

The Linguistic versus Cognitive Role of Classifying Nouns

Piek Vossen

0. Introduction

Semantic classifications play a role in both the organization of our world knowledge and our vocabulary. Lexicons with semantic information are likewise often organized as taxonomies in which specific words are related or decomposed to (a small set of) more general words. Conceptual knowledge bases are structured as networks in which redundancies are predicted by more general concepts for more specific concepts. Even though lexical knowledge and conceptual knowledge do not necessarily coincide, in practice, most databases do not make a clear distinction between these two types of knowledge. In semantic lexicons and dictionaries it is not clear what is being described -- knowledge of words or knowledge of worlds -- and conceptual databases often just store information for the same words in capital letters, suggesting that a definition of CAR is automatically different from a definition of "car".¹ Because no clear distinction is made in the role of the semantic information, it is also not clear what criteria are to be used to evaluate that information or, more specifically, what is the role of classification structures in these specifications.

In this paper I will argue that the linguistic role of 'classifying nouns' in communication and their cognitive role to capture redundancies in conceptual networks cannot be described in the same way. I will first discuss some of their linguistic semantic properties and show that these are not very well reflected by a strict cognitive function of these nouns as categories in a conceptual network. From this it follows that words such as "object", "liquid", "vehicle", "food", "medicine" will need a very different specification of the semantics when we consider their linguistic usage or the cognitive role of the concepts *object*, *liquid*, *vehicle*, *food*, *medicine* in capturing redundancies in a world-knowledge base. Next I will describe the organization of a lexical knowledge base which I am developing for English and Dutch nouns in which the lexical semantic level and a cognitive level are kept separate. The lexical semantic level indicates how these entities are conceptualized and the cognitive level specifies the cognitive entities a word can name. Such a dual model of semantics has the advantages that an exhaustive description of world knowledge is restricted to the entities at the cognitive level, that we have more flexibility to express differences in the range of cognitive entities a word can name (the denotation) and, finally, that the linguistic implicature associated with a word is described separately and explains its linguistic usage. In the final section I will then describe some of the results of studying the semantic organization of two systematically analyzed machine readable dictionaries (English and Dutch). From these findings it follows that the vocabulary of language is not organized as an analytically-based (cognitive) taxonomy as claimed by Berlin (1972) but shows very different lexicalization patterns which support the view that a linguistic account of the semantics of nouns should be kept separate from a cognitive account.

1. The linguistic usage of classifying nouns

In lexicons with semantic information, such as ordinary dictionaries and databases (e.g. the Princeton WordNet, Miller et al 1990), the meaning of words is defined in terms of other words. Likewise, the vocabulary can be semantically organized by relating the entry words to the syntactic heads of their definitions in their appropriate meaning, as in the following examples taken from the Longman Dictionary of Contemporary English (LDOCE, Procter 1978):

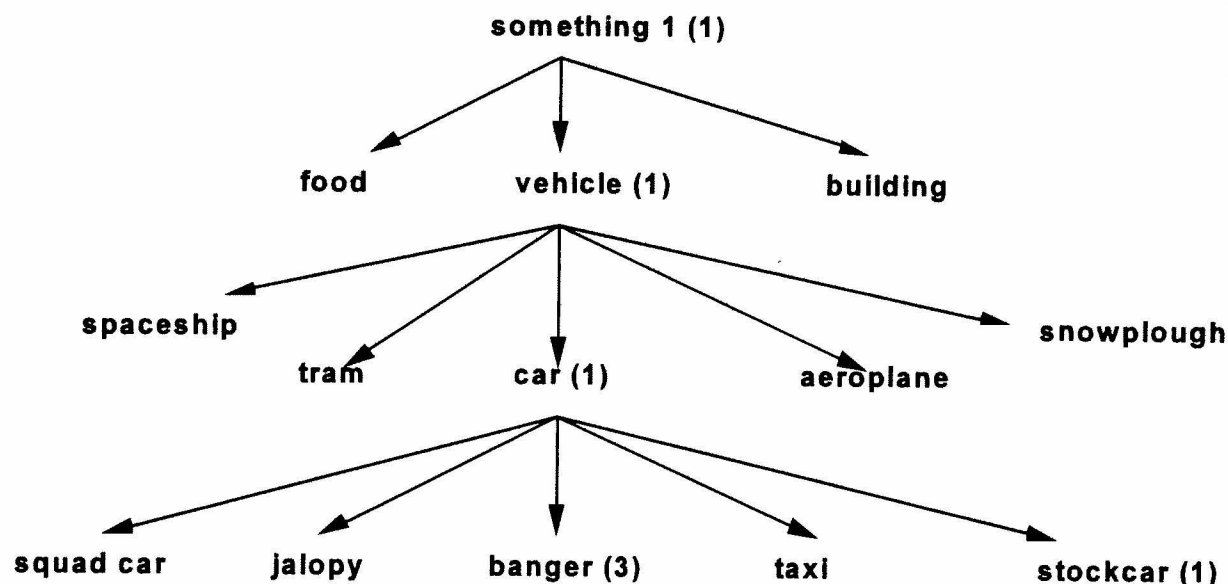
¹ From now on I will use double quotes (e.g. "car") to indicate that I refer to the word and italics (e.g. *car*) to indicate that I refer to the concept. So the word "car" can be used to refer to instances of the concept *car*. Other words such as "vehicle" and "traffic" can refer to similar instances of the concept *car*.

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

Entry Word	Definition LDOCE 1978
(1) banger (3)	an old <i>car</i> , esp. one not in very good condition; jalopy
car (1)	a <i>vehicle</i> with 3 or usu. 4 wheels and driven by a motor
vehicle (1)	<i>something</i> in or on which people or goods can be carried from one place to another
something 1 (1)	some unstated or unknown <i>thing</i>
thing (1)	any material <i>object</i>
object 1 (1)	a <i>thing</i>

By tracing these relations between the entry words, the whole vocabulary can be organized as a hierarchy which, at first sight, shows much resemblance to a classical taxonomy:

Taxonomy fragment derived from LDOCE 1978



The taxonomic relations between the words can then be used to gather the information expressed by the modifiers or differentiae in the definitions which can be inherited by default: i.e. for "banger" we not only can infer that it is *old*, and *in bad condition*, but also via "car" that it is *driven by a motor*, via "vehicle" that it is *used for transport*, and via "thing" that it is a *material*. These additional inferences can not only be made for "banger" but for all the words which are defined as "car". For each word we thus get an overview of the semantic relations it has with the other words in the vocabulary and a list of all the properties that can be associated with the concept named by the word. Because words are defined in terms of their relations with other words, this model of meaning is fully relational. From a cognitive point of view, the definition heads function as redundant categories that predict properties of more specific concepts. In this respect, a semantic lexicon can be seen as a conceptual network as well (Collins and Quillian 1969, Brachman 1979).

To what extent does this model then account for the linguistic usage of words? From a linguistic point of view, semantically related words can be described in terms of overlap in denotation. Denotational overlap is again defined in terms of substitutability of one predicate for another: whenever a word X can be

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

used to name some entity, its category or hyperonym Y can be used to name it (e.g. when something can be called "banger" it can also be called "car"). This principle is transitive so that we can substitute "banger" with "vehicle" and "thing" as well.

Which word from these hierarchical alternatives is the most appropriate is then said to be determined by the ambiguity of context in combination with the Gricean Maxim of Quantity (Grice 1975): do not specify more information than strictly necessary. In a situation where a noun is used by a Speaker to direct an Addressee to some intended referent out of several entities, the noun should be specific enough to avoid ambiguity without being unnecessarily informative. Suppose you want to buy a specific car from a dealer who sells various types of cars. The terms "the vehicle" or "the car" will never be appropriate to select the intended referent. Ambiguity can only be avoided by encoding sufficient information to select a unique individual in the given context. This can either be by locating the car you want (e.g. "that car/vehicle over there") or by supplying more properties such as "the black one", "the banger", etc. No more properties need to be coded than are required by the context. Various researchers (especially from the cognitive area) have suggested that words at the cognitive basic or generic level (e.g. the specificity level of "car") are most likely to be used to refer to an entity: i.e. in most contexts, more general nouns will be too ambiguous and more specific nouns will be unnecessarily specific or even incorrect in reference to a particular entity (Brown 1958, Berlin 1972).

This use of nouns in terms is dubbed the cognitive functional use by Cruse (1986) because it exploits the semantic properties of a word to pick out some referent. However, when there is only one vehicle in a context, such a cognitive use is no longer required (the non-functional identifying use). In such situations, Cruse observes that any noun in a taxonomy is odd except for words at the supposed generic or basic level:

- (2) I'm going to wash my
- | |
|----------|
| ?vehicle |
| car |
| ?banger |

In a context in which there is only one *vehicle* (or *car*) around to be washed (the thing in front of your house), "vehicle" will suffice to uniquely select the intended referent in the specific context. However, using "vehicle" in (2) is still odd. The same holds for "banger", even though it may very well be the case that the *car* in front of your house is *old and in poor condition*. From a cognitive functional point of view, "vehicle" would be sufficient and "banger" would match the entity best and is most informative (although perhaps less relevant). Nevertheless, the only unmarked usage in this case will be the noun "car" despite the fact that in this context it is too specific: we don't really need the information that it is a "car" to pick out the correct referent.

What makes "vehicle" and "banger" odd is that they imply a loading or perspective which is not required given the context in which the term is used. The noun "car" does not lead to an odd sentence because it is the most obvious name for the entity, while the context requires just that. Apparently, words such as "car" simply seem to name obvious things in the world without predicating any property of these things in particular. This explanation is confirmed by the different communicative effects of these nouns in terms used predicatively:

- (3)
- a This car is a banger
 - b This banger is a car
 - c This wreck/vehicle/menace/threat is a car
 - d This car is a wreck/menace/threat/vehicle

In 3(b) and 3(c) we the opposite effect that "car" is odd whereas the other predicates in 3(a) and 3(d) make

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

sense. The predicative effect of non-basic or non-generic words, such as "banger", "menace", "threat" and "wreck", is clearly recognized, whereas in the case of (3)b and (3)c it is not clear at all what we are predicating by using "car". Instead of predicating a property of the subject, these sentences seem to express the identity of some entity despite what is claimed by the subject term. Intuitively, the predicative effect of sentences (3)b and (3)c is somehow reversed, especially when "is" or "car" has stress.

One way of accounting for the implicature of meaning is by making a distinction between presupposition and assertion (Miller 1969, Fillmore 1971: 382). Given a classical definition in terms of a genus and differentiae, such as we have seen above, the genus can be seen as the presupposed concept, while the differentiae can be seen as the asserted properties of this concept. To take another example, "drinking water" can be described in terms of the genus "water" and the distinguishing feature *used for drinking*. The genus "water" is then the presupposition and its use is the assertion. The claim is then that by using the predicate "drinking water" we do not express that something is *water* and is *used for drinking* but we say *of water* that it *is used for drinking*. The distinction can thus be seen as a kind of prescription for when to use a word. The Speaker should first verify whether something in the situation falls within the class of things denoted by the presupposition. In the case of "drinking water", this would be something identified as "water". Secondly, s/he should consider the appropriateness of predicating or highlighting the asserted property of this thing. The distinction between presupposition and assertion perfectly explains the above use of specific predicates. When used predicatively, the presupposition can be seen as a selection restriction, while the assertion is the predicated property. In the case of non-functional identifying reference, on the other hand, the strongly implied property becomes a non-relevant property, possibly distracting the Addressee from the relevant message.

However, this distinction cannot be used for all words in a hierarchical chain. In the case of more general words the difference gets less clear. In LDOCE, "water" is defined as "the most common liquid, without colour, taste, or smell, which falls from the sky as rain, forms rivers, lakes, and seas, and is drunk by people and animals". We could interpret this information in a similar way as above. The presupposed concept would be "liquid" and the assertion would be formed by the other properties in the definition. The usage of "water" would be formulated as: to identify something in the context as a "liquid" and next see if it makes sense to predicate the asserted properties. In this case the distinction is less appealing:

- First of all, it is very difficult to exhaustively define so-called natural kind words such as "water". Defining the full meaning of such concepts comes close to storing world knowledge. The definition given in LDOCE characterizes it in terms of some typical properties but leaves out many others (such as its use to *clean*, *cook* or *extinguish fire*). It is doubtful whether any verbalized definition can fully capture the concept named by "water" (Quine 1960, Putnam 1975, Haiman 1980, Cruse 1986, Fillmore 1982, Wierzbicka 1980). Whereas the specific types of "water" seem to be distinguishable using a relatively limited set of features (therefore they are called nominal kind terms by Cruse (1986)), natural kind terms, e.g. "water" and "car", are said to differ in infinitely many properties.
- Secondly, it is not intuitively clear what we are predicating or asserting when we call something "water". None of the many properties associated with "water" seems to be asserted in particular, and, given some context, we can choose any conceptualization that fits the requirements (e.g. think of the many different uses of "water" in comparison to the single use implied by "drinking water").
- Thirdly, the presupposition "liquid" is not a good selection restriction for "water". It seems more difficult to determine that something is a "liquid" than determining that it is "water". Many people (e.g. little children) would not know how to identify a "liquid", but still know very well when to use the word "water". Instead of a presupposition, "liquid" rather represents some analytic classification or inferencing scheme.

Apparently, no obvious distinction between a necessarily implied presupposition and assertion can be made

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

for these natural kind terms. Their pragmatic usage is not to assert particular properties, but to name something which is identified in an independent way. For these reasons many researchers define natural kind terms as self-evident terms: a cat is an animal of the kind that we call "cats" (Wierzbicka 1980).

In the case of the more general word "liquid", the definition of the presupposition is even more problematic. A liquid is defined in LDOCE as "substance not solid or gas, which flows and has no fixed shape". Determining whether something is a *substance* comes close to a chemical analysis. Things get worse when we look at even more abstract presuppositions, such as "material", "anything", "thing", "object", to which such decomposition leads. On the other hand, the asserted part can be defined in a relatively satisfying way, e.g. applying the predicate "liquid" implies "fluidity". Another problem for more general words is that the notion of presupposition incorrectly predicts that they can be used as alternatives for all the words that strictly belong to the class of entities they represent (see the above substitution principle). This may be true at more specific levels: we can use "water" to refer to "drinking water" as well, but this does not seem to hold for the superordinates. We do not seem to use the word "substance" for all things that are substances in a strict sense but typically for *solid substances*. Referring to *gases* or *liquids* as "substances" is rather odd (where *liquids* are more acceptable than *gases*). Similarly, the word "object" could be defined as: a concrete thing having a definite outline or shape. In that case *animate* things, such as *animals*, *plants* and *humans*, are strictly speaking *objects*. However, using the word "object" for *animate beings* is also rather odd. Moreover, "object" seems to be restricted to *inanimate movable concrete things with a restricted volume and which have a definite outline*. Pursuing the distinction between presupposition and assertion at higher levels, therefore, leads to incorrect predictions about the substitutability of these words.

Intuitively, these abstract and general nouns do not presuppose the identification of a more general entity. The denotation is more naturally captured by listing the things that fall into the class it represents. In some cases these may be very diverse or open; in other cases they can be described in terms of a limited more homogeneous list. Rather than describing the use of superordinates as asserting properties of even more general superordinates, it is better to describe them as generalizing this property over more specific concepts. Just as with specific concepts, the meaning of superordinates can therefore be defined in a relatively satisfying way in terms of a generalization, although the presupposition is stated in a different way. This usage of superordinate nouns is especially clear when they occur as the head of an indefinite term used to introduce an entity in discourse:

- (4) a There was some liquid
 b There was something like water

The term in (4)a should either be interpreted as:

- the Speaker has perceived something which has not been fully determined but which is similar to other things which share the property "fluid"; or
- the Speaker considers information about the specific type of entity irrelevant or distracting and only wants to stress the *fluid* character of something.

As such, there is strictly speaking no tangible information given to the Addressee with which to identify instances of the concept except for the explicitly implied part of the meaning: "fluid". Only because we have knowledge about typical specific types of "liquid" can we build up some additional provisional representation. The identification, therefore, to some extent depends on the specific subtypes it classifies. Nevertheless, using the word "liquid" in (4)a is not equivalent to "something like water" in (4)b because the latter expression does not indicate in what respect it is similar to water (it could have been the *colour*, *taste*, *fluidity*, etc.), whereas "liquid" definitely (and only) implies the generalized property "fluid".

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

sense. The predicative effect of non-basic or non-generic words, such as "banger", "menace", "threat" and "wreck", is clearly recognized, whereas in the case of (3)b and (3)c it is not clear at all what we are predicating by using "car". Instead of predicating a property of the subject, these sentences seem to express the identity of some entity despite what is claimed by the subject term. Intuitively, the predicative effect of sentences (3)b and (3)c is somehow reversed, especially when "is" or "car" has stress.

One way of accounting for the implicature of meaning is by making a distinction between presupposition and assertion (Miller 1969, Fillmore 1971: 382). Given a classical definition in terms of a genus and differentiae, such as we have seen above, the genus can be seen as the presupposed concept, while the differentiae can be seen as the asserted properties of this concept. To take another example, "drinking water" can be described in terms of the genus "water" and the distinguishing feature *used for drinking*. The genus "water" is then the presupposition and its use is the assertion. The claim is then that by using the predicate "drinking water" we do not express that something is *water* **and** is *used for drinking* but we say **of water** that it *is used for drinking*. The distinction can thus be seen as a kind of prescription for when to use a word. The Speaker should first verify whether something in the situation falls within the class of things denoted by the presupposition. In the case of "drinking water", this would be something identified as "water". Secondly, s/he should consider the appropriateness of predicating or highlighting the asserted property of this thing. The distinction between presupposition and assertion perfectly explains the above use of specific predicates. When used predicatively, the presupposition can be seen as a selection restriction, while the assertion is the predicated property. In the case of non-functional identifying reference, on the other hand, the strongly implied property becomes a non-relevant property, possibly distracting the Addressee from the relevant message.

However, this distinction cannot be used for all words in a hierarchical chain. In the case of more general words the difference gets less clear. In LDOCE, "water" is defined as "the most common liquid, without colour, taste, or smell, which falls from the sky as rain, forms rivers, lakes, and seas, and is drunk by people and animals". We could interpret this information in a similar way as above. The presupposed concept would be "liquid" and the assertion would be formed by the other properties in the definition. The usage of "water" would be formulated as: to identify something in the context as a "liquid" and next see if it makes sense to predicate the asserted properties. In this case the distinction is less appealing:

- First of all, it is very difficult to exhaustively define so-called natural kind words such as "water". Defining the full meaning of such concepts comes close to storing world knowledge. The definition given in LDOCE characterizes it in terms of some typical properties but leaves out many others (such as its use to *clean*, *cook* or *extinguish fire*). It is doubtful whether any verbalized definition can fully capture the concept named by "water" (Quine 1960, Putnam 1975, Haiman 1980, Cruse 1986, Fillmore 1982, Wierzbicka 1980). Whereas the specific types of "water" seem to be distinguishable using a relatively limited set of features (therefore they are called nominal kind terms by Cruse (1986)), natural kind terms, e.g. "water" and "car", are said to differ in infinitely many properties.
- Secondly, it is not intuitively clear what we are predicating or asserting when we call something "water". None of the many properties associated with "water" seems to be asserted in particular, and, given some context, we can choose any conceptualization that fits the requirements (e.g. think of the many different uses of "water" in comparison to the single use implied by "drinking water").
- Thirdly, the presupposition "liquid" is not a good selection restriction for "water". It seems more difficult to determine that something is a "liquid" than determining that it is "water". Many people (e.g. little children) would not know how to identify a "liquid", but still know very well when to use the word "water". Instead of a presupposition, "liquid" rather represents some analytic classification or inferencing scheme.

Apparently, no obvious distinction between a necessarily implied presupposition and assertion can be made

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

for these natural kind terms. Their pragmatic usage is not to assert particular properties, but to name something which is identified in an independent way. For these reasons many researchers define natural kind terms as self-evident terms: a cat is an animal of the kind that we call "cats" (Wierzbicka 1980).

In the case of the more general word "liquid", the definition of the presupposition is even more problematic. A liquid is defined in LDOCE as "substance not solid or gas, which flows and has no fixed shape". Determining whether something is a *substance* comes close to a chemical analysis. Things get worse when we look at even more abstract presuppositions, such as "material", "anything", "thing", "object", to which such decomposition leads. On the other hand, the asserted part can be defined in a relatively satisfying way, e.g. applying the predicate "liquid" implies "fluidity". Another problem for more general words is that the notion of presupposition incorrectly predicts that they can be used as alternatives for all the words that strictly belong to the class of entities they represent (see the above substitution principle). This may be true at more specific levels: we can use "water" to refer to "drinking water" as well, but this does not seem to hold for the superordinates. We do not seem to use the word "substance" for all things that are substances in a strict sense but typically for *solid substances*. Referring to *gases* or *liquids* as "substances" is rather odd (where *liquids* are more acceptable than *gases*). Similarly, the word "object" could be defined as: a concrete thing having a definite outline or shape. In that case *animate* things, such as *animals*, *plants* and *humans*, are strictly speaking *objects*. However, using the word "object" for *animate beings* is also rather odd. Moreover, "object" seems to be restricted to *inanimate movable concrete things with a restricted volume and which have a definite outline*. Pursuing the distinction between presupposition and assertion at higher levels, therefore, leads to incorrect predictions about the substitutability of these words.

Intuitively, these abstract and general nouns do not presuppose the identification of a more general entity. The denotation is more naturally captured by listing the things that fall into the class it represents. In some cases these may be very diverse or open; in other cases they can be described in terms of a limited more homogeneous list. Rather than describing the use of superordinates as asserting properties of even more general superordinates, it is better to describe them as generalizing this property over more specific concepts. Just as with specific concepts, the meaning of superordinates can therefore be defined in a relatively satisfying way in terms of a generalization, although the presupposition is stated in a different way. This usage of superordinate nouns is especially clear when they occur as the head of an indefinite term used to introduce an entity in discourse:

- (4) a There was some liquid
 b There was something like water

The term in (4)a should either be interpreted as:

- the Speaker has perceived something which has not been fully determined but which is similar to other things which share the property "fluid"; or
- the Speaker considers information about the specific type of entity irrelevant or distracting and only wants to stress the *fluid* character of something.

As such, there is strictly speaking no tangible information given to the Addressee with which to identify instances of the concept except for the explicitly implied part of the meaning: "fluid". Only because we have knowledge about typical specific types of "liquid" can we build up some additional provisional representation. The identification, therefore, to some extent depends on the specific subtypes it classifies. Nevertheless, using the word "liquid" in (4)a is not equivalent to "something like water" in (4)b because the latter expression does not indicate in what respect it is similar to water (it could have been the *colour*, *taste*, *fluidity*, etc.), whereas "liquid" definitely (and only) implies the generalized property "fluid".

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

2. A lexical database organized around cognitive core concepts

The main problem with relational models of meaning is that a word is always defined in terms of its relations to other words and not to any notion of denotation outside the language-system. As such these models cannot explain why some predicates are better names for entities than others or why some predicates have a clearly distinguishable predicative loading while other predicates have not. To account for the linguistic usage of nouns, we thus not only have to distinguish the implicature of their meaning relative to some presupposition, but the presupposition has to be defined independently of the relations between the predicates.

The existence of such a level of (relatively) language-independent concepts is supported by extensive evidence in cognitive science. From a cognitive point of view, conceptual categories are needed to reduce the cognitive load of processing information by reducing the infinite differences between stimuli to behaviourally and cognitively usable proportions. Categories thus function as mini-theories generalizing over instances; these mini-theories serve to make predictions, to focus on other aspects than the redundant properties, and to overcome gaps, disturbances and mutilations of incoming information. According to Rosch (1977), there are two opposing principles that determine the value of concepts as categories for classifying incoming stimuli:

- The more properties that can be predicted on the basis of category membership, the better. This principle would lead to a large number of categories applying to a restricted number of instances.
- The fewer categories that are needed, the better. This leads to a small number of categories applying to as many instances as possible.

Furthermore, Rosch assumes that there is a basic level of categorization in which these opposing principles are in balance. This level "is the most general and inclusive level at which categories are still able to delineate real-world correlational structures" (Rosch 1977: 213). At this level, a significantly greater number of attributes are listed, whereas at the more general level only a few common attributes are given. More specific levels do not receive significantly more attributes.

Taking the findings together we can set up a model of meaning which is based on this *cognitive* notion of denotation. Instead of organizing the lexicon in a strictly analytic and taxonomic way in which concepts are explained by means of ever more abstract elements, it should be organized around a set of core concepts at a basic level of specificity, while the specific use of words should be described relative to those core concepts. As such, the fundamental elements of the lexicon will not be the most abstract primitives resulting from continuous decomposition (ad infinitum), but the most rich and self-evident concepts at a level where we tend to individuate things from a cognitive point of view (or given some Natural Language Processing (NLP) application).

The most important implications of this model are that words at the basic level do not decompose into more abstract concepts at the denotation level, but will simply name basic level concepts as such while more general, superordinate words will get a reverse dependency to more specific concepts rather than being decomposed in a more abstract presupposition:

Relative Decomposition model:

hyperonym/ hyponym --> hyperonym/ hyponym --> hyperonym/ hyponym-->

Conceptual dependency model:

subordinate --> basic concept <-- superordinate

By reversing the dependency of superordinates the conceptual dependency model is anchored at the basic level. Superordinates, such as "liquid", may then express a specific property (e.g. *fluidity*) but depend on the specific subtypes for other properties (e.g. *wetness*). To some extent even the properties captured by the generalization may ultimately depend on properties of its most stereotypical subtypes. Saying of *liquids*

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

that they "can flow" is not very helpful unless we know what "flow 1 (1)" is: "(of liquid) to run or spread smoothly" (LDOCE 1978). Not surprisingly, *flowing* is restricted to *liquids*, whereas "run and spread smoothly" do not really explain the meaning unless we know what *liquids* are. In general, there could be many things that "run or spread smoothly" which are not *liquid*. Verification that something may be called "a liquid" may thus still be based on a partial similarity with "water".

To account for these phenomena, I developed lexical representations for nouns which are likewise organized around a set of cognitive core concepts. These representations have been defined in a Lexical Knowledge Base (LKB, Copestake 1993) which was developed in the Acquilex-project² and which makes use of a unification-based, typed feature structures (compare Carpenter 1990). Typed feature structures consist of feature-value pairs and can be organized hierarchically, where feature-value pairs of more general types inherit to more specific types. The following feature structure (FS) is a simplified example of the general representation of concrete nouns, where features are listed in capitals before the colon, their values in bold after the colon and the FS is headed by its type-name in bold:³

² The Acquilex project was a joint enterprise of the Universities of Amsterdam, Barcelona, Cambridge, Pisa and Dublin and the publishers Cambridge University Press, Van Dale Lexicography and Bibliograf. It was funded by the EC (Esprit, BRA-3030 and BRA-7315) from 1989 to 1995.

³ When I refer to features and values in the text this typography will be taken over as well.

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

- (5) [concrete-noun
ORTH: **form**
DEN: **denotation**
CON: **conceptualization**]

Here the three features ORTH, DEN and CON capture information on the orthographic form, the denotation and the conceptualization that is associated with a predicate. At the denotation level, we specify the range of cognitive entities a predicate can name and, at the conceptual level, what the naming relation is with that denotation.

Words at the basic level will thus get a specification as follows, where there is no dependency on any other concept:

- (6) [basic-noun
ORTH: **water**
DEN: [**water**
TELIC: (**drink, clean, cook, extinguish fire**)
FLUIDITY: **true**
COLOUR: **colourless**
TASTE: **tasteless**
SHAPE: **shapeless**]
CON: **name**]

A word such as "water" simply names a self-evident cognitive entity in the world without highlighting or marking any property of this denotation. As suggested by many authors, these concepts are self-evident, given by axiom. The definition of this denotation is then a cognitive issue. Here, for simplicity, an arbitrary selection of properties is given (where TELIC represents all typical activities in which something may be involved; Pustejovsky 1991). In principle we should store a pointer to an external definition of the denotation. From a cognitive point of view, these concepts may be described in terms of perceptual images, sensory-motor movements, analytic epistemic models, etc.. From an NLP point of view, these concepts may be defined in terms of the ontology of some application. The following phenomena have been established for words at the basic level that function as neutral names:

- No assertion or generalization is expressed in the conceptualization, which means that no property of the denotation is implied more than any other property when the word is used to name some denotation.
- There is no presupposition of a more obvious or basic concept on which the denotation depends.
- There is no finite set of semantic traits or properties that exhaustively define the denotation.
- They are the best names in most contexts, which means that they normally do not impose a particular loading, conceptualization or clustering on what is perceived. This makes these nouns very suitable for identifying reference and less suitable for predicative, non-referring use.

More general and more specific words can now be defined as follows:

- | | | |
|-----|--|--|
| (7) | [superordinate-noun
ORTH: stem
DEN: den
CON: [superordinate
ASSERTION: proposition]] | [subordinate-noun
ORTH: stem
DEN: den
CON: [subordinate
ASSERTION: proposition]] |
|-----|--|--|

The asserted property of a superordinate and subordinate noun is stored at the conceptualization level (CON). These types of nouns then only differ in their dependency on some denotation. This information is

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

stored separately in the lexical knowledge base by means of so-called 'psort relations' and operators (Copestake 1993):

- = equality operator: two FSs have the same values
- < default inheritance operator: values are inherited from one FS to another by default
- V intersection operator: the intersection of two FSs is taken of all shared properties

By means of such a psort relation, the denotation of these predicates can be initialized by reference to some other predicate. In the case of a specific predicate there is a single cognitive entity to which the predicate is restricted and properties of this concept are inherited by default from this denotation: i.e. they can be overwritten by local features. This is expressed in the database using the default operator < as follows: <den> < basic-noun <den>, meaning that the values of the denotation of some predicate can be taken over by default from the denotational values found at the entry of some basic-noun. In the case of a superordinate, we will get a diverse range of concepts from which the intersection can be taken. This is expressed in the database using the default and the intersection operator as follows: <den> < (basic-noun₁ <den> V basic-noun₂ <den> V basic-noun_n <den>) <den>, where the denotation is initialized by the properties shared by a whole range of basic concepts_{1..n}.

An example of a subordinate noun is "drinking water". There is nothing which distinguishes *drinking water* from ordinary *water*, except the purpose *used for drinking*. The denotation specification of "drinking water" is thus fully determined by the DEN properties of the "water" entry given above in (6). This is expressed at the bottom of the left-hand entry by the psort relation with "water". The asserted property is stored as a value for TELIC. In the left column we thus find the lexical entry for "drinking water" as stored in the lexical knowledge base and in the right column we find the result of expanding the denotational dependency with the presupposition "water":

<p>(8) Unexpanded entry</p> <p>[subordinate ORTH: drinking water DEN:<1>[denotation TELIC: <0>] CON:[subordinate ASSERTION: [denotation TELIC:<0>drink]] <DEN> < water <DEN></p>	<p>Expanded entry</p> <p>[subordinate ORTH: drinking water DEN:<1>[water TELIC: <0> FLUIDITY: true COLOUR: colourless TASTE: tasteless SHAPE: shapeless] CON: [subordinate ASSERTION: [denotation TELIC:<0> drink]]</p>
---	--

Naturally, what is claimed at the DEN level should be in correspondence with the conceptualization implied. In some cases this may overwrite the default properties of the presupposition; in other cases it may simply add unspecified information. The semantic properties implied by the assertion should thus be related to the values filled in for DEN. In this example, this is indicated by the 'reentrancy of the value', as defined by Pollard and Sag (1987), at the denotational level with the conceptualized level (indicated by the 'reentrancy index': <0>). Reentrancy of values is used by Pollard and Sag to indicate that there is a correlation of values of different features.

As can be observed, all the knowledge we have of "water" (see the FS in (6) above) is available via the presupposition to the concept "drinking water" as well. However, in a normal inheritance-based database, the asserted purpose of "drinking water" will overwrite the default value of TELIC for "water": **drink, clean, cook, extinguish fire**. Since *drink* was also stored as a use of *water* in general, the relation between the specific concept and the more basic concept should rather be seen as a highlighting of a particular domain of interest from all the different epistemic models applying to *water*, as described by

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

Lakoff (1986). Other uses of *water* such as *extinguishing fire*, *cleaning* or *cooking* are not excluded by conceptualizing it as "drinking water". Unfortunately the LKB cannot deal with such a highlighting relation. No distinction is made between inherited properties and local properties. The latter always overwrite the former. In practice, such a rigid default inheritance seems to be the exception rather than the rule. Only in rather marginal cases, such as the classical *penguins* (which are birds that cannot fly), is knowledge truly overwritten. For some further discussion on the inheritance mechanisms in lexical knowledge bases, see Briscoe et al. (1992), Briscoe et al. (1994), Lascarides et al. (1994), Daelemans et al. (1992). The above representations account for the following facts:

- There is a clear distinction between presupposition and assertion.
- The asserted meaning can be defined in a relatively satisfactory way in addition to the presupposed concept.
- Even though these concepts give rich information (via the presupposition), they are inappropriate in situations in which the specifically implied properties are not relevant. In this respect they are less useful for identifying reference and particularly useful for predicative use.

For a superordinate such as "liquid" the lexical dependencies can now be described as follow:

<p>(9) Unexpanded "liquid" [superordinate ORTH: liquid DEN:[denotation FLUIDITY:<0>] CON:[superordinate ASSERTION: [denotation FLUIDITY:<0> true]]</p>	<p>Expanded "liquid" [superordinate ORTH: liquid DEN: [substance TELIC: telic FLUIDITY:<0> COLOUR: colour TASTE: taste SHAPE: shapeless] CON: [superordinate ASSERTION: [denotation FLUIDITY:<0> true]]]</p>
--	--

<DEN> <(water <DEN> V coffee <DEN> V petrol <DEN> V blood <DEN> V beer <DEN>) <DEN>

Just as with subordinates, the implied property of the generalization (FLUIDITY has the value **true**) is related to the denotation via reentrancy. The DEN properties are further initialized by taking the intersection of the DEN of the stereotypes, indicated in the bottom line by the intersection operator V. If we compare the properties of "coffee", "petrol", "blood", "wine" and "beer" with the previous properties of "water" we see that they only share the values **true** for FLUIDITY and **shapeless** for SHAPE. Only these values can then be filled in for *liquids* by default, with FLUIDITY being already implied by the generalization. The above representation of superordinates explains the following facts:

- There is a clearly definable implied meaning but not a clearly definable presupposed meaning.
- The implied meaning can be described in a relatively satisfactory way even though there is no easily identifiable presupposed concept.
- To some extent, the denotation can be described in terms of the intersection of the properties of the stereotypical subtypes or hyponyms.
- These nouns can typically be used:
 - to abstract from properties and stress the generalized property in case of identifying reference (to refer to "water" with the term "the liquid");
 - to predicate the generalized property in predicative use, e.g. "this car is a menace";
 - to introduce entities which have not fully been identified (e.g. "There is some liquid in it")

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

but which are considered sufficiently similar to the class members.

In this account, the only difference between subordinate and superordinate concepts is the range of denotation to which each applies. In this respect we can expect that there exist minimal pairs of words that express exactly the same conceptualization but only differ in range. A typical example is "drink", which carries the same implied conceptualization as the above "drinking water" but differs only in the fact that the presupposed concept is more general. Strictly speaking, anything which is drunk can be called a "drink", but it is practically restricted to *liquids*. The difference between the specific predicate "drinking water" and the general predicate "drink" is then only reflected in the fact that the former presupposes a single more basic concept, whereas the latter generalizes over *liquids*. Another example of such a pair is formed by the Dutch nouns "blusmiddel" (fire extinguishing agent) and "bluswater" (fire extinguishing water). These two nouns express the same conceptualization but differ only in the range of denotation as well. In this case, however, the more general noun "blusmiddel" is not restricted to *liquids* but is typically restricted to *foam* and *powders*. Here we see that the class to which the conceptualization can physically apply is larger than the denotation that is typically associated.

Using these representations for nouns, we can now organize the vocabulary as sets of predicates that can name the same or overlapping sets of cognitive core concepts. In this way we would get for each of these core concepts (e.g. *car* or *water*) an overview of the predicates that can name it (e.g. "car", "vehicle", "banger", "menace", "wreck", or "water", "drinking water", "dishwater", "rainwater", "liquid") together with a specification of the conceptualizations they imply. In this way, the linguistic lexicon and especially the classifying nouns in it are freed from the burden of providing an exhaustive list of semantic features or properties for each entry and access to this information can be gained by expanding the dependency to some range of denotation that is provided.

3 Lexicalization and classification patterns in dictionaries

Further support for such a non-classical organization of the lexicon has been found in machine readable dictionaries (MRDs). In several research projects the definitions of these MRDs have been parsed and the head of the definitions have been disambiguated so that the classification schemes, such as in (1), can automatically be generated for the whole dictionary. The inventory of these classification schemes and definitions structures yields two important results:

- Instead of a few top concepts the vocabulary consists of many relatively general and abstract predicates.
- The classification of similar concepts varies considerably in the same dictionary and across dictionaries.

These findings are discussed in more detail in the following two sections.

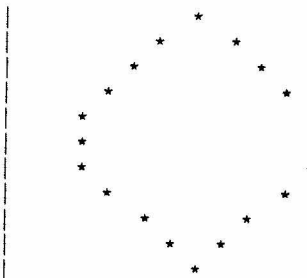
3.1 Lexicalization patterns of superordinates

A typical example of a classical organization of the vocabulary is described by Berlin (1972). He reports that folk taxonomies of animal and plant names in various languages exhibit at least three (and at most five) different levels, where most names for animals and plants are to be found at the generic level in the middle (500 categories in each language), and that there are considerably fewer life form and unique beginner terms. Furthermore, he claims that generic names are always morphologically unmarked, whereas specific and varietal names are always morphologically marked. We can schematically represent his taxonomic distribution of the lexicon as follows:

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

Taxonomic distribution of the vocabulary according to Berlin 1972

Thing
 Unique beginner: Plant
 life form: Tree
 generic name: Pine
 specific name: Ponderosa Pine
 varietal name: Northern Ponderosa Pine



The top of this picture suggests that there are a few superordinate or classifying nouns into which the rest of the vocabulary can be decomposed. If these claims can be extended to the whole vocabulary, we expect that the genus words or definition heads in dictionaries which are close to the top have a relatively low frequency and that relatively many of their children are also genus words of again more specific words. This is, however, not the case.

From inspecting the taxonomies automatically generated from the definitions in LDOCE and a Dutch dictionary (the Van Dale Dictionary of Contemporary Dutch, Sterkenburg and Tops 1984), it followed that we find many generalizing nouns in both languages. The following table gives the most frequent concrete nouns at the top of the taxonomy derived from LDOCE. After each word (most of which have an indication of their homograph and sense), the number of words is given which it classifies as a definition head (its direct children). The final column indicates how many of the direct children classify again other words (and are therefore non-leaf children):

Percentage of non-leaf children of concrete definition heads in LDOCE, where the average percentage of non-leaf words is 13%.

Definition Head	Total no. of direct children	Number of non-leaf children	Definition Head	Total no. of direct children	Number of non-leaf children
instrument (1)	122	31 25%	woman (1)	175	22 12%
money (1)	127	16 13%	system (1,2)	181	29 16%
container (1)	127	21 16%	apparatus (1,2)	204	25 12%
someone	130	8 6%	material	212	35 16%
room 1 (1)	131	24 18%	substance	239	47 19%
covering	133	33 25%	thing	275	34 12%
building (1)	143	28 19%	man 1 (1)	316	41 13%
people 1 (1)	143	16 11%	place	523	78 15%
machine 1 (1)	146	29 20%	something	869	63 7%
object	149	24 16%	person	2534	267 10%

If we look at the distribution of the heads (their frequency as head) we see that many of these still have no more than an average number of non-leaf children (13%). Clearly above average are "instrument", "covering", "room", "building", "machine" and "substance". Remarkably, however, in many other cases it does not seem to make much difference whether we are dealing with relatively specific words such as "woman" or "apparatus" or extremely vague cases such as "something" and "thing".

Given the fact that about 70% of all the noun senses are defined by the 400 most-frequent definition heads (Vossen 1995), we thus find a large proportion of the vocabulary very close to the top of the hierarchy although many of these words do not differentiate into more specific concepts. Especially in

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

the case of void heads such as "something" and "thing", the fact that many of their children are leaf words can only mean one of the following things:

- these leaves are also very general words;
- they have very rich definitions to distinguish rich concepts with the same head from each other; or
- they are inadequately defined by sparse definitions.

There is further evidence that these leaf words at the top of the hierarchy tend to be general words themselves. Many of these words have the following definition structure:

	Entry word	Definition LDOCE 1978
(11)	buzzer	a <i>thing</i> that buzzes
	stiffener	a <i>thing</i> which stiffens
	annoyance	<i>something</i> which annoys
	attraction	<i>something</i> which attracts

In all these cases, the specific information is expressed in a post-modifying verb phrase or clause in which some state of affairs is described. They are very often derived from verbs or adjectives one way or another, but, rather than referring to the quality or the event designated by the verb or adjective, in these senses they refer to any entity involved in the type of eventuality designated by the verb or adjective as one of its arguments. In this respect, most of these entry words defined by void *heads* can be described as cases of argument nominalizations. What these words have in common is that (almost) all of the semantics is expressed in the modifying phrase, whereas the definition head or genus is void. They can be defined in a satisfying way by co-reference of the denotation of the noun with the argument variable of the related verb or adjective. This means that these words express a specific conceptualization but can, in principle, apply to a variety of things: their denotation is often diverse or unrestricted. As stated above, this characterizes these words as superordinates. The denotation of some of these words is limited by the selection restriction of the verb on the argument: e.g. "drink" is restricted to "liquid". In such cases, the selection restriction can be seen as a denotational class to which the predicate is restricted. However, for many other nominalizations the restriction is either empty or, at best, filled in stereotypically. There is no restriction on the argument or the denotation variable and the usage and meaning of these nouns has to be accounted for in terms of the involvement they express. You can call anything an "acquisition"; you have to verify only that it has been acquired.

The denotational vagueness of argument nominalizations is also indicated by the fact that in many cases (40% of all head-occurrences of "thing") these heads are coordinated with completely different heads, e.g. "person or thing":

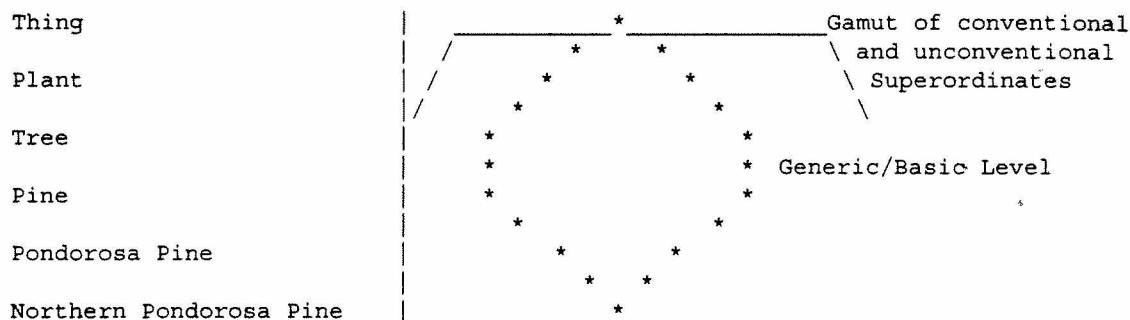
	Entry Word	Definition LDOCE 1978
(12)	arrival	a person who or thing that arrives or has arrived
	threat	a person, thing or idea regarded as a possible danger
	puzzler	a person who or thing that puzzles one

It is of little importance whether the entity denoted is human or not, as long as it fulfills the role described. The coordinated list of heads can be interpreted as indicating the stereotypical fillers of the slots. Strictly speaking, this means that the denotation of the defined word is more general than the denotation of the genus words: i.e. these definitions exhibit reverse conceptual dependency relations with their genus words, which, as we stated above, is typical of superordinates. In this respect, coordination of genus words as alternatives can be seen as the most natural way of accounting for the denotation of superordinates or words that express generalizations rather than specifications.

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

Given the relative frequency of these structures, we can conclude that the actual distribution of words at hierarchical top levels deviates considerably from the distribution suggested by Berlin's folk taxonomies. Whereas the latter started with a small set of unique beginners, language, in fact, has productive devices for generating unique beginners. As a result, we do not get a neat top but a whole range or gamut of superordinates at the top:

Vocabulary distribution at the top of the taxonomy in dictionaries



There are many rather general words in both English and Dutch that represent particular generalizations over concepts rather than naming specific things in the world. The size of a vocabulary can therefore not only be explained as the result of attaining deeper levels of detail or specialization but as also due to derivation and generalization⁴. Note also that Berlin's claim about morphological complexity is falsified. Many of the above argument nominalizations all have a morphologically complex structure and are still superordinate.

Some of these abstract words occur as conventional superordinates in classifications (e.g. "liquid", "substance" and "object"), but most superordinates are not conventionally used to define or describe other words, e.g. "threat", "menace", "waste", "valuables", being both leaf words and close to the top. Regardless of the circumstantiality of the predicated property, all these words are strict superordinates with respect to more specific words. The lexical-semantic test for hyponymy relations (Cruse 1986) succeeds in all cases:

- (13) a It's a *car* therefore it is a *threat/ menace/ waste*.
 b *Cars* and other *threats/ menaces/ wastes*.
 c *It's a *threat/ menace/ waste* therefore it's a *car*.
 d **Threats/ menaces/ wastes* and other *cars*.

Although *being a threat* is not the usual way of classifying *cars*, still it is a possible way of classifying them. Apparently, these superordinates provide alternative (non-conventional) ways of referring to analytic classes or diverse collections of objects and substances. Given the productivity of the superordinate classes, conventional analytic relations between superordinates are thus the exception rather than the rule. In biological or chemical classifications, we find deep taxonomies with many subclasses and types but only a few tops. Most superordinates in everyday language, however, are not ordered as deep taxonomies but form a flat level of alternatives which generalize over (often similar) things from very different

⁴ The extreme frequency of top level words could also be explained by the use of the controlled vocabulary in LDOCE. However, examination of a Dutch dictionary without a controlled vocabulary resulted in similar lexicalization patterns.

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

perspectives with void presuppositions, e.g. words such as "jewelry", "groceries", "threat", "waste".

3.2 Variation of classifications in dictionaries

Comparing the classification of similar concepts in the same dictionary and across dictionaries revealed that classifications vary considerably (Vossen 1991, Vossen 1995). For example, in the following definitions the same properties are expressed, but different choices have been made with respect to which property is captured by the head word and which property is expressed in the differentiae:

	Entry Word	Definition LDOCE 1978
(14)	armour	strong protective <i>covering</i> on fighting vehicles,...
	blanket	a thick, esp. woollen <i>covering</i> ...
	carpet	heavy woven often woollen <i>material</i> for covering floors or stairs
	daub	(a) soft sticky <i>material</i> for covering surfaces like walls

In the case of "armour" or "blanket" the function is derived from the head "covering"; in the case of the others, it is expressed in the differentiae. To some extent, this variation in classification is due to the non-empirical and non-scientific practice of making dictionaries. However, there are also more fundamental explanations for variation. First of all, we may expect conceptualization to vary at the basic level. Basic-level words simply name a very rich concept irrespective of the way it is conceptualized. Because there is a lot of information available, any property that applies can be taken as a point of departure for its classification, using all or any selection of the other properties to differentiate the concept from other concepts that share the classifying property. Variation could thus be seen as an indication that the concept is relatively rich and therefore also basic. To put it more strongly, because we can conceptualize these entities in various ways depending on the context, they should be defined in a flexible way, preferably differentiated for different epistemic models:

	Basic words	Conventional Superordinates
(15)	soup	liquid & food & artefact & grocery & etc..
	water	liquid & drink & fire extinguishing agent & cleaner etc..

This also implies that defining basic-level words in one particular way or another, as is done in dictionaries, can never be correct. For other words (superordinates, subordinates, pejoratives) the conceptualization is an essential part of their meaning and does not leave room for variation, e.g. "drink" and "drinking water".

Another value of the seemingly arbitrary classification of relatively specific words is that it tells us something about the superordinates themselves. Not all things that are *edible* are also called "food". Most edible things can be divided into two major classes: natural edible things and artificial edible things. Remarkably, the latter are almost always substances which are classified as food (in 16(a) "food" is the definition head), but the former are usually defined differently (in 16(b) the usage as food is expressed in the differentiae):

	Entry Word	Definition LDOCE 1978	
(16)	a	cheese (1) dish 1 (2) bread (1) pudding (1) cake 1 (1)	(any of the many kinds of) soft or firm solid <u>food</u> made from... (an amount of) cooked <u>food</u> of one kind a common <u>food</u> made of baked flour the sweet <u>dish</u> in a meal a <u>food</u> made by baking a sweet mixture of
	b	vegetable spinach	a (<u>part of a</u>) <u>plant</u> that is grown for food a type of widely-grown <u>vegetable</u> whose broad green leaves are eaten

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

asparagus	a <u>plant</u> whose young green stems are eaten as a vegetable
fruit 1 (1)	an <u>object</u> that grows on a tree or bush, contains seeds, is used for food, ...
(2)	these <u>objects</u> in general, esp. considered as food
meat	the <u>flesh</u> of animals, apart from fish and birds, which is eaten
milk	a white <u>liquid</u> produced by human or animal females for the feeding of their young, ...

We saw that these stereotypical restrictions are important if we are to describe the appropriate usage of superordinates. In practice they are not used to generalize over all things that fall under their denotation proper; rather, this denotation is restricted by the stereotypicality of the subtypes. Those specific concepts that have been called "food" in dictionaries can then be seen as representing the stereotypical subtypes which tend to be called "food" in general. Those things will be called "food" that not only are "edible" but also resemble typical foodstuff in other aspects. *Plants, fruits* and *animals* may be (partially) edible but are, unless the context explicitly requires it, not typically called such because they are not *processed* or *cooked substances*. We could say that the classifications in dictionaries give more information on the classifying superordinate than on the classified subtypes. The single classification of a subtype in dictionaries only reflects one way of conceptualizing it, whereas in many cases we can vary our perspectives relative to the context. The superordinates, on the other hand, express only one such conceptualization as part of their lexicalized pragmatic function in language.

4 Conclusion

I have described a model of meaning in which a distinction is made between the denotational level and the lexical semantic level of meaning. Denotation is then defined in terms of cognitive basic concepts rather than abstract meaning components and the denotational dependency of superordinates has been reversed. In such a model, we can give a better account of the pragmatic, linguistic usage of taxonomically related nouns (words that are neutral names and words that carry a predicative loading) and we can explain stereotypical substitution restrictions of classifying nouns. Finally, it has the advantage that world-knowledge can be defined at the denotation level, outside the linguistic lexicon, whereas dependency on these concepts (from which further semantic properties can be derived) can be stated in various ways.

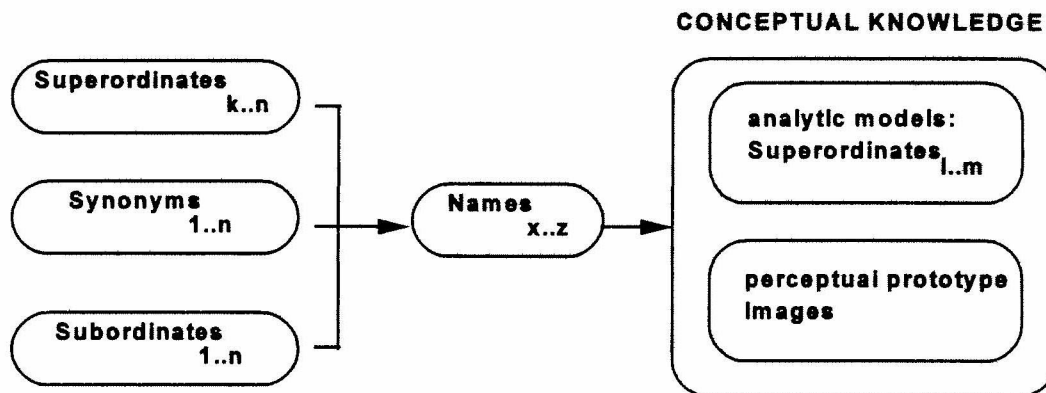
Furthermore, we saw that lexicalization patterns and definition structures in dictionaries suggest that the vocabulary of a language contains many conventional and unconventional nouns that can be used to classify various things in the world. The usage of these nouns can better be defined by reference to the more specific things that they can name, as is also often done in dictionary definitions. Finally, I have argued that variation in classification in dictionaries can be seen as an indication that cognitive classification of basic level concepts should not be construed according to a classical model but should give access to rich world-knowledge which can be organized in different, simultaneously applying epistemic models, reflecting all kinds of classification schemes. The stereotypically restricted classification in dictionaries only presents a single category which tells us more about applicability of the classifying noun than about properties of the more specific concepts it classifies.

From a knowledge engineering point of view, we would still want to distinguish various general classes of denotations with which to group related properties (Weigand 1989, Pustejovsky and Bergler 1991, Vossen 1991, Calzolari 1991, Vossen and Copestake 1994). However, in that case assignment to classes would have a purely technical role in describing world-knowledge, no longer capturing the notion of presupposition as used before. An example of such classes is given by the complementary denotational types *object* and *substance*, which represent technical labels for things with and without SHAPE. These labels are not the same as the predicates "substance" and "object" since, as we saw, these do not strictly apply to all shaped and shapeless things, respectively.

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

From an epistemic point of view, we may still want to store the class of a basic-level concept in order to represent a particular conventional inferencing scheme. One of these schemes would indicate that *persons* fall into the class of objects just as *cars* and *houses* do. However, such inferencing schemes within our conceptual knowledge should not be mixed up with the way denotation has been defined here. Our notion of denotation links up with our general conception of reality, in which analytic inferencing schemes are just one aspect. Only a small subset of superordinates has such a special role in representing the conventional analytic classes for basic-level concepts, whereas all superordinates have a stereotypical linkage with such concepts that explains their typical pragmatic use:

Linguistic and cognitive dependencies



The superordinates which are used to define the encyclopaedic information of basic-level concepts ($\text{Superordinates}_{l..m}$) are a subset of all the generalizing terms in the vocabulary of languages ($\text{Superordinates}_{k..n}$). Many of the unconventional superordinates have diverse or void presuppositions so that the generalizations they express are all that can be said about them. At the cognitive level, the database still gives access to knowledge of the world which can take various forms (perceptual, propositional), can have various functions and on which different operations can apply to infer properties of these entities. Furthermore, common sense inferencing can also make use of the stereotypical linkage with more specific concepts. If no other information is known beyond an abstract classification such as "a vehicle", we may build up much richer expectations on the basis of the more specific subtypes associated with the class, such as "car" and "train", than by deducing properties from a more abstract superclass such as "something" or "object".

References

- Berlin B.
1972 "Speculations on the growth of etnobotanical nomenclature", in: *Language in Society*, 51-86.
- Brachman R.
1979 "On the epistemological status of semantic networks", in: N. Findler (ed.) *Associative networks: representation and use of knowledge by computers*. New York: Academic Press: 3-50.
- Briscoe E.J. - A. Copestake - A. Lascarides
1992 *Blocking*. Computer Laboratory, University of Cambridge and Human Communication Research Centre, University of Edinburgh.
- Briscoe E J. - A. Copestake - V. de Paiva (eds.)
1994 *Default inheritance in unification based approaches to the lexicon*. Cambridge: Cambridge University Press.
- Brown, R.

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

- 1958 "How shall things be called?", in: *Psychological Review* 65, 14-21
- Calzolari N.
1991 "Acquiring and representing semantic information in a lexical knowledge base", in: J. Pustejovsky and S. Bergler (eds.) *Lexical Semantics and Knowledge Representation: Proceedings of the first SIGLEX Workshop Berkeley*. New York/Berlin: Springer Verlag: 235-244 (also available as *Esprit BRA-3030 Working Paper* 016).
- Carpenter R.
1990 "Typed feature structures: inheritance, (in)equality and extensionality", in: *Proceedings of the Workshop on Inheritance in Natural Language Processing*. Tilburg: 9-18.
- Collins A.M. - M.R. Quillian
1969 "Retrieval time from semantic memory", in: *Journal of Verbal Learning and Verbal Behavior*, 8. New York: Academic Press: 240 - 248.
- Copestake, A.
1993 *The Compleat LKB*, Acquilex-II Deliverable, 3.1
- Cruse, D.A.
1986 *Lexical Semantics*. Cambridge: Cambridge University Press.
- Daelemans W. - K de Smedt - G. Gazdar
1992 "Inheritance in natural language processing", in: *Computational Linguistics*, 18, 2. Cambridge MA: MIT Press: 205-219.
- Fillmore C.
1971 "Types of lexical information", in: D.D. Steinberg and L.A. Jakobovits (eds.) *Semantics, an interdisciplinary reader in philosophy, linguistics, and psychology*, p. 370 - 392. London, Cambridge University Press.
- Fillmore, C.
1982 "Towards a Descriptive Framework for Spatial Deixis", in: Jarvella and Klein (eds.) *Speech, Place and Action*, London: John Wiley.
- Grice H.P.
1975 "Logic and conversation", in: P. Cole & J.L. Morgan (eds.) *Syntax and semantics*, 3. New York/London: Academic Press: 41-58.
- Haiman J.
1980 "Dictionaries and Encyclopedias", in: *Lingua* 50, 1980: 329-357.
- Lakoff G.
1986 "Classifiers as a reflection of mind", in: C. Craig (ed.) *Noun classes and categorization*. Amsterdam: Benjamin: 13-51.
- Lascarides, A. - T. Briscoe - N. Asher - A. Copestake
1994 "Order independent and persistent typed default unification", *Esprit BRA-7315 Acquilex Working Paper* 33.
- Meijs W.J. - P. Vossen
1991 In so many words: Knowledge as a lexical phenomenon, in: *Proceedings of the ACL Siglex Workshop on Lexical Semantics and Knowledge Representation*, Berkely, California: 113 - 126, also available in: J. Pustejovsky and S. Bergler (eds.) *Lexical Semantics and Knowledge Representation*. Berlin: Springer-Verlag, 1992.
- Miller G.A.
1969 "A psychological method to investigate verbal concepts", in: *Journal of Mathematical Psychology*, 6. New York: Academic Press: 169-191.
- Miller et al.
1990 "Introduction to WordNet: An On-line Lexical Database, in: *International Journal of Lexicography*, Vol 3, No.4 (winter 1990), 235-244.
- Pollard C. - I. Sag
1987 *An information-based approach to syntax and semantics: Volume 1 fundamentals*. CSLI Lecture Notes 13. Stanford CA.
- Procter, P.
1978 *The Longman Dictionary of Contemporary English*, Longman, London.
- Pustejovsky J.
1991 "The generative lexicon", in: *Computational Linguistics*, 17, 4. Cambridge MA: MIT Press: 409-442.
- Pustejovsky J. - S. Bergler (eds.)
1992 *Lexical semantics and knowledge representation: Proceedings of the First SIGLEX Workshop Berkeley*. USA, June 1991. Lecture Notes in Artificial Intelligence nr. 627. New York/Berlin: Springer Verlag.
- Putnam H.
1975 "The Meaning of Meaning", in: H. Putnam (ed.) *Philosophical Papers Vol. II*, Cambridge: 215-271.
- Quine W.V.
1960 *Word and Object*. Cambridge MA: MIT Press.
- Rosch E.

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

- 1977 "Classification of real world objects: origins and representation in cognition", in: P.N. Johnson-Laird and P.C. Wason (eds.) *Thinking: Readings in cognitive science*. Cambridge: Cambridge University Press: 212-222.
- Sterkenburg P. van - W.J.J. Pijnenburg
1984 *Groot woordenboek van hedendaags Nederlands*. Utrecht: Van Dale Lexicografie.
- Vossen P.
1991 "Comparing noun-taxonomies cross-linguistically", *Esprit BRA-3030 Aquilex WP* No. 014, University of Amsterdam.
- Vossen P. - A. Copestake
1994 "Untangling definition structure into knowledge representation", in: Briscoe, E J. A. Copestake and V. de Paiva (eds.) *Inheritance, defaults and the lexicon*, Cambridge: Cambridge University Press, UK.
- Vossen P.
1995 Grammatical and Conceptual Individuation in the Lexicon, PhD. Thesis, University of Amsterdam, IFOTT.
- Weigand H.
1989 *Linguistically motivated principles of knowledge base systems*. PhD. Thesis, the Free University, Dordrecht: Foris.
- Wierzbicka A.
1980 *Lingua Mentalis*. London: Academic Press.

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP