Prototype theory and lexical relations

D.A.Cruse

The aim of this article is to consider what contribution prototype theory can make to the study of lexical relations. Three areas of relevance suggest themselves: the characterisation of relations as natural concepts; the nature of the relata; and the predictability of relations from representations of the relata. Sense relations have usually been defined by means of logical, or pseudo-logical formulae. These have certain drawbacks: excessive rigidity, lack of congruence with native-speaker intuitions, and a failure to account for goodness-of-exemplar judgements. A standard prototype representation in terms of a set of weighted, non-necessary features, all of which contribute to membership and centrality, offers some advantages. But the most satisfying picture for all the main relations seems to require a 'core' feature which controls membership, and a set of 'cosmetic' features which affect only 'goodness' within the category. The core feature is perhaps best seen as a Lakoffian image-schema, rather than a logical definition. The standard prototype model may well be valid for certain types of category: it need not be assumed that all categories are built to the same model. However, problems in predicting relations from representations suggest that the standard model may require modification. The relata of lexical relations are most usefully viewed as conceptual entities that receive distinct (prototype or other) representations.

1. Introduction.

1.1. Prototype theory.

Prototype theory is now a familiar part of the mental landscape of most linguists; no attempt will therefore be made to expound it in any detail here. Its origins can be traced at least as far back as Wittgenstein's notion of 'family resemblance', which was the first major challenge to the Aristotelian conception of categories as being characterisable in terms of necessary and sufficient criteria. Rosch and others took this idea forward, showing that natural conceptual categories have a number of properties which are inexplicable on the classical view; these are the so-called 'prototype effects': fuzzy boundaries; graded goodness-of-exemplar scores; early and frequent mention by subjects asked to list

Rivista di Linguistica 6.2 (1994) pp 167-188 *(ricevuto nell'ottobre 1994)*

members of categories; early acquisition; faster verification of category-member pairs; stronger priming effect of category name on member, etc. Such prototype effects are very robust, and are accepted now by most scholars. The question of how such concepts are represented in the cognitive system (or, alternatively, in the mental lexicon) is, however, more controversial.

1.2 Prototype theory and lexical semantics.

When prototype theory is incorporated into lexical semantics, representation usually takes the form of a modified feature specification (see, for example, Coleman & Kay 1981); some psychologists also adopt this format (Hampton 1992 provides a recent example). (The meaning of a word is usually identified with the concept that it gives access to; for the most part, sense relations will be treated here as conceptual relations.) Unlike the semantic features proposed by Katz & Fodor (1964), prototype features are not necessary and sufficient. Although the details of prototype representations vary from scholar to scholar, those who adopt a feature-based approach agree that the more of the prototype features of a category an item possesses, the better an example of the category it is. In what follows, I shall at least to begin with assume some kind of feature representation.

1.3 Lexical relations.

The main concern here will be with abstract paradigmatic relations such as hyponymy, incompatibility, meronymy, oppositeness, and so on, which formed the bulk of the subject-matter of Cruse (1986). In that work a prototype approach was not adopted, although the notion was well-established at the time of writing. It seems therefore worthwhile to investigate what contribution prototype theory can make to the study of sense relations, in comparison to what has hitherto been the more standard logical approaches, and conversely, to see whether the study of sense relations has any repercussions on prototype theory.

2. Standard definitions of lexical relations.

There are two standard approaches to the definition of lexical sense relations in what might loosely be described as a logical manner, namely, the extensional approach and the intensional approach (see Lyons 1968). I shall first of all illustrate extensional definitions of a range of relations.

2.1 Extensional definitions.

2.1.1. Hyponymy.

A typical extensional definition of hyponymy runs as follows (from Cann 1993, but slightly modified so as to exclude synonymy):

(1) X is a hyponym of Y iff there exists a meaning postulate relating X' and Y' of the form: $\forall x[X'(x) \rightarrow Y'(x)]$, but none of the form: $\forall x\,[Y'(x) \rightarrow X'(x)]$.

X' and Y' in the above definition represent the logical constants corresponding to the lexical items X and Y. The definition thus states that the extension of X is included in the extension of Y, but not vice versa.

2.1.2. Synonymy.

Synonymy is defined as bilateral hyponymy:

(2) X is a synonym of Y iff there exists a meaning postulate relating X' and Y' of the form: $\forall x\,[X'(x) \leftrightarrow Y'(x)]$.

This definition stipulates that the extensions of two synonyms must be identical. This is a very strict criterion.

2.1.3. Incompatibility.

The definition of incompatibility specifies an implication from one term to the negation of the other:

(3) X and Y are incompatibles iff there exists a meaning postulate relating X' and Y' of the form: $\forall x[X'(x) \rightarrow \sim Y'(x)]$.

Notice that it is not necessary to specify the reverse relationship:

(4) $\forall x[Y'(x) \rightarrow \sim X'(x)]$

because the definition entails that x cannot belong simultaneously to X' and Y', so the reverse relation follows automatically.

2.1.4. *Complementarity.*

Complementarity is a sub-variety of incompatibility: its definition requires extensional identity between one term of the pair and the negation of the other term:

(5) X is a complementary of Y iff there is a meaning postulate relating X' and Y' of the form: $\forall x[X'(x) \longleftrightarrow \sim Y'(x)]$.

This is a version which is equivalent to Cann's definition, but strictly x should be required to belong to some domain Z. For instance, *dead* and *alive* are complementaries only when applied to living things: we cannot say *The pencil is not alive, therefore it is dead.* A better definition might therefore be:

(6) X is a complementary of Y in domain Z iff there exists a meaning postulate relating X', Y' and Z' of the form:
$\forall x[Z'(x) \rightarrow [X'(x) \longleftrightarrow \sim Y'(x)]]$

2.1.5 *Converses.*

Although the full range of converse types discussed in Cruse (1986) is difficult to capture in a logical formula, Cann (1993) offers the following definition which covers the most obvious cases:

(7) X is a converse of Y iff there is a meaning postulate relating X' and Y' of the form:
$\forall x ... \forall y[X'(y)...(x) \rightarrow Y'(x)...(y)]$

This is designed to cater for predicates with different valencies; in contrast to most of the logical definitions given above this one is relatively relaxed in the sense that it allows pairs like *daughter* and *parent* to count as converses, even though *X is Y's parent* does not imply *Y is X's daughter.* A stricter definition would require bilateral implication.

2.1.6. *Meronymy and oppositeness.*

Two important meaning relations cannot be captured by definitions of the kind we have been considering. One of these is meronymy, the relation between *hand* and *finger.* The other is oppositeness (in the everyday sense in which *black* and *white* but not *red* and *yellow* are opposites). (See, however, the definition of antonymy in Lehrer & Lehrer 1982).

2.1.7. *Problems with the extensional approach.*

There are two main problems with standard extensional definitions of sense relations. Both stem from the fact that basically they do not deal with meanings. One problem (see Lyons 1968) is that some terms, e.g. *unicorn*, have no extensions, and it is therefore somehow unsatisfactory to attempt to work out their sense relations in this way; however, although they have no extensions, this does not mean that they have no meaning - an adequate account of the meaning of *unicorn* would show that it is the name of an animal, albeit a non-existent one. The second drawback concerns the nature of the material implication connective. *P implies Q* makes no reference to the meanings of P and Q, but merely to the way their truth conditions are related: *P implies Q* is therefore true provided that Q is never false when P is true. On this definition, *The door is open* implies *All bachelors are unmarried.* What is needed is a notion of semantic entailment (see, for example, Seuren 1985:212) whereby the truth of one proposition follows inescapably from the truth of another by virtue of the internal semantic make-up of the propositions. This is the essence of an intensional approach (as intended by Lyons).

2.2. *Intensional approaches.*

2.2.1. *Hyponymy.*

An example of an intensional definition of hyponymy is given in (8) (after Cruse 1986):

(8) X is a hyponym of Y iff $F^a(X)$ entails but is not entailed by $F^a(Y)$.

In (8), F^a represents a sentential function satisfied by X and Y; entailment is expressed here as a relation between the sentences $F^a(X)$ and $F^a(Y)$. Strictly speaking, entailment is a relation between propositions, not sentences, so (8) is a kind of shorthand for something like:

(9) X is a hyponym of Y iff for every situation in which $F^a(X)$ expresses a true proposition, $F^a(Y)$ also expresses a true proposition (assuming referential constancy) whose truth follows from the truth of the proposition expressed by $F^a(X)$, but not vice versa.

It should be pointed out that in the above definitions restrictions must be placed on the nature of the sentential function (F^a stands for a sentential function of the appropriate type). This is because the relation of entailment does not hold in all sentential functions: for example, *The material is scarlet* entails *The material is red*, but *The material turned*

scarlet does not entail *The material turned red,* because it might have been another shade of red to start with. Although it is intuitively clear which sentential functions are appropriate, it is not easy to specify them in general terms, and no attempt will be made to do this here (notice that this is not a problem for the extensional definition).

2.2.2. *Synonymy.*

Synonymy can again be defined as bilateral hyponymy:

(10) X and Y are synonyms iff F(X) entails and is entailed by F(Y).

This, too, must be interpreted as a shorthand version of a longer, more precise definition, as with hyponymy. However, interestingly, there are no restrictions on the nature of F (hence F in the definition rather than F^a)

2.2.3. *Incompatibility.*

The definition of incompatibility, like that of hyponymy, involves a restricted set of 'appropriate' sentential functions, but since they are not the same as those required for hyponymy, they will be symbolised as F^b:

(11) X and Y are incompatibles iff the truth of $F^b(X)$ entails the falsity of $F^b(Y)$.

(Notice that *The material turned red* entails *The material did not turn blue*, so sentences of this form are appropriate for incompatibility.)

2.2.4. *Complementarity.*

The definition of complementarity shares with that of synonymy the characteristic that no restrictions are necessary on the nature of F:

(12) X and Y are complementaries iff the truth of F(X) entails and is entailed by the falsity of F(Y).

Hence, since *A is dead* entails and is entailed by *A is not alive*, *dead* and *alive* are complementaries. Notice, however, that the restriction to a particular domain, in this case that of the organic world, which was introduced into the extensional definition, is still necessary here.

2.2.5. *Converseness.*

Converses are not easy to characterise in this way; the following definition will serve for present purposes:

(13) X and Y are converses iff $F^X(A,B)$ entails and is entailed by $F^Y(B,A)$, where F^X and F^Y are two-place sentential functions containing, respectively, X and Y in parallel syntactic positions.

Hence, if F^X = *..is above..*, and F^Y = *..is below..*, then *A is below B* entails and is entailed by *B is above A*, showing that *above* and *below* are converses. As with synonymy and complementarity (all involving bilateral entailment) there are no restrictions on the nature of F other than that of yielding grammatical sentences with appropriate lexical arguments.

2.2.6. *Meronymy.*

Meronymy presents problems of definition similar to, but more acute than, those of hyponymy. The simplest definition in terms of entailment in sentential functions is as follows:

(14) X is a meronym of Y iff $F^c(X)$ entails $F^c(Y)$ and $F^d(Y)$ entails $F^d(X)$.

F^c denotes a sentential function of an appropriate type, such as *The boil is on John's -: The boil is on John's elbow* entails *The boil is on John's arm*, so *elbow* is a meronym of *arm*. F^d denotes a sentential function which gives the reverse entailment, such as *I examined the whole of John's -*, which gives, for instance, *I examined the whole of John's hand.* This entails *I examined the whole of John's palm.* Notice that the set of appropriate functions for meronymy overlaps with that of appropriate functions for hyponymy: *The boil is on one of John's elbows* entails both *The boils is on one of John's arms* and *The boil is on one of John's joints*. This is not necessarily a problem since the two relations will be distinguished by the distribution of entailments within the domain of sentential functions. More serious problems are: firstly, the difficulty of characterising appropriate sentential functions in general terms rather than simply listing them, which would hardly be satisfactory (on the face of it, this would seem more difficult than the corresponding task for hyponymy); and secondly, the fact that in many cases it seems that a modification in the sentential function is required, in addition to intersubstitution of the lexical items in question. For instance, it does not seem quite right to say that *I examined the whole of John's hand* entails *I examined the whole of John's finger.* It might be better to say that *I examined John's fingers* was a better entailment, but

that changes the sentential function, and admitting such cases into the definition requires us not only to specify in general terms the appropriate functions but also appropriate modifications.

2.2.7. *Oppositeness.*

The intuitively strong relationship of oppositeness does not seem to be amenable to a sentential-function-plus-entailment definition at all.

2.3. *Problems with logical approaches in general.*

There are two main shortcomings with logical definitions of sense relations, and these would not disappear even if the definitions could be made 'watertight'. The first is that they are too strict, in that they do not correspond to native speaker intuitions. For instance, the notion of synonymy is a familiar one in lexicography. Not only are there dictionaries of synonyms, but synonyms are often used and signalled in the definitions which appear in standard dictionaries. People generally speaking seem to know what to expect from a list of synonyms. However, an inspection of any such sets of synonyms reveals that they go well beyond what would be sanctioned by either of the logical definitions offered above. Likewise, students asked (in an introductory session on lexical relations) to pick out pairs which show the same relationship as *apple:fruit* will quite happily pick out *dog:pet*, even though *It's a dog* does not entail *It's a pet*; and speakers will use this pair in expressions which seem to demand a relation of this sort: *dogs and other pets* is just as normal as *apples and other fruit*. Furthermore, hyponymy, if defined logically, ought to be a transitive relation; however, intuitive judgement does not confirm this. For instance, speakers typically will assent to (15a) and (15b), but not to (15c):

(15) a. A hang-glider is a kind of glider
b. A glider is a kind of aeroplane
c. ?A hang-glider is a kind of aeroplane

Similarly, to appear normally in coordinated lists, it is not necessary for items to be strict logical incompatibles: *You meet all sorts here --- students, nurses, prostitutes.* In other words, strict logical relations seem not to have any particular intuitive status.

The second shortcoming of logical definitions is that they can offer no account of the so-called 'prototype effects', in particular, fuzzy boundaries and graded 'goodness-of-exemplar' ratings.

For instance, *stallion:horse, spaniel:animal and horse:animal* all pass the logical tests for hyponymy, but speakers typically judge the first two to be less good examples of the relationship than the last.

3. *Prototype definitions: relations as natural categories.*

A possible strategy for achieving more satisfactory definitions of lexical relations is to treat them as natural concepts of a relational kind and attempt to apply the prototype model of category structure to them. This will be more convincing if they can be shown to be natural concepts in everyday use, and even more so if they can be shown to be expressed by common linguistic forms.

3.1. *Definition by list of prototypic features.*

We shall first of all attempt to apply the feature model of prototype structure. This involves producing a set of features not unlike the Katz-Fodor type, with the important difference that they are not to be taken as necessary and sufficient, as in the classical approach to category specification, but have the property that the more of them some item manifests or possesses, the more central that item is in the relevant category. The highest scoring items in a category will be the prototypical examples.

3.1.1. *Hyponymy.*

I shall claim that the lexical relation of hyponymy corresponds to an everyday relational concept, which will be referred to as the [KIND OF] concept, and which is expressed by, for instance *An X is a kind/type of Y*. There is evidence that this concept has a prototype structure; indications of particular importance are (a) the fact that there are good and less good examples (of pairs of items manifesting the relation), and (b) the fact that the concept has no clear boundaries. Consider the following as prototype features of [KIND OF] (they can be tested by how well they predict the appropriateness and normality of, for instance, *X is a kind/type of Y*. I leave it open as to whether or not there is a finite list of such features):

(16) (i) F'(X) entails F'(Y)
(ii) X and Y have the same 'perspective'
(iii) X and Y occupy adjacent nodes within a coherent taxonomy
(iv) X is one of a set of incompatibles such that they represent the 'best' sub-division of Y

I use 'perspective' to refer to the fact that for instance, *stallion* is in some sense oriented towards the animal's sex - maleness is highlighted, and equinity is backgrounded or presupposed; this shows up in the fact that *That's not a stallion* is likely to be interpreted as referring to a

mare. *Horse,* on the other hand, has no such special orientation, hence the perspectives are different. The effect of this is that *stallion:horse* are less good as an example of [TYPE OF] than *horse:animal.* Other examples of differences of perspective are *blonde:woman*, and *waiter:man.* (Compare the examples in (17), where only (17a) shows constancy of perspective:

(17) a. ?A waiter is a type of man
b. ?A blonde is a type of woman
c. An ash-blonde is a type of blonde

Feature (iii) is meant to capture the fact that, for example, *alsatian:dog* exemplify the relation better than *alsatian:animal* or *alsatian:creature,* where, intuitively, one or more taxonomic levels have been skipped. The explanation of why, for instance *dog:animal,* do not give the impression of a level-skip, in spite of the existence of the intermediate term *mammal*, presumably lies in the fact that *mammal* does not belong to the same register as the other two terms: it is not part of the folk taxonomy, but is a technical term. Feature (iv) should be interpreted in the psychological sense: a good category is one whose members resemble one another to a high degree, and show a high degree of difference from members of other categories. It follows from this criterion that by and large the best examples of [TYPE OF] involve basic level concepts as hyponyms and 'kind' concepts as superordinates, hence *apple:fruit* exemplify hyponymy better than *ash-blonde:blonde.*

The features suggested above account reasonably well for some of the graded goodness-of-exemplar judgements for hyponymous pairs. In particular, they account for varying centrality among pairs that are indubitably examples of hyponymy.

But as they stand, they are not entirely satisfactory, because they do not account for border-line phenomena: they do not show, for instance, why *dog:pet* are better hyponyms than, say, *hammer:weapon.* The problem is that if the feature of entailment is absent, the remaining features do not guarantee the right sort of relation. We shall return to this point.

3.1.2. Hyponymy: a conceptual or a lexical relation?

Perhaps we should distinguish between the conceptual relation [TYPE OF] and the lexical relation of hyponymy. If we do this, then there are additional prototype features for hyponymy:

(18) (v) X and Y belong to the same syntactic category
(vi) X and Y have the same 'word-specific' properties

Feature (v) would predict that *hammer:tool* are better than *knife:cutlery*, because the former pair are both count nouns, whereas *cutlery* is a mass noun. Feature (vi) is based on the assumption that, for instance, *father* and *daddy* both evoke the concept [FATHER], but *daddy* has additional expressive features, specific to the word rather than properties of the concept, which 'modulate' the concept. Hence, because their modulatory properties are not the same, *doggy:animal,* for instance, exemplify the lexical relation of hyponymy less well than, say, *dog:animal.*

3.1.3. Incompatibility.

The relation of incompatibility can be treated in a way parallel to that adopted for hyponymy, that is to say, as a natural category with a prototypic structure. In this case the concept will be referred to as [DIFFERENT TYPE OF]; this is a three-place relation:

(19) [(X)- A DIFFERENT TYPE OF-(Z)-FROM- (Y)].

It can be expressed in everyday language by means of sentences such as: *X is a kind of Z; Y is a different/another kind.* Obviously this will inherit some of its properties from [TYPE OF]. For instance, other things being equal, the GOE rating will be higher if X, Y and Z have the same perspective. The distinctive prototypical properties of this relation will include:

(20) (i) The truth of F"(X) entails the falsity of F"(Y)
(ii) X and Y are at the same taxonomic level

As with hyponymy, and for similar reasons, it may be profitable to distinguish a conceptual and a lexical relation.

3.1.4. Meronymy.

Meronymy and oppositeness fare rather better under this approach.

The conceptual relation corresponding to meronymy is [PART OF]. The main prototypical features are as follows:

(21) (i) Integrality
(ii) Congruence
a. range
b. type
c. stage
(iii) Motivated boundaries
(iv) Sharp boundaries

(v) X has a determinate function relative to Y
(vi) Necessity
(vii) Immediacy

Feature (i) is the feature which *handle:spoon* manifest but *handle:door* do not:

(22) a. The handle is attached to the door
b. ?The handle is attached to the spoon

The features under (ii) refer to whether two entities are 'matched' or not. Feature (iia) refers to matching in respect of specificity, or range; compare *finger:hand* and *lid:pan*: all hands have fingers, and fingers only occur on hands; however, not all pans have lids, and lids do not occur only on pans.

Feature (iib) refers to matching in respect of type; for instance, *finger:hand* are of the same type, but *vein:hand* are not (this distinction is described in Cruse 1986 as one between 'segmental' and 'systemic' parts). Feature (iii) is one of those which distinguishes between pieces and parts, the latter having arbitrary boundaries. Feature (iv) would predict that *tip:tongue* would have a lower GOE for [PART OF] than, say, *nail:finger*.

Feature (v) is the second main feature, along with feature (iii), which distinguishes pieces from parts. Feature (vi) refers to the distinction between 'canonical' parts and 'facultative' parts (the possession of a canonical part is not a logical necessity): a well-formed hand must have fingers, therefore *finger* is a canonical part of *hand*, but a door does not have to have a handle to qualify as well-formed. Feature (vii) is exactly parallel to feature (iii) for hyponymy. Once again, a purely conceptual relation can be distinguished from a lexical relation.

3.1.5. Opposites as a general category.

The general category of opposites can be treated as a prototypic category, but the individual sub-types - complementaries, antonyms, reversives, etc. - present a methodological problem. The overall category is one about which untrained native speakers have clear intuitions; anyone can answer questions like:

(23) What's the opposite of *long/hot/good/undress/up/give*?

The factors which affect the 'goodness' of a pair of words as examples of opposites would appear to include the following:

(24) (i) diametric opposition
(ii) binarity
(iii) exhaustiveness of superordinate domain
(iv) symmetry

Thus, for instance, *tall* and *squat* are not prototypical opposites because they are not diametrically opposed (one can become more squat without becoming less tall), nor are they symmetrical; *huge* and *small*, and *prolong* and *shorten*, are diametrically opposed, but are not symmetrical; *huge* and *tiny* are symmetrical, but do not exhaust the superordinate domain to the same extent that *large* and *small* do, etc. We can again distinguish the conceptual relation from the lexical relation: *clean* and *mucky* are good conceptual opposites, but (assuming that *dirty* and *mucky* access the same concept) less good lexical opposites, because they do not belong to the same register. Once the general properties of opposites are established, the various sub-varieties seem to be characterisable by one or two features. Thus, antonyms are gradable adjectives, reversives are change-of-state verbs, and for converses, a logical definition will capture the essence.

3.2. Problematic aspects of the prototype approach.

Although the prototype approach to sense relations has some advantages over the logical definition approach, in the form in which it has been presented above, it is still not fully satisfactory. By having a set of features from which a selection must be made, and which must be either present or absent, an important point about the organisation of these relational concepts is obscured, and that is that in most cases (and perhaps all), there is a single controlling feature which determines membership, and the remaining features are what will be called 'cosmetic'. Furthermore, that feature has to be seen as not simply present or absent, but in some sense gradable, with a threshold value corresponding to membership, and values below the threshold controlling degree of membership. It is true that in the prototype model the features can be weighted differentially, but that still allows them all to contribute to membership, and it is arguable that in the case of sense relational concepts we are looking at something slightly different.

4. The 'core definition plus cosmetic features' model.

A possible alternative to the standard prototype model is a 'core definition plus cosmetic features' model, in which category membership is governed by the core definition, and the cosmetic features control only centrality. A model of this sort was put forward by Osherson & Smith (1982); their motivation, however, rather different from ours, was

the existence of the so-called 'guppy effect'. A similar proposal was made in Persson (1990). For hyponymy this would involve elevating feature (i) in definition (16) to the status of core definition, and giving the remaining features the status of cosmetic features. A similar move would be possible for the other relations defined (in each case, the first feature in the list would be the core definition). However, what would not be captured in this way would be the necessary gradability of the core definition: we need to provide for degrees of relaxation of the core definition. This problem will be discussed in connection with synonymy.

4.1. Relaxing the core definition: a detailed study of synonymy.

It is puzzlingly difficult to draw up a prototype definition for synonymy; however, the notion of a graded core definition holds out some promise. As we have seen, strict versions of the core definition are fairly straightforward (repeated here for convenience):

(25). a. X is a synonym of Y iff there exists a meaning postulate relating X' and Y' of the form: $\forall x[X'(x) \leftrightarrow Y'(x)]$.

b. X and Y are synonyms iff F(X) entails and is entailed by F(Y).

Both of these definitions ignore non-propositional meaning and therefore are unable to account for the fact that *pass away* and *decease* are more synonymyous than either *pass away:kick the bucket* or *decease:kick the bucket*. They do, however, capture the important point that propositional meaning is of greater significance in determining synonymy than non-propositional meaning. It has already been argued that these definitions are too strict; we must therefore consider possible ways of relaxing them and making them gradable.

4.1.1. Relaxing the extensional definition.

Let us look first at the extensional definition. Cann (1993) suggests that this can be relaxed by replacing the universal quantifier in the formula by "most", which defines synonymy by the meaning postulate "Most X are Y and most Y are X". Obviously this can be interpreted as gradable according to the interpretation given to "most": the higher the qualifying percentage, the greater the degree of synonymy guaranteed by the definition. Another way of relaxing the extensional definition is to stipulate not that everything that is X must be Y (and vice versa), but that for X and Y to count as synonyms, a prototypical X must be Y and a prototypical Y, X. It may not be immediately obvious how this can be made gradable, but it can be. In many versions of prototype theory

there are ways of quantifying degrees of prototypicality (often expressed as 'degrees of membership'). Suppose we assign a percentage score of 100% to that item in a category with the highest goodnes-of-exemplar rating, and proportionately lower scores to items with lower ratings. The degree of synonymity between X and Y could then be expressed as being inversely related to the lowest qualifying percentage score for an X to be guaranteed to be a Y and vice versa. That is, the lowest degree of synonymy will hold if only the prototype of X can be sure of being a Y, whereas a higher degree of synonymy will hold if, for instance, an 80% member of X can be sure of being a Y. (The highest degree of synonymy is when the mere fact that an item reaches the criterion for X guarantees that it also reaches the criterion for Y.) These two measures of degree of synonymity are not equivalent: it is not logically necessary for the prototypical subcategory of a category to be the most frequent.

4.1.2. Relaxing the intensional definition.

There are also various non-equivalent ways of relaxing the intensional definitions. One way is to weaken the entailment relation so as to correspond to "strength of expectation", ranging from certainty downwards. Degree of synonymity would then be measured by some combined function of the degree of expectation from X to Y and that from Y to X. The thesaurus incorporated into the Microsoft Word word-processing programme gives *strange*, *unusual*, *odd*, *rare*, *peculiar* as synonyms. It will be noticed that these fall into two conceptually distinct groups, those referring to the objective fact of low frequency, and those referring to a subjective feeling of unfamiliarity. What unites these two groups seems to be the fact that something which does not happen very often is likely to give a feeling of strangeness, and vice versa - in other words they are united by a certain strength of expectation. Another way is to think in terms of the number or proportion of entailments of Y which are also entailments of X, and vice versa. This in effect is a measure of the degree of overlap of content between X and Y, which, at least intuitively, is a significant component of synonymy. Again these are at least to some extent independent measures; they are also (to some extent) independent of the extensional measures discussed earlier. Another point that should be mentioned is that although propositional meaning is more significant for synonymy than non-propositional meaning, it is less significant the less salient it is. So, for instance, *mare* and *stallion* are far less synonymous than *pretty* and *handsome*, although both might be said to differ in respect of the features [MALE] and [FEMALE], the reason being that in the latter pair these features are backgrounded (to the extent of being presupposed).

4.2. *An 'Ideal example plus resemblance' approach.*

We have examined a number of explicit ways of relaxing and grading strict definitions of synonymy. How are we to decide which way is the appropriate one? Arguably all of them are relevant. But we cannot be certain of having identified all the alternatives, or even that there is a finite list, and even if this were not a problem, we would need to decide on a way of combining the various methods of relaxation. This way of looking at synonymy seems doomed. Perhaps the best strategy is simply to leave the core definition intact (in either form?), as an ideal and think in terms of approximations to it, but without trying to specify in advance all the possible ways in which an approximation may be achieved, which may well be impossible in principle. I suggest a two-part characterisation of synonymy. First, a specification of an ideal, which functions as the end-point of a scale of synonymity (perhaps never realised):

(26) X and Y are absolutely synonymous if their meanings are identical.

(There are various ways of giving substance to the notion of 'identical meaning'. I offer Haas's notion of "equinormal in all grammatical contexts"; for further discussion, see Cruse 1986:268). Words will then be synonymous to the extent that they approximate to the ideal. However, we still need to characterise the everyday notion of synonymy; in other words, we need to set a boundary of some sort on the scale:

(27) X and Y are synonyms if their differences are insignificant.

Here it has to be recognised that the significance of meaning differences will be contextually sensitive; perhaps we can invoke the notion of relevance, and say that synonyms are words whose differences are irrelevant. This will yield different answers in particular contexts; but we might hypothesise a kind of default context, evoked in the case of isolated words, which governs what we judge to be synonyms in a lexicographic setting.

4.3. *Synonymy and other concepts.*

Pictured in this way, the notion of [SYNONYM] has some resemblances both to that of [ODD NUMBER] and that of [CIRCLE]. In all of these there is an ideal at the core. [SYNONYM] is more like [CIRCLE] than like [ODD NUMBER]. The latter has a strict core definition, (together with cosmetic features, which cause *5* to be judged a better example of the category than *44267*). However, there is no intelligible notion of "degree of oddness" for numbers. [CIRCLE] *is*

almost exactly like [SYNONYM], with an ideal (unattainable?) core, and membership governed by a contextually sensitive degree of resemblance threshold.

4.4. *Sense relations and schemas.*

What of other sense relations? Much of what has been argued for synonymy is valid for other relations. There are differences, however. The notion of 'cosmetic' features seems not to be so relevant for synonymy; the reason is perhaps that such features are subsumed under the heading of resemblance or approximation. Whatever the reason, we seem not to need both a notion of variable resemblance to an ideal, and additional cosmetic features. A consideration of other relations than synonymy suggests a further theoretical move. It was relatively easy to give a logical-type definition of synonymy. It is also possible for hyponymy, but here we have the problem of specifying relevant sentential functions. In the case of meronymy and, *a fortiori*, oppositeness, logical definitions do not seem to be the most appropriate. Can we find a more general approach? One possible avenue is in terms of schemas of the Lakoffian sort. These have the advantage of intuitive plausibility, but the disadvantage of being (at least at present) imprecise, impossible to tie down, or to base quantifications upon (cf. Brugman 1990). For instance, hyponymy and meronymy could be based on a containment schema, applied to classes in the former case, and individuals in the latter case. Incompatibles would be based on an exclusion schema. Opposites would have as their basis a schema of diametric opposition (manifested in, for example, the set-up in a tug-of-war). Synonymy would then have the simplest schema of all, namely, sameness. Instead of resemblance to the schema as the grading principle, we would have ease of imposing the schema on an area of content, or alternatively, of construing a situation or state of affairs in terms of the schema.

4.5. *What do lexical relations relate?*

There is a little-discussed problem concerning the entities which stand in relations of the lexical sort.

For Lyons, these are lexemes. But this does not make a great deal of sense in the case of polysemous lexemes (consider *dog* and *bitch*, which as lexemes are simultaneously superordinate and hyponym, and complementaries). In Cruse (1986) the entities related by sense relations were taken to be senses, established by ambiguity tests. This rationalises the *dog:bitch* anomaly, but leaves other problems. Take the relation between, on the one hand, *book* and *novel*, and on the other hand, *book* and *paperback*.

The word *book* is not ambiguous in the normal sense, but its meaning has two distinct components (which I call 'facets'), namely [TOME] and [TEXT]. *Novel* is strictly a hyponym of the [TEXT] facet, and *paperback* a hyponym of the [TOME] facet. Notice that *paperback* and *novel* are not incompatibles, as co-hyponyms of a single superordinate usually are. Another difficult case is *knife*. Consider the sentences in (28):

(28) a Use your knife and fork to eat your meat
b The intruder threatened me with a knife
c There was a box containing knives of various sorts

The reading of *knife* which appears in (28a) is hyponymous to *cutlery*, but the reading which appears in (28b) is not. Yet *knife* is not in the normal sense ambiguous, in fact even less so than *book*. And the reading which appears in (28c) is superordinate to the other two. In some cases, the relata of relations seem to crucially involve prototypes. Consider the relation illustrated in (29):

(29) fish:swim lizard:crawl bird:fly

The relation is intuitively constant here, yet not all birds fly, not even all well-formed birds. A suggestion with some plausibility is that sense relations hold fundamentally between prototypes, or at least, between distinct representations, at least some of which will be in the form of prototypes. I think it is possible to argue that [TEXT] and [TOME] require separate prototypes; and the *knife* that is hyponymous to *cutlery* is by all criteria a basic level term and requires its own prototype (the superordinate *knife* is more like a 'kind' term).

5. *Predicting lexical relations from prototype representations of relata.*

In order to achieve a fully satisfactory account of lexical relations in prototype theoretical terms, one further step must be taken (at least); it may well be the most difficult one, and no general solution will be offered here. It ought, in principle, to be possible, given prototype specifications of the meanings of two words (at least where this form of representation is the most appropriate one), to compute what relationship holds between them.

The requirement of relation computability places potentially interesting constraints on possible representations: other things being equal, a type of representation which allows relations to be read off or computed is to be preferred to one which does not allow this. Ideally, one should be able to read off goodness-of-exemplar values for the relation, too, but it is at present far from clear how this might be done.

The observations offered here are no more than preliminary; only two relations will be discussed, hyponymy and incompatibility.

5.1. *Computation of relations from 'classical' features.*

5.1.1. *Hyponymy.*

If the meanings of lexical items are defined in terms of classical necessary and sufficient features, à la Katz and Fodor, then the diagnosis of hyponymy is straightforward. X is a hyponym of Y iff all the features which define Y are included in the set which define X.

5.1.2. *Incompatibility.*

Unlike hyponymy, the relation of incompatibility cannot be computed simply on the basis of two feature representations. It is necessary also to know (and this must be stipulated as part of the theory) which features belong to the same set of 'antonymous n-tuples', or mutually exclusive sets. Lexical items X and Y are incompatibles iff any of the features defining X belongs to the same set of antonymous n-tuples as any of the features defining Y. Sets such as ("male", "female"), and ("equine", "bovine", "feline", "canine", etc.) form antonymous n-tuples.

5.2. *Computation of relations from prototype representations.*

Let us assume that a prototype representation consists of a set of weighted features, and that membership of a category is determined by whether a feature count which takes account of the weightings reaches an appropriate threshold value.

5.2.1. *Incompatibility.*

It is clear that it will not be possible to diagnose the relation of incompatibility between two lexical items on the basis of prototype representations of the above sort. For this to be possible, we would need the prototype equivalent of antonymous n-tuples. That is to say, we would need to know, for each pair of features (F(1), F(2)), to what degree the possession of F(1) leads to the expectation of the absence of feature F(2). On the basis of such information it might be possible to devise an algorithm to compute a global degree of mutual exclusion between two lexical items; incompatibility could then be defined in terms of a threshold value of this measure. To the best of my

knowledge, no proposal within prototype theory has taken this requirement on board.

5.2.2. *Hyponymy.*

One might expect hyponymy to be simpler. However, this is so only for prototypical hyponyms. Hampton claims that assenting to *An X is a kind of Y* does not involve a commitment to the truth of *All Xs are Ys*, but only to the truth of *Prototypical Xs are Ys*. (This is why transitivity failures are possible.) As a first approximation, the whole set of prototype features of a category can be regarded as necessary and sufficient for being a prototypical member of that category. Hyponymy might therefore be interpreted as follows:

(30) X is a hyponym of Y iff any item which possesses all the prototype features of X is guaranteed to reach the membership threshold weighted feature count for Y

This would seem to work well enough provided that X is the prototype subcategory of Y. The problem is that a prototype representation of category Y is going to be heavily biased towards the features of its prototype subcategory X. But consider the case of an item Z, which is a non-prototypical hyponym of Y and an incompatible of X. Some feature or features of Z must be co-members of antonymous n-tuples with some feature or features of X. But since the features of Y must reflect at least to some extent those of X, it is likely that there will be incompatibility between Z and Y, too. In other words, there is on the face of it a problem in devising a computational algorithm which will be able to distinguish beween incompatibles of Y and non-prototypical hyponyms of Y. This brief excursus into relation computation suggests that there are many problems awaiting solution.

6. *Prototypes and sense relations: an assessment.*

We are perhaps now in a position to assess the relevance of the prototype model of natural concepts for the study of sense relations. Firstly, it seems reasonable to assume that the relata of sense relations are the cognitive entities which qualify for distinct representations (of whatever sort). Secondly, it is clear that the classical prototype model has something to contribute to an understanding of lexical relations which a purely logical approach misses, in particular, the fact that goodness-of-exemplar ratings are at least partly determined by a set of non-necessary features which must be regarded as properties of the concept. But a closer examination suggests that while it is useful to think of lexical relations as natural concepts with a prototype structure, these

concepts differ in some significant ways from more familiar concepts such as [VEHICLE], for which a set of features can be drawn up which contribute jointly to the determination of membership of the category. In the case of lexical relations, for each relation there is one 'master' feature which is the main determinant of membership, and which can in many, but not all, cases, be plausibly viewed as a logical definition. The resultant model is thus close to that suggested by Osherson & Smith (1982); there are, however, some differences. First, Osherson & Smith suggested that this model should be generalised to all concepts, arguing that the standard prototype model was unable to explain the 'guppy effect' (a guppy is neither a prototypical fish, nor a prototypical pet, but is a prototypical pet fish). However, Hampton (1992) shows that the guppy effect can be explained using a more-or-less standard prototype model. There has been a (natural) tendency to seek for a universal model of concept structure. What the present investigation suggests is that not all concepts are built to the same model: there is room both for a standard prototype model and a core-plus-cosmetic-features model. (However, problems which emerge from a preliminary study of the possibility of computing relations from standard prototype representations indicate that some modifications might be necessary.) Another difference concerns the nature of the core. The Osherson & Smith core is a logical definition which is either satisfied or not. It is suggested here, first, that not all cores can plausibly be seen as logical definitions and that perhaps for sense relations an image-schema in the sense of Lakoff (1990) might be more generally applicable; and second, that it is more appropriate to think of degree of goodness-of-fit of the core, rather than satisfaction versus non-satisfaction.

Address of the Author:

D.A.Cruse
Department of Linguistics
University of Manchester
Oxford Road
Manchester M13 9PL
United Kingdom

References

Brugman, C. (1990), "What is the invariance hypothesis?", *Cognitive Linguistics* 1,2:257-266.

Cann, R. (1993), *Formal Semantics*, Cambridge, Cambridge University Press.

Coleman, L. & Kay, P. (1981), "Prototype semantics: the English word *lie*", *Language* 57:26-44.

Cruse, D.A. (1986), *Lexical Semantics*, Cambridge, Cambridge University Press.

Cruse, D.A. (1990), "Prototype theory and lexical semantics", in Tsohatzidis (1990:382-402).

Fodor, J.A. & J.J.Katz, eds. (1964), *The Structure of Language: Readings in the Philosophy of Language*, Englewood Cliffs, N.J., Prentice Hall.

Hampton, J.A. (1992), "The combination of prototype concepts", in Schwanenflugel (1992:91-116).

Katz, J.J. & J.A.Fodor (1964), "The structure of a semantic theory", in Fodor & Katz (1964:479-518).

Lakoff, G. (1987), *Women, Fire and Dangerous Things*, Chicago, University of Chicago Press.

Lakoff, G. (1990), "The invariance hypothesis: is abstract reason based on image-schemas?", *Cognitive Linguistics* 1.1:39-74.

Lehrer, A. J. & K. Lehrer (1982), "Antonymy", *Linguistics and Philosophy* 5:483-501.

Lyons, J. (1968), *Introduction to Theoretical Linguistics*, Cambridge, Cambridge Universty Press.

Lyons, J. (1977), *Semantics*, Cambridge, Cambridge University Press.

Osherson, D.N. & E.E.Smith (1981), "On the adequacy of prototype theory as a theory of concepts", *Cognition* 9:35-58.

Osherson, D.N. & E.E.Smith (1982), "Gradedness and conceptual conjunction", *Cognition* 12:299-318.

Persson, G. (1990), *Meanings and Metaphors: A Study in Lexical Semantics*, Stockholm, Almqvist & Wiksell International.

Pulman, S.G. (1983), *Word Meaning and Belief*, London, Croom Helm.

Rosch, E.H. (1975), "Cognitive representations of semantic categories", *Journal of Experimental Psychology: General* 104:192-233.

Schwanenflugel, P.J, ed. (1992), *The Psychology of Word Meanings*, Hillsdale, New Jersey, Lawrence Erlbaum Associates.

Seuren, P.A.M. (1985), *Discourse Semantics*, Oxford, Basil Blackwell.

Taylor, J.R. (1989), *Linguistic Categorization: Prototypes in Linguistic Theory*, Oxford, Clarendon Press.

Tsohatzidis, S.L, ed. (1990), *Meanings and Prototypes: Studies in Linguistic Categorization*, London, Routledge.