acquilex Working Paper No. 014, ESPRIT BRA-3030, University of Amsterdam, May 1991.
Weigand H.,
"Linguistically Motivated Principles of Knowledge Base Systems", Dordrecht, Foris, 1989.
Wilks Y., Fass D., Guo C., McDonald E., Plate T., Slator B.,
"Machine tractable dictionaries as tools and resources for natural language processing", In: Proceedings of the 12th Conference on Computational Linguistics, Budapest, 1988, pp. 750-755.
Woolf H.B.
"The Merriam-Webster Dictionary", G.&C. Merriam Co., New York, 1974.