ISSUES IN PRAGMATICS (PLIN 3001) 2006-07

LEXICAL PRAGMATICS

5. Metaphysical and cognitive functions of concepts

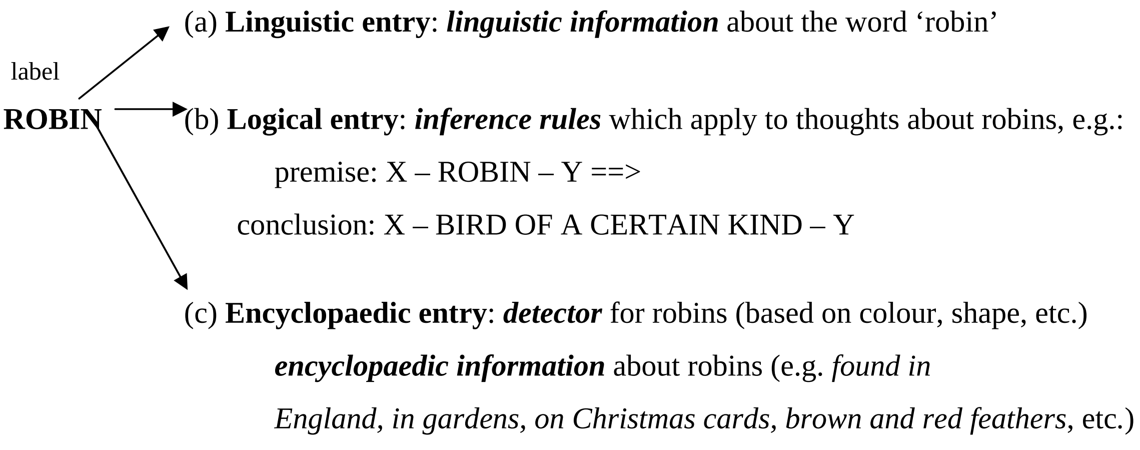
1. Introduction

We've now considered several accounts of concepts, and I'd like to summarise some tentative conclusions I think we've reached. Let's look first at the role of concepts as word meanings. Here, Fodor has some good arguments against the classical decompositional approach. Notice, though, that his own experimental evidence suggests that what he called 'implicitly negative' expressions such as *doubt* and *deny* (unlike *bachelor*) do behave as if they're semantically negative, and might therefore be mentally represented as BELIEVE NOT and SAY NOT. So at least some words decompose. Moreover, even a classical theorist must concede that decomposition has to stop somewhere, and hence that at least some concepts are simple and unanalysable. So the dispute between Fodor and the classical theorists is really about what proportion of the vocabulary encodes simple, unanalysable concepts, and what proportion of the vocabulary can be defined and has its meaning represented as a definition. Notice, too, that as we saw in lecture 3, even concepts which Fodor treats as 'simple and unanalysable', such as BIRD, BACHELOR and KILL, may provide access to linguistic, encyclopaedic or logical information (e.g. meaning postulates, encyclopaedic entries) which plays a role in categorisation, inference and language processing; so a 'simple, unanalysable' concept isn't as unstructured as we might have thought. I'll return to this point.

In lecture 3, we looked at the prototype approach to concepts, and saw that there are good arguments against the view that the meanings of words, phrases and sentences are **generally** prototypes. For example, when a phrase like *pet fish* has a prototype, it generally isn't constructed out of the prototypes for the constituent words; moreover, most phrases and sentences don't have prototypes, and we want words to have the sort of meanings that can be compositionally combined into phrase and sentence meanings. The conclusion seems to be that prototypes contribute not to the **semantic** or **logical** function of concepts, but to their **categorising** function, and we'll continue this argument today.

As regards the role of concepts in categorisation, we've seen so far that there are two possible positions: (a) according to the classical view, categorisation is a matter of satisfying a definition; or (b) according to the approaches we looked at last week, categorisation is a matter of fitting a prototype or stereotype (i.e. a set of typical, often perceptual, features), or, more generally, depends on encyclopaedic information accessible in memory. We’ve also seen that these views aren’t incompatible: a category like BACHELOR or ODD NUMBER may have both a definition and a prototype. What conclusion should we come to about this? Today I want to argue that there are in fact two different views of what categorisation is for, and two corresponding views of the relation between concepts and objects: the **cognitive** view, which is the one we've been concerned with so far, and a more philosophical, or **metaphysical**, view. It’s this distinction between cognitive and metaphysical approaches that we’ll look at today.

Notice, first, that all the literature we've looked at so far has dealt with the **psychological** or **cognitive** roles of concepts. We've been taking for granted that concepts are mental representations, which can be acquired and lost, which can act as the input to mentally-represented inference rules, and perform other functions within the individual's cognitive system. We could use the following model from relevance theory to show how a mentally represented concept might be structured in order to perform such functions (see Sperber & Wilson 1986/95, chapter 2, section 4):



label

ROBIN

(a) **Linguistic entry**: ***linguistic information*** about the word ‘robin’

(b) **Logical entry**: ***inference rules*** which apply to thoughts about robins, e.g.:

premise: X – ROBIN – Y ==>

conclusion: X – BIRD OF A CERTAIN KIND – Y

(c) **Encyclopaedic entry**: ***detector*** for robins (based on colour, shape, etc.)

encyclopaedic information about robins (e.g. *found in England, in gardens, on Christmas cards, brown and red feathers*, etc.)

Here, all the information, in all the entries, is seen as mentally represented and accessible to the individual who hears the word 'robin', or is processing the concept ROBIN. The **label ROBIN** is the mentally-represented conceptual address that is activated by hearing the word ‘robin’, or seeing a robin. The **linguistic entry** links the word ‘robin’ to the conceptual address ROBIN, which in turn gives access to logical and encyclopaedic entries. The **logical entry** contains one-way inference rules (e.g. Fodor’s **meaning postulates**), which draw conclusions from thoughts about robins whose truth is guaranteed as long as the premises are true. The **encyclopaedic entry** contains an **identification procedure**, or **detector,** for robins, which is automatically activated by the sight or sound of a robin; as we saw last week, it also contains a vast reservoir of **encyclopaedic information** about robins, which may be added to the context in which thoughts about robins are processed.

This model raises a number of questions which have not generally been discussed by linguists or psychologists, and which fall within the province of philosophy. Today I want to look at these questions, which are discussed in the article 'Concepts and Stereotypes', by the philosopher Georges Rey. This article was written in response to Smith & Medin's work on prototype theory, and accuses psychologists of ignoring two further properties, or functions, of concepts that are crucial to philosophers.

2. Rey on the stability of concepts

A crucial feature of concepts, according to Rey, is that they are recurring ingredients of thought. They impose a structure on our mental life: they enable a given person to have the same thought at two different times, and different people to have the same thought at the same or different times. We can understand his claim better by drawing two contrasts. The first is between **concepts** and **percepts**, or having the **concept** RED and having the **perceptual** ability to discriminate different shades of red without actually conceptualising them. The second is between having the **stable** concept RED and having the **temporary** conceptual ability to form different thoughts about, e.g., two different shades of red, two neighbouring blades of grass, two ants, two dinner plates in a rack, without setting up a permanent conceptual address for them. Philosophers have emphasised that rational thought, rational argumentation, would simply not be possible if we did not have concepts that (a) have logical (inferential, as opposed to perceptual) properties, and are thus capable of playing a role in valid inference, and (b) are stable across individuals and times. Thus, consider the following argument:

Premise 1: If it's raining, I'll stay at home.
Premise 2: It's raining.
Conclusion: I'll stay at home.

This argument would not be valid if the word *raining* merely encoded a percept (with no logical properties), or if the word *rain* changed its meaning between Premise 1 and Premise 2, or the phrase *stay at home* changed its meaning between Premise 1 and the Conclusion. But if the words are not to change their meaning, the contents of the concepts RAINING and STAY

AT HOME must also remain stable. More generally, two people could not agree or disagree about anything unless the thoughts they agreed or disagreed about had the same content. Thus, any attempt to study human cognition or communication rests on the assumption that concepts do have **contents** which remain relatively **stable** across individuals and times. We must explain how this stability is achieved.

Let's return now to our cognitive model of the concept ROBIN. This consists of a conceptual address, or label, ROBIN, and various types of information, linguistic, logical and encyclopaedic, that this label gives us access to. All this information is represented inside the head. According to the **internalist** approach to concepts, we can account for their stability and their contents by looking only at cognitive information, information represented inside the head. We might say, for example, that two people share the concept ROBIN when they share the conceptual address, or label, ROBIN. But how do we know when two people share a label? How do we know that your label for ROBIN doesn't refer to sparrows while mine refers to robins? How do we know that my label, which used to refer to robins, has not now wandered off to refer to sparrows? Could someone be said to have the concept ROBIN if he'd acquired the label but thought it referred to typewriters? It seems that we can’t account for the stability or content of concepts without looking at the objects in the world that a given concept picks out: in other words, by looking outside the head.

A slightly different version of the internalist approach might claim that the content of concepts, and hence their stability, is guaranteed by their mentally-represented encyclopaedic entries, which are linked to the cognitive categorising function. On this approach, two people would share the concept ROBIN when they share not just the label but also the encyclopaedic information that distinguishes robins from sparrows and typewriters, for example. One problem with this is that encyclopaedic entries themselves consist of labels, whose content needs to be established and stabilised in the way described above. Another problem is that encyclopaedic entries change all the time. My encyclopaedic entry for ROBIN changes every time I see another robin; even prototypes seem to vary from situation to situation, as we saw last week. Moreover, it’s unlikely that any two individuals have identical encyclopaedic entries for any concept. So if we define *concept* in this very broad sense, we will lose the stability across individuals and times that seems to be an essential feature of the role we want concepts to play in an explanatory psychology. Again, what seems to be needed is a notion of conceptual content which is somehow independent of individual psychology, with all its variations.

This discussion of the stability function of concepts thus raises a further question: how can concepts have stable enough contents to account for their role as recurring ingredients of thought? A possible answer is to take an **externalist** view of concepts, which claims that their contents are somehow independent of individuals, and what individuals know or believe about the objects that fall under them. It's this more abstract notion of concepts that philosophers have mainly used. Georges Rey, in his article on concepts and stereotypes, talks of concepts as having a **metaphysical** function in addition to their cognitive functions. This is what we'll look at next.

3. Rey on the metaphysical function of concepts

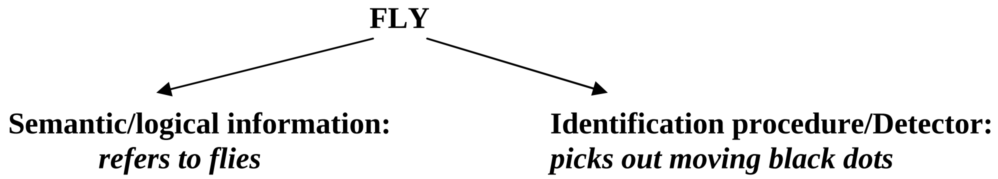
To introduce the metaphysical function of concepts, let's return to the question of how objects are categorised. Rey points out that there are in fact two different questions that we can ask in categorising an object – say, a bird:

(a) How do we **recognise** birds?
(b) When **is** something a bird?

Prototype theory is a response to question (a): it's designed to explain the prototypicality effects which arise when individuals are asked to *recognise* objects as belonging to a certain category. Question (a), as we've seen, is a **cognitive** question. Philosophers would call it an **epistemological** question: epistemology is the study of human knowledge, its acquisition and exploitation.

Question (b), by contrast, is a **metaphysical** question. Metaphysics is the study not of how we *think* the world is, but of how it really is. Metaphysics and epistemology do not always coincide: for example, the world might be a certain way without our knowing it, or we might think the world is a certain way when it's not. In particular, they need not coincide in categorisation: for example, we might all think a certain animal is a zebra, when in fact it's a horse with painted stripes.

There are well-known examples of such mismatches, from both philosophy and cognitive science. Here's an example from cognitive science. Consider the frog, which is interested in catching flies. We might think of it as having the concept FLY, which denotes the set of possible and actual flies, and whose function is to trigger an appropriate plan of action whenever a fly crosses the frog's field of vision. But what is the mentally-represented identification procedure, or detector? How does the frog recognise a fly? Experiments have shown that the frog’s fly-catching reflex is triggered by any appropriately-sized black dot that moves across the fly’s field of vision. In other words, though the **function** of the frog’s concept is to pick out all and only flies, the **identification procedure** picks out moving black dots, not all of which will turn out to be flies:



FLY

Semantic/logical information:
refers to flies

Identification procedure/Detector:
picks out moving black dots

The question that then arises is: Why do we say that the frog has the concept of a fly, rather than the concept of a moving black dot? Some philosophers want to say that it’s the concept of a fly (assuming it’s a genuine concept rather than a percept – see above) because its **function** is to pick out actual flies. If there were no flies in the world, the frog would not have this concept. The frog isn’t interested in little black dots except as indicators of the presence of flies. For that reason, we can describe the identification procedure as a ***fly***-detector rather than a ***little-black-dot*** detector, even though it is triggered by all and only little black dots.

What this suggests is that the information stored in the identification procedure for a concept is not necessarily the best guide to the **content** (or **denotation**) of the concept, and that someone who lacks the correct identification procedure might still be said to have the concept and understand the meaning of the associated words. In any case, this seems to be Fodor’s position. He does not deny that concepts may have associated (learned) prototypes or identification procedures: he merely claims that these are not essential to the possession of the concept, and do not determine its **content.**

There are also some standard examples used in the philosophical literature to underline the distinction between metaphysics and epistemology (or cognition) (e.g. in Putnam 1975). One is the case of GOLD. We all know that gold is a metal of a certain kind, and we also know roughly what it looks like, that it's a precious metal, used to make rings, jewellery, gold plates, etc. That is, we have some sort of prototype or identification procedure for gold. However, most of us probably couldn't tell real gold from various sorts of fake. The same is true of silver, diamonds, watches, paintings, designer clothes, and so on. This is an epistemological, or cognitive, problem, a problem of **recognition**. None of us feels that our inability to recognise gold, etc. means that there's no such thing as gold.

If the answer to the cognitive question (how do we recognise gold?) is that we use prototypes (or detectors, or identification procedures), what is it that answers the metaphysical question? That is, what makes gold gold? Recall that at the beginning of the lecture I suggested two possible approaches to categorisation:

(a) Objects are categorised in terms of their **defining features**;
(b) Objects are categorised in terms of **similarity to the prototype**.

Rey suggests that while method (b), based on prototypes, may be used to answer the cognitive question, it's method (a), a theory of the essential features of objects (at least of ‘natural kinds’ of objects, e.g. horse, tulip, gold, etc.), that answers the metaphysical question. In the case of gold, our current theory of the essential features of gold is a scientific one: gold is the metal with atomic number 79. Similarly, our theory of the essential features of horses as opposed to zebras will be a genetic one; with colours, we will be told a story about wavelengths, and so on. These theories may be wrong, and we can imagine not having a theory at all. But the idea behind the metaphysical approach to concepts is that at least some of them – particularly the natural-kind ones – *do* have essential features, whether anyone knows them or not, and it is this link between concepts and their contents (or denotations)– which are external to any particular individual – that accounts for the stability of concepts across individuals and times.

Returning to our two possible positions on the role of concepts in categorisation, it now appears that they are not incompatible after all. Concepts play two quite separate roles in categorisation: a metaphysical role, in which categorisation (at least for natural kinds) involves an appeal to essential features, and which accounts for the stability of concepts, and a cognitive role, in which categorisation involves an appeal to encyclopaedic information, perhaps including prototypical features. An adequate theory of concepts should deal with both these roles.

4. <u>Metaphysics and word meaning</u>

Let's return now to questions of word meaning. So far in this course, I've been taking a resolutely cognitive approach to word meanings. When I said that concepts function as the meanings of words, I was thinking of concepts as mentally represented entities that play a cognitive or epistemological role. This is the notion of concept that Fodor and the decompositionalists have in mind: they are disagreeing about our **knowledge** of word meanings, and about whether this knowledge amounts to a set of mentally represented necessary and sufficient conditions, a simple, unanalysable mental address, or what.

Now recall that one of Fodor's main arguments against the decompositionalists rested precisely on 'natural-kind' terms of the type we've been looking at today: *gold*, *horse*, *red*, and so on. Putnam, Rey, etc. argue that such natural-kind terms do have metaphysical necessary and sufficient conditions: that is, natural kinds have an underlying **essence** which makes them what they are. This raises two questions: (a) does the existence of such conditions add anything to the debate between Fodor and the classical theorists? and (b) does it alter our views on the meanings of words in any way?

(a) Fodor, the classical theorists and metaphysical definitions

Fodor's knock-down argument against the classical view of word meaning went as follows. To capture the intuition that *red*, *blue*, etc. are semantically related, you should decompose *red* into COLOURED + X, where X is the feature that distinguishes *red* from all the other colours. But this commits you to finding a feature X that means RED BUT NOT COLOURED – a logical impossibility. Similar arguments apply to *horse*, *gold*, etc.

Given today's claim that natural-kind terms have metaphysically necessary and sufficient conditions, you might wonder whether a classical theorist couldn't argue that the missing feature 'X' can be defined in terms of these. Thus, *red* might be decomposed into COLOURED + OF WAVELENGTH X, *gold* might be decomposed into METAL + OF ATOMIC NUMBER 79, and so on. This may be what many decompositionalists have in mind, but such a solution would not be adequate, because it confuses epistemology with metaphysics. We want to be able to say that *gold*, *red* and *horse* mean the same now as they did in Shakespeare's day, when our current theories of genetic structure, atomic structure, etc. did not exist, and so no-one knew them. We want to say that people can know the meanings of *gold*, etc. even though they don't know the associated scientific theories. This is, of course, the whole point of distinguishing metaphysics from epistemology. To try to rescue a cognitive theory – the classical view of word meanings – by appealing to metaphysically necessary and sufficient conditions would be a mistake.

By the same token, Fodor's claim that the meanings of words like *red*, *horse*, *gold*, etc. are simple, unanalysable concepts is quite compatible with the philosophers' arguments that these terms have metaphysically necessary and sufficient conditions. Fodor is talking about the structure of our **mental representations**; the philosophers are not.

(b) Word meaning and metaphysically necessary and sufficient conditions

The metaphysical notion of a concept, though, does raise a more general question about the nature of word meanings. As mentioned above, I've been assuming so far this term that word meanings are purely cognitive, mentally represented entities. Meanings, on this account, are **internalised**: 'in the head'. To establish whether Peter and Mary assign the same meaning to the word *gold*, all we need to do, in principle, is look inside their heads. As we saw, there are problems about this, because it's hard to see whether two people share a concept by looking at cognitive factors alone.

A number of philosophers have argued that the best solution to these problems is to move to a metaphysical notion of concepts, and an externalist conception of word meanings. Putnam is one of these. What really determines the meaning of a word, he says, is its metaphysically necessary and sufficient conditions (i.e. the set of objects it picks out in the actual world and other possible worlds). Since people may not know these metaphysically necessary and sufficient conditions, it follows that 'meanings aren't in the head'.

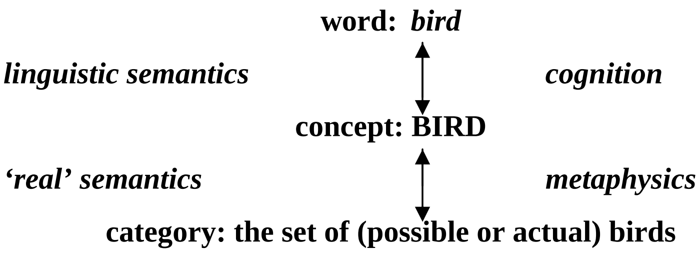
Here are two of Putnam's examples:

(a) Putnam knows the words *elm* and *beech*, and he knows that they are kinds of tree, but he doesn't know how to tell them apart. He leaves that to the experts. The experts know the metaphysically necessary and sufficient conditions. They are the ones who know the meanings of the words. In other words (he says), meanings are independent of the individual's cognitive system. What makes my concept ELM the same as yours is that fact that they both refer to elms (whether we know how to recognise elms, i.e. whether we have an identification procedure for elms, or not). This argument applies to natural-kind terms and any other terms with metaphysically necessary and sufficient conditions.

(b) Imagine that there is a planet called Twin Earth, which is like Earth in all respects but one: the colourless, tasteless liquid which on Twin Earth is called *water*, which runs in rivers, comes out of taps and falls from the sky as rain, is composed not of H_2O but of some different chemical formula, say XYZ. **Question**: Does the word *water* on Twin Earth mean the same as the word *water* on Earth? The answer, according to Putnam, is 'no'. *Water* on Earth refers to water, i.e. the substance whose metaphysically necessary and sufficient condition is that it is H_2O. On Twin Earth, by contrast, *water* refers to XYZ – an entirely different substance, with a different metaphysical definition. Even if no-one on Earth or Twin Earth is aware of the difference, the words would still have different meanings, because they refer to different substances. Even if we imagine identical twins, Oscar on Earth and Twin-Oscar on Twin Earth, who have had identical experiences, identical life histories and are in identical brain states, with identical mentally-represented concepts WATER, when Oscar says *water* he means H_2O, but when Twin-Oscar says *water*, he means XYZ. Hence, meanings aren't in the head.

These arguments in turn suggest an **externalist** view of meaning, with meanings being determined by something external to the individual, in contrast to the **internalist** (cognitive) view of meaning which I have been assuming so far. One way of reconciling the two positions would be to break down the analysis of word meaning into two stages: (a) **linguistic semantics** would be a purely cognitive matter, relating words to mentally represented concepts (b) **'real' semantics** (i.e. the semantics of mental representations) would be a metaphysical matter, relating mentally represented concepts to their contents, which are sets of actual or possible objects in the world. The debate between Fodor and the decompositionalists arises at stage (a), and is about the nature of mentally represented concepts. The Putnam arguments apply to level (b), at which we have to provide a theory of how our mental representations relate to objects in the world, and account for their stability across individuals and times. Here, Fodor broadly agrees with Putnam and takes an externalist aproach.

On this approach, the analysis of sentence meaning would involve the same two stages: (a**) linguistic semantics** would involve translating natural-language sentences into a mentally represented conceptual representation system, or language of thought; (b) **truth-conditional semantics** would be a theory of how our conceptual representations get their contents, by relating to states of affairs in the world. (Both Fodor and Sperber & Wilson take this two-stage view.):



word: ***bird***

linguistic semantics ***cognition***

concept: BIRD

'real' semantics ***metaphysics***

category: the set of (possible or actual) birds

5. Conclusion

The main point that I've tried to make today might be summed up as follows. When we talk of the categorising function of concepts, we may have two very different things in mind. We may mean a cognitive theory of how humans **recognise** objects as falling under concepts. This is what prototype theorists have generally been concerned with. Or we may mean a metaphysical theory of what objects actually **do** fall under concepts. This is what philosophers in general, and truth-conditional semanticists in particular, have been concerned with. I've suggested that in order to provide a fully adequate account of concepts, both metaphysical and cognitive functions will have to be taken into account.

Let me end by comparing what internalist and externalist approaches to concepts would say about the Twin Earth case. As we've seen, from a purely internalist point of view, Oscar and Twin-Oscar have the **same** concept, which performs the same cognitive role for both of them. From an externalist point of view, by contrast, Oscar and Twin-Oscar have **different** concepts, which refer to different substances. If externalists are right, two concepts will differ if they pick out different objects, even though they play identical cognitive roles. For internalists, two concepts cannot differ without playing different cognitive roles. Much of Fodor's recent work has been devoted to defending the externalist view of concepts, on which the content of a concept is determined by the objects that fall under it, and criticising the internalist view, on which the content of a concept is exhausted by its cognitive (or 'inferential') role. For Fodor, then, the claim that concepts are simple and unanalysable amounts to the claim that of all the mentally-represented information they carry, only their labels contribute to determining their contents (i.e. their metaphysically necessary-and-sufficient truth conditions), and these labels have no internal structure.

Homework

(a) Make sure you understand the notion of metaphysically necessary and sufficient condition, and the Twin Earth case.

(b) What might be the metaphysically necessary and sufficient conditions for WOLF, CAR, CHAIR, BOEING 747, PROMISE, BURGLARY, ANGER? How did you decide?

Reading

Rey, G. 1983. Concepts and stereotypes. *Cognition* 15: 237-62.

Background references

Putnam, H. 1975. The meaning of "meaning". In K. Gunderson (ed.) *Language, Mind and Knowledge*. University of Minnesota Press.

Sperber, D. & Wilson, D. 1986/95. *Relevance: Communication and Cognition*. Blackwell: Oxford.