ISSUES IN PRAGMATICS (PLIN 3001) 2006-07

LEXICAL PRAGMATICS

3. Concepts and categorisation

1. Introduction

Last week we looked at Fodor's arguments against the classical theory of concepts, construed as a theory of word meaning. According to the classical theory, the meanings of most words are complex concepts with the internal structure of a definition. Fodor had a battery of arguments against the classical approach. In some cases, he argued that no satisfactory definition is available; in others, he argued that even though a satisfactory definition is available, there is experimental evidence to show that it does not function as our mental representation of the meaning of the word. His conclusion was that most word meanings are simple, unanalysable (atomic) concepts with no internal structure, which cannot be further decomposed. He added that, since only complex concepts can be learned, most word meanings are innate.

Today I want to look at **prototype** theories of concepts, which seem to offer an alternative approach. Prototype theorists also reject the classical theory of concepts, but for different reasons from Fodor's. Recall that in lecture 2 I noted that the classical theory of concepts has been attacked from both sides: as a theory of word meaning, and as a theory of categorisation. While Fodor criticises it as a theory of word meaning; prototype theorists criticise it as a theory of categorisation. The classical account of categorisation says that we recognise someone as a bachelor by checking that he fits the definition. Prototype theorists provide a wealth of experimental evidence against this account, and conclude instead that concepts have an internal structure composed of **typical features** rather than **necessary and sufficient conditions**. We'll start by summarising the evidence, and then consider its implications for lexical semantics and pragmatics.

2. Prototypicality effects

Prototypicality effects (or typicality effects) are the data that seem to cast doubt on the classical view of categorization and support an alternative, prototype-based view. According to this alternative view, saying whether an object belongs to a category is not a simple yes-no matter but a matter of degree: there are more or less typical members, better or worse examples. As Smith & Medin (1981) put it,

"Typicality effects reveal that not all members of a concept are equal, or to put it more positively, that concepts possess an internal structure that favours typical members over less typical ones."

Here are some illustrations, all of which have been experimentally confirmed:

1. Typicality rankings. People can not only make yes-no decisions about category membership, but also rank category members for typicality; e.g. for Americans, the most typical birds are robins and sparrows, less typical birds are hawks and eagles, and even less typical birds are ducks and geese; among mammals, deer and horse are most typical members, lion and cow are mid-typical, and mouse and pig low-typical (Smith & Medin: 34).

2. Category membership decisions. Decisions about category membership are fastest and most accurate for typical members. Americans answer 'Yes' faster to the question 'Is a robin a bird?' than to the question 'Is a duck a bird?', and respond faster to a picture of a robin than to a picture of a duck when asked, 'Is this a bird?'

3. Acquisition. Children are able to categorise typical category members earlier than less typical members: e.g. they learn that a robin is a bird before they learn that a duck is a bird. More generally, concept acquisition is normally based on experience with typical members.

4. Listing. When people are asked to list members of a category, they name the most typical members first. With birds, they list sparrows and robins before ducks and geese.

5. Inference. Typicality judgements seem to play a role in inference. For example, people are more likely to infer that a disease will spread from typical to less typical members of a category than the other way around (e.g. they judge that a disease will spread from the robins on a certain island to the ducks, rather than from the ducks to the robins).

6. Listing typical properties. People can list not only typical **members** of a category, but also typical **properties** of a category member. They will tell you that a typical chair is made of wood, has four legs and so on, even though these are not defining features (necessary and sufficient conditions) of chairs. They can tell you about a typical soldier, or mother, or bachelor, or restaurant, or best-seller writer. Indeed, they are remarkably good at doing this, while, as we've seen, they are on the whole quite bad at producing definitions. Such lists of typical properties are sometimes called **stereotypes**.

There is also a wide range of intuitive evidence based on **problems of categorization** – cases where we would hesitate for a moment or even be unable to decide which category an object belongs to – which seems to back up these experimental findings. I'll roughly classify the examples in the way that many prototype theorists do.

1. Scalar or comparative concepts. Take a concept like RICH, or FAR, or RELEVANT, which admits of degrees. The richer someone is, the easier we find it to judge that he is rich, and so on for other comparative concepts. Presumably we'd all agree that the Queen is rich; difficulties and hesitations arise the *lower* someone is on the scale of richness.

2. Non-typical cases. If we have a mental stereotype of a bachelor – a mental representation of the typical bachelor – then the closer someone comes to fitting the stereotype, the less hesitation we will have in saying that he is a bachelor. Questions like 'Is the Pope a bachelor?' create hesitation because he's non-typical on a number of counts – age, eligibility, freedom to marry, lifestyle, etc. Similarly, people hesitate when asked 'Is Tarzan a bachelor?'

3. Borderline cases. Take an object that is midway between being a cup and a mug – i.e. it is not a typical member of either category, but shares equal numbers of typical properties of both. We'll be likely to hesitate about which category to put it in.

4. 'Family resemblance' cases. The term 'family resemblance' comes from Wittgenstein, who argued that there is no common set of defining features that apply to all and only games, but that our various uses of the term resemble each other in the way family members do. Lakoff generalises this idea to a wide variety of cases. For example, he argues that a word like 'mother' has not a single prototype but a cluster of different prototypes: biological, social, legal, etc. He mentions stepmothers, adoptive mothers, birth mothers, foster mothers, biological mothers, surrogate mothers, unmarried mothers and genetic mothers as clustering round a central case where all the prototypes converge and we have someone who fits all the biological, social and legal roles simultaneously. There is no common thread which runs through all cases of mothers – each group shares some properties with others, and differs in some properties – hence the name 'family resemblance'.

5. Conflicting criteria. People may hesitate when asked whether a tomato is a fruit or a vegetable, because we have conflicting sets of criteria for deciding. A tomato is a fruit by biological criteria, but ranks for culinary purposes as a vegetable. This conflict leads to categorisation problems. Similarly, we might hesitate about a piece of furniture which was designed as a dressing table but is used as a desk.

Prototype theorists argue that prototypicality effects are best explained by assuming that the internal structure of a concept consists, not of a set of necessary and sufficient conditions for category membership, as on the classical view, but of a set of typical properties of category members (what we called a **stereotype** in the last section). The crucial difference between a stereotype and a definition is that stereotypical features are not necessary but

merely probable: for example, a stereotypical (but not a defining) feature of ‘bird’ would be FLIES; a stereotypical (but not a defining) feature of ‘lemon’ would be YELLOW, TASTES BITTER (it is merely a statistical likelihood, not a certainty, that a given bird will fly, or a given lemon will be yellow). On this approach, categorisation is seen as based on recognising similarities (i.e. sharing of features) between the item to be categorised and the stereotype, or prototype. The closer the similarity, the easier the categorisation; the greater the dissimilarity, the greater the hesitation.

3. <u>Prototypes and semantics</u>

Prototypicality effects clearly exist. The question is, what light do they shed on the nature of word meanings, the structure of concepts and the lexical-pragmatic processes we looked at in Lecture 1? Prototype **semanticists** claim that prototypicality effects give us direct insight into the nature of word meanings. Word meanings are prototypes or stereotypes, they say. On this approach, both the classical theory of word meaning and Fodor’s atomic concept approach are wrong. We can thus avoid Fodor’s worrying conclusion that most word meanings are innate.

Let's consider, then, whether prototype theory provides an adequate account, not just of **categorisation**, but of **word meaning**. The crucial requirement on an account of word meaning is that it makes it possible to construct a **compositional semantics**, which assigns meanings to an infinite set of phrases and sentences on the basis of word meaning plus syntactic structure. The question is whether prototype theory provides an adequate basis for a compositional semantics. Fodor and others have argued convincingly that it does not.

The main problem is that if the meaning of a word is a prototype (or stereotype), then the meaning of a phrase should be a complex prototype built up from the prototypes of its constituents. But this assumption leads to unacceptable results. So, for example, we all have a prototype of 'pet' – something furry, four-legged, cuddly, etc., and we all have a prototype of 'fish' – a trout, for example. The problem is that we also have a prototype of a 'pet fish' – say, a goldfish – and this is *not* made up by combining the prototypes of its constituent parts. A goldfish is not a furry, four-legged, cuddly trout. This is the **combinatorial problem** for prototype theory, and appears to disqualify it as a serious proposal about word meanings. (See Fodor, 'Present status of the innateness controversy', pp. 292-8.)

Fodor gives a further type of counterexample, where we seem entirely to lack a prototype for a complex expression. We may, he says, have a prototype for grandmothers, but we surely don't have a prototype for 'grandmothers most of whose grandchildren are married to dentists'; we may have a prototype for cities, or even American cities; but we surely don't

have a prototype for 'American cities situated on the East Coast just a little south of Tennessee.' Nonetheless, we can understand these phrases: they have linguistic meanings. So prototypes must be distinct from linguistic meanings, and prototype theory doesn't provide an adequate basis for linguistic semantics.

The classic paper on the combinatorial problem for prototype theory is by Osherson and Smith (1981). They take the predictions of fuzzy set theory, which had been claimed to provide an adequate compositional semantics for prototype theory, and show that it is inadequate to deal with the creation of complex concepts out of simpler concepts. A basic claim of fuzzy set theory is that if something is a bad example of category A, and a bad example of category B, it can't be a *good* example of a complex category A+B. One counterexample that Osherson and Smith consider is an apple with stripes. They point out that this is a good example of a STRIPED APPLE, but it is not a good example of an APPLE (since apples aren't typically striped) and not a good example of a STRIPED THING (since striped things aren't typically apples). So the predictions of fuzzy set theory are wrong.

There are many further problems. As we've seen, prototype theorists tend to reject the classical view of concepts. They argue that there are no necessary and sufficient conditions for category membership, no strict boundaries around a category, but merely degrees of similarity to the prototypical representation. Now consider some predictions that seem to follow from this view. Notice first that squirrels have some degree of resemblance to grizzly bears. Then if there really are no category boundaries, people should think of squirrels as grizzly bears to some degree; should consider squirrels as examples when asked to provide evidence for or against a claim about grizzly bears, should perhaps answer 'yes' instead of 'no' to the question 'Is a squirrel a grizzly bear?', and so on; which of course they don't. Indeed, among the findings on prototypicality effects is the fact that typicality judgements generally make no difference to people's ability to tell you, for example, that a robin or a chicken is *not* a fish. The fact that, however much they hesitate, people are generally able to tell you that a duck *is* a bird and a butterfly or an aeroplane *isn't*, suggests that mere similarity to a prototype isn't the only thing that is determining their judgements.

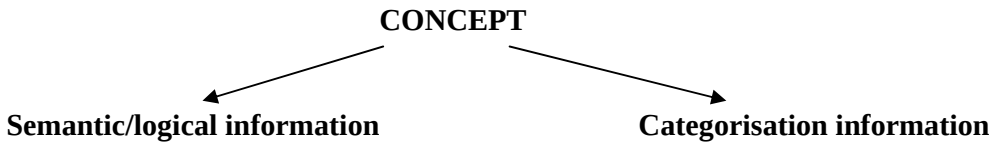
The other classic refutation of prototype theory extends these ideas. Armstrong, Gleitman and Gleitman (1983) showed that clearly definable expressions such as 'odd number' exhibit prototypicality effects. That is, in experimental situations, people judge some odd numbers to be more typical, or better examples of the category, than others. Now not only is 'odd number' definable, but people generally know the definition. So there cannot be an

incompatibility between a concept's being definable and a concept's having a prototype structure. This is obvious, too, from the fact that 'bachelor' has a prototype, and creates prototypicality effects, even though we all know its definition.

The best conclusion seems to be that prototype theory doesn't tell us anything about **word meanings**. Moreover, if prototype representations exist, they exist not as a **substitute** for necessary-and-sufficient conditions but as a **supplement**. Still, the existence of prototype effects does show us *something* about the structure of the human mind, and these effects need to be explained. In the next section, we'll look at an alternative view of the place of prototypes in cognition, not as word meanings, but as part of the encyclopaedic entries of concepts.

4. Prototypes and the encyclopaedia

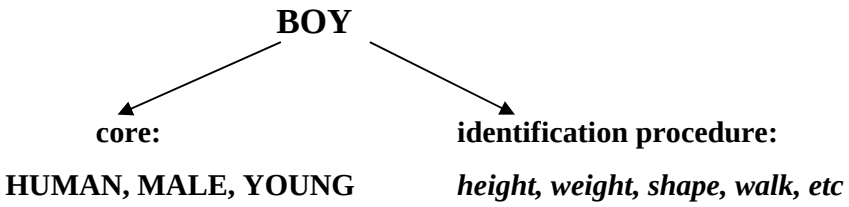
The failure of prototype semantics points to an obvious conclusion: that what we have been calling the concept of BIRD, or BOY, or CAR, gives us access to two different types of information: (a) semantic or logical, and (b) classificatory or categorising. That is, the role of concepts as word meanings and their role in categorisation come apart:



The semantic/logical information associated with a concept could be seen as a set of necessary and sufficient conditions, as in the classical view, or a set of meaning postulates, as in Fodor's account, while the categorisation information would look more like a prototype or stereotype, and would account for prototypicality effects. It would follow that the information we use in categorisation needn't necessarily be semantic or logical, or give us any insight into word meanings. To that extent, we would have to abandon the classical view of concepts, on which their linguistic, logical and classificatory functions all depend on their definitions; but for semantic purposes we could still use some elements of the classical or Fodorian accounts.

In recent years, it's become fairly standard to talk of the information we use in categorisation tasks as **encyclopaedic** rather than semantic (where encyclopaedic information is linked to the concept in some way, but doesn't tell us anything directly about the **meaning** of the associated word. And if prototype theories of concepts aren't suitable for a semantic role, it seems reasonable to consider whether they can perform an encyclopaedic role, and contribute to an account of how encyclopaedic knowledge is represented and accessed. You'll find a sketch of an account along these lines in Smith and Medin (pp 18-21 and 59-60). They

distinguish between the **core** of a concept and its **identification procedure**, where the core seems to represent semantic knowledge and the identification procedure is encyclopaedic. Thus, they equate the core of a concept with Frege's notion of sense, and analyse it in terms of necessary and sufficient conditions: for example, on their account, the core for 'boy' consists of the features HUMAN, MALE and YOUNG, as on the classical view of word meaning. The identification procedure, by contrast, is what explains our categorization abilities: for example our ability to recognise boys and distinguish them from other things in the world. This typically involves much more superficial properties than necessary and sufficient conditions. For example, the identification procedure for boys might contain information about typical heights, weights, body proportions and other perceptual characteristics of boys. This is where prototypes would come in:



Let's consider, then, the view that in your encyclopaedic entry for the concept BIRD, apart from all the idiosyncratic knowledge you have about particular birds – the bird you saw on your lawn this morning, and so on – you have a representation of the prototypical bird. The idea that encyclopaedic entries contain knowledge of typical objects and events is very widely accepted, in both psychology and AI. Some people talk of **schemas**, others of **frames**, **scenarios, scripts** or **stereotypes,** but the basic idea is the same: our representations of the world include information about what is **expected, typical, normal**. Generally, this information, unlike our more idiosyncratic knowledge, is seen as stable across individuals and across times, and is sometimes seen as underlying our ability to understand each other at all.

You might find it hard to see how anyone could object to such a picture, but if you read Larry Barsalou's paper 'The instability of graded structure: implications for the nature of concepts', I doubt if you'll find it so appealing by the end. By 'graded structure', Barsalou simply means the typicality rankings, from most to least typical members, which are what prototype accounts of concepts are designed to explain. His experimental findings considerably extend our knowledge of prototypicality effects; in my view, they also cast serious doubt on the idea that encyclopaedic entries simply contain ready-made prototypes which are automatically activated when the associated word is heard. I'll summarise the most

striking points below.

(a) Variation across individuals

Early experiments on prototypicality effects reported a very high agreement between subjects on typicality rankings: over .9, which is about as high as you can get: it means that if you know one person's typicality rankings you will be able to predict about 90% of other people's rankings. Barsalou questions the statistics used in those experiments, and ran experiments of his own, where he found that agreement between subjects averages at .5 – a substantially lower result. He also points out that these results were obtained from fairly homogeneous populations of experimental subjects (American undergraduates), and that with less homogeneous populations, agreement between subjects might be expected to drop. His first conclusion is that 'Across individuals, graded structure is relatively unstable'.

(b) Variation within individuals

He then decided to test the stability of a given person's typicality rankings over time. Again, he expected to find little variation. What he found was that if you know someone's typicality rankings at a given moment, you will be able to predict 85% of their rankings after an hour, 75% after a day, and only 65% after a week. The greatest instability was among the middle-ranking members of a category, and the greatest stability was in most and least typical members. So within individuals, graded structure is relatively unstable.

(c) Variation across contexts

Barsalou also shows that a variety of contextual factors can alter typicality rankings, and that linguistic (or discourse) context plays a decisive role. For example, in the context of a discussion of milking, 'cow' and ‘goat' are judged more typical animals than 'horse' and 'mule', but in a discussion of riding, 'horse' and 'mule' are judged more typical than 'cow' and 'goat'.

(d) Variation in criteria used

He also reports an experiment which showed that different criteria for judging typicality were used in different contexts. So far, we've been assuming that similarity to the prototype is the sole measure of typicality, but Barsalou suggests that there are at least two more. He distinguishes between **prototypical** and **ideal** members of a category – think of the difference between the typical man and the ideal man, or the typical woman and the ideal woman – and argues that in certain types of context, similarity to the ideal rather than to the prototype is the best predictor of prototypicality effects. He also shows that, again contrary to what early experiments suggested, **frequency of exposure** to certain items in a category will affect prototypicality judgements: for example, the more you are exposed to eagles as an example of

birds, the more you will think eagles are typical birds.

(e) Variations in point of view

More surprisingly, he shows that subjects are quite capable of generating typicality judgements not only from their own point of view but from (what they take to be) the point of view of others. So, for example, subjects asked to rank birds for typicality from the American point of view ranked 'robin' and 'eagle' as typical, whereas those asked to rank birds for typicality from the Chinese point of view ranked 'swan' and 'peacock' as typical. Subjects asked to rank animals for typicality from the forest ranger's point of view saw the typical animal as wild and large, whereas when asked to rank animals from the pet shop owner's point of view, the typical animal came out as tame and small. (There is no suggestion, of course, that these predictions are accurate: merely that people have no difficulty making them.) Although Barsalou doesn't note it, this is further evidence for a 'mind-reading' ability (the ability to attribute mental states to others), and it would be interesting to investigate whether it develops and breaks down along similar lines to other such abilities (Baron-Cohen 1995).

(f) Ad hoc concepts

Barsalou also found that people were quite capable of providing typicality rankings for ad hoc concepts (e.g. THINGS THAT CAN FALL ON YOUR HEAD, or GOOD THINGS TO STAND ON TO CHANGE A LIGHTBULB). An ad hoc concept is one that is made up on the spot, that the subject is most unlikely to have had ready-made before the experiment started. In this case, then, typicality rankings can't be explained by appeal to a ready-made prototype.

(g) Different prototypicality effects affected by different factors

Finally, Barsalou checked whether the range of prototypicality effects we looked at in section 2 do indeed vary together, so that they could all be explained by appealing to a single, ready-made prototype. Here again the answer was 'no'. In one experiment, for example, he found that the order in which individuals list items in a category (e.g. particular birds) correlates only .27 with their individual judgements about which birds actually are typical.

5. <u>Implications of Barsalou's findings</u>

What do these experimental findings show? They seem to Barsalou (and to me) to show clearly that the full range of prototypicality effects is not best explained by assuming that prototypes are stored ready-made in encyclopaedic entries of concepts. Arguments against the ready-made prototype view are:

l. It would require a very large number of prototypes for each category, since you would need prototypes for all the possible points of view and contexts you were likely to encounter.

2. It is extremely unlikely that American undergraduates have constructed ready-made prototypes for the 40 categories they were tested on from the Chinese point of view, the American point of view, the housewife's point of view, the farmer's point of view, and so on.
3. The ability to generate typicality judgements for novel categories such as THINGS THAT COULD FALL ON YOUR HEAD is not plausibly explained in terms of memorised prototypes.
4. The disagreement between results of typicality gradings and results of listing experiments (mentioned above) again suggests that no single prototypical account will work.

Barsalou suggests that his findings on prototypicality can be best explained, not by postulating ready-made prototypes in memory, but by assuming that the 'content' of a concept (including whatever determines typicality rankings) on a given occasion of use is constructed ad hoc out of the huge range of encyclopaedic information we have at our disposal – each use of a concept resulting in a slightly different combination of assumptions from encyclopaedic memory. When taking account of point of view, we simply assemble different sets of assumptions; when performing different types of typicality ranking, we assemble different sets of assumptions, when using a concept in inference or in classification, we assemble different sets of assumptions, and so on.

This notion of an ad hoc concept seems to connect up with some of the variations in lexical-pragmatic interpretation we found in Lecture 1. Recall examples like (4a-e):

(4) a. As I worked in the garden, a *bird* perched on my spade.
b. *Birds* wheeled above the waves.
c. A *bird,* high in the sky, invisible, sang its pure song.
d. At Christmas, the *bird* was delicious.
e. John opened the birdcage, and the *bird* flew across the room.

As we've seen, each occurrence of 'bird' is interpreted in a slightly different way, and it seems that Barsalou's notion of an ad hoc concept might help us to explain what is going on.

Barsalou's solution, of course, immediately raises a further question. What is it that determines which set of assumptions we will assemble for a given concept on a given occasion? Here Barsalou has some suggestions to make. As some of his experiments show, linguistic **context** clearly has an effect, and the **accessibility** of assumptions also clearly has an effect. These are, of course, typical pragmatic factors, which relevance theory (among other pragmatic theories) has looked at in some detail. And what determines when you have assembled the right set of assumptions? Here, Barsalou suggests that considerations of **relevance** play a decisive role. Though he says nothing about how relevance can be defined, this again suggests that relevance theory might be able to shed some light on at least some of

the problems that arise in this area. Thus Barsalou's work on categorisation, combined with relevance-theoretic pragmatics, might provide new insight into the working of lexical-pragmatic processes.

Homework

1. Write down the first 15 (a) flowers, (b) items of furniture, (c) musical instruments, and (d) makes of car you can think of. (We'll compare the results in backup class).

2. Make up a prototype representation (in terms of features) which might predict these results. What problems (if any) did you find in constructing these prototypes?

3. Suggest a possible core and identification procedure (along the lines of Smith & Medin's for BOY), for (a) ANIMAL (b) DOG and (c) ALSATIAN. Would the identification procedures remain the same in all circumstances and for all purposes? If not, how might they vary, and why?

Reading

Margolis, E. & S. Laurence 1998. *Concepts: Core Readings,* Chap 1, section 3, The prototype theory of concepts.
Fodor, J. 1981. The present status of the innateness controversy, pp. 292-8. (arguments against prototype theory)
Barsalou, L. 1987. The instability of graded structure: implications for the nature of concepts. In U. Neisser (ed) *Concepts and Conceptual Development.* CUP.

Background references

Rosch, E. 1978. Principles of categorization. In Rosch & Lloyd (eds) *Cognition and Categorization*. (Also in Margolis & Laurence 1998) (the classic work on prototypes)
Lakoff, G. 1987. Cognitive models and prototype theory. In U. Neisser (ed.) *Concepts and Conceptual Development*. (Also in Margolis & Laurence 1998) (dept file) (pro-prototypes)
Armstrong, S., L. Gleitman & H. Gleitman 1983. On what some concepts might not be. *Cognition* 13: 263-308. (Also in Margolis & Laurence 1998) (arguments against)
Osherson, D. & Smith, E. 1981. On the adequacy of prototype theory as a theory of concepts. *Cognition* 9, 1: 35-58. (Also in Margolis & Laurence 1998) (arguments against)