9 The mapping between the mental and the public lexicon

Dan Sperber and Deirdre Wilson

1   Introduction

There are words in the language we speak and concepts in our minds. For present purposes, we can use a relatively common-sense, unsophisticated notion of a linguistic word. A bit more needs to be said, now and later, about what we mean by a concept. We assume that mental representations have a structure not wholly unlike that of a sentence, and combine elements from a mental repertoire not wholly unlike a lexicon. These elements are mental concepts: so to speak, 'words of Mentalese'. Mental concepts are relatively stable and distinct structures in the mind, comparable to entries in an encyclopaedia or to permanent files in a data-base. Their occurrence in a mental representation may determine matching causal and formal (semantic or logical) relationships. On the one hand, there are relationships between the mind and the world. The activation of a concept may play a major role in causal interactions between the organism and external objects that fall under that concept. On the other hand, there are relationships among representations within the mind. The occurrence of a concept in a mental representation may play a causal role in the derivation of further representations, and may also contribute to the justification of this derivation.

2   Three types of mapping

What kind of mapping is there (if any) between mental concepts and public words? One extreme view is that natural languages such as English or Swahili are the sole medium of thought. In this case, obviously, there is a genuine one-to-one correspondence between public words and mental concepts. An opposite extreme view is that there are no such things as individual mental concepts at all, and therefore no conceptual counterparts to public words. We will ignore these extreme views. We assume that there are mental concepts, and that they are not just internalisations of public words, so that the kind and degree of correspondence between concepts and words is a genuine and interesting empirical issue.

In principle, the mapping between mental concepts and public words might be exhaustive (so that every concept corresponds to a word and conversely), or partial. If it is partial, this may be because some concepts lack a corresponding word, because some words lack a corresponding concept, or both. The mapping between words and concepts may be one-to-one, one-to-many, many-to-one, or a mixture of these. However, the idea that there is an exhaustive, one-to-one mapping between concepts and words is quite implausible.

Some words (for instance the third person pronoun 'it') are more like place-holders and do not encode concepts at all. Many words seem to encode not a full-fledged concept but what might be called a pro-concept (for example 'my', 'have', 'near', 'long' – while each of these examples may be contentious, the existence of the general category should not be). Unlike pronouns, these words have some conceptual content. As with pronouns, their semantic contribution *must* be contextually specified for the associated utterance to have a truth-value. For instance, 'this is my book' may be true if 'my book' is interpreted as meaning *the book I am thinking of*, and false if it means *the book I wrote* (and since there are indefinitely many possible interpretations, finding the right one involves more than merely disambiguation). Similarly, whether 'the school is near the house' is true or false depends on a contextually specified criterion or scale; and so on. We believe that pro-concepts are quite common, but the argument of this chapter does not depend on that assumption (or even on the existence of pro-concepts). What we will argue is that, quite commonly, all words behave *as if* they encoded pro-concepts: that is, whether or not a word encodes a full concept, the concept it is used to convey in a given utterance has to be contextually worked out.

Some concepts have no corresponding word, and can be encoded only by a phrase. For instance, it is arguable that most of us have a non-lexicalised concept of *uncle-or-aunt*. We have many beliefs and expectations about uncles-or-aunts (i.e. siblings of parents, and, by extension, their spouses). It makes sense to assume that these beliefs and expectations are mentally stored together in a non-lexicalised mental concept, which has the lexicalised concepts of *uncle* and *aunt* as sub-categories. Similarly, people who do not have the word 'sibling' in their public lexicon (or speakers of French, where no such word exists) may nonetheless have the concept of *sibling* characterised as *child of the same parents*, and object of many beliefs and expectations – a concept which has *brother* and *sister* as sub-categories. So it seems plausible that not all words map onto concepts, nor all concepts onto words.

The phenomenon of polysemy is worth considering here. Suppose Mary says to Peter:

(1) Open the bottle.

In most situations, she would be understood as asking him to uncork or uncap the bottle. One way of accounting for this would be to suggest that the general meaning of the verb 'open' gets specified by the properties of its direct object: thus, opening a corked bottle means uncorking it, and so on. However, this cannot be the whole story. Uncorking a bottle may be the standard way of opening it, but another way is to saw off the bottom, and on some occasion, this might be what Mary was asking Peter to do. Or suppose Mary says to Peter:

(2) Open the washing machine.

In most situations, she will probably be asking him to open the lid of the machine. However, if Peter is a plumber, she might be asking him to unscrew the back; in other situations, she might be asking him to blow the machine open, or whatever.

The general point of these examples is that a word like 'open' can be used to convey indefinitely many concepts. It is impossible for all of these to be listed in the lexicon. Nor can they be generated at a purely linguistic level by taking the linguistic context, and in particular the direct object, into account. It seems reasonable to conclude that a word like 'open' is often used to convey a concept that is encoded neither by the word itself nor by the verb phrase 'open X'. (For discussion of similar examples from alternative perspectives, see Caramazza and Grober 1976; Lyons, 1977; Searle, 1980, 1983; Lehrer, 1990; Pinkal, 1995; Pustejovsky, 1995; Fodor and Lepore, 1996; Pustejovsky and Boguraev, 1996.)

So far, we have argued that there are words which do not encode concepts and concepts which are not encoded by words. More trivially, the existence of synonyms (e.g. 'snake' and 'serpent') shows that several words may correspond to a single concept, and the existence of homonyms (e.g. 'cat' or 'bank') shows that several concepts may correspond to a single word. So the mapping between concepts and words is neither exhaustive nor one-to-one.

Although the mapping between words and concepts is imperfect, it is not haphazard. Here are three contrasting claims about what this imperfect mapping might be like:

(3) Nearly all individual concepts are lexicalised, but many words encode complex conceptual structures rather than individual concepts. So there are fewer concepts than words, and the mapping is partial mostly because many words do not map onto individual concepts.

(4) Genuine synonyms, genuine homonyms, non-lexicalised

concepts and words that do not encode concepts are all
relatively rare, so there is roughly a one-to-one mapping
between words and concepts.

(5) The mapping is partial, and the main reason for this is that
only a fraction of the conceptual repertoire is lexicalised.
Most mental concepts do not map onto words.

In *The Language of Thought* (1975), Jerry Fodor famously argued against
(3) and in favour of (4). According to the version of claim (3) he criticised,
words correspond in the mind to definitions couched in terms of a relatively
compact repertoire of primitive concepts: for example, the word 'bachelor'
might have as its conceptual counterpart the complex mental expression
unmarried man; the word 'kill' might have as its conceptual counterpart the
complex mental expression *cause to die*, and so on. Many words – perhaps
most – would be abbreviations for complex conceptual expressions, rather
than encoding individual concepts. Against this view, Fodor argued that
most words have no plausible definitions and must therefore correspond to
mental primitives. There are psycholinguistic reasons for thinking that even
a word like 'bachelor', which seems to have a definition, is best treated as
encoding a single concept *bachelor*, which would itself be a (mental rather
than public) abbreviation for a complex mental expression (Fodor, 1975, p.
152).

As Fodor points out, verbal comprehension is fast, and unaffected by the
alleged semantic complexity of lexical items. 'Kill' is no harder to process
than 'die', and 'bachelor' is no harder than 'unmarried', even though it
might be argued that the meaning of 'die' is included in the meaning of
'kill', and the meaning of 'unmarried' is included in the meaning of 'bach-
elor'. All this suggests to Fodor that the structure of mental messages is very
close to that of the public sentences standardly used to communicate them.
'It may be that the resources of the inner code are rather directly represented
in the resources we use for communication' (Fodor, 1975, p. 156).

Fodor's argument against (3), combined with a rather traditional view of
linguistic communication, seems to weigh in favour of (4). Fodor does, as
he says himself, view language 'the good old way':

A speaker is, above all, someone with something he intends to communicate. For
want of a better term, I shall call what he has in mind a message. If he is to com-
municate by using a language, his problem is to construct a wave form which is a
token of the (or a) type standardly used for expressing that message in that language.
(Fodor, 1975, p. 106.)

Here, Fodor is adopting an updated version of what we have called the code
theory of verbal communication (Sperber and Wilson, 1986/1995). The
classical code theory was based on the following assumptions:

(6) For every thought that can be linguistically communicated, there is a sentence identical to it in content.

(7) The communication of a thought is achieved by uttering a sentence identical to it in content.

Assumption (7) is clearly too strong. Sentences with pronouns are obvious counter-examples: they are used to communicate different thoughts on different occasions, and are not identical in content to any of these thoughts. The updated code theory accepts (6), but rejects (7) in favour of the weaker assumption (8):

(8) The communication of any thought *can be* achieved by uttering a sentence identical to it in content.

For the classical code theory, the only way to communicate thoughts is to encode them. For the updated code theory, this is still the basic way, but there are also inferential short-cuts. The updated theory admits that the basic coding-decoding process can be speeded up, supplemented, or even by-passed by use of contextually informed inferential routines. Though full encoding is possible, the theory goes, it is often unnecessary. By exploiting shared contextual information and inferential abilities, communication can succeed even when a name or description is replaced by a pronoun, a phrase is ellipsed, or a whole thought is indirectly suggested rather than directly encoded.

Still, on both classical and updated versions of the code theory, the semantic resources of a language must be rich enough to encode all communicable thoughts. Every concept that can be communicated must be linguistically encodable. There may be a few non-lexicalised concepts (e.g. *uncle-or-aunt*) which are encodable only by a phrase; but it is reasonable to think that, in general, the recurrent use of a concept in communication would favour the introduction and stabilisation of a corresponding word in the public language.

Because Fodor uncritically accepts the code theory of communication, and because he does not even consider claim (5), let alone argue against it, his excellent arguments against claim (3) do not unequivocally point to the conclusion in (4). Claim (5) might still be correct. We want to argue that it is, and hence that most mental concepts do not map onto words.

There are two interpretations of claim (5) on which it would be trivially true, or at least easily acceptable. First, it is clear that the number of perceptual stimuli that humans can discriminate is vastly greater than the number of words available to name them: for instance, it has been claimed that we can discriminate anything up to millions of colours, while English has a colour vocabulary of a few hundred non-synonymous terms, only a

dozen of which are in frequent use (Hardin, 1988, pp. 182–3). If we have a concept for every colour that we can discriminate (or even for every colour that we have had the opportunity to discriminate), it is clear that we have many more concepts than words. However, a discrimination is not the same as a conceptualisation of the items discriminated. Someone may discriminate two shades of vermilion, and even think, *here are two shades of vermilion*, without forming a distinct mental structure, let alone a stable one, for each of these two shades.

A concept, as we understand the term, is an enduring elementary mental structure, which is capable of playing different discriminatory or inferential roles on different occasions in an individual's mental life. We are not considering ephemeral representations of particulars (e.g. an individual tree, an individual person, a particular taste), attended to for a brief moment and then forgotten. Nor are we considering complex conceptual structures, built from more elementary mental concepts, which correspond to phrases rather than words, and are not stored in long-term memory. Even so, it might be argued that people do form many idiosyncratic, non-lexicalised concepts on the basis of private and unshareable experience. For example, you may have a proper concept of a certain kind of pain, or a certain kind of smell, which allows you to recognise new occurrences, and draw inferences on the basis of this recognition, even though you cannot linguistically express this concept, or bring others to grasp and share it. More generally, it is arguable that each of us has ineffable concepts – perhaps a great many of them. This would again make claim (5) trivially true.

We will return to this point later, and argue that effability is a matter of degree. For the time being, we will restrict ourselves to effable concepts: concepts that can be part of the content of a communicable thought. We want to argue that, even on this interpretation, claim (5) is true: there are a great many *stable* and *effable* mental concepts that do not map onto words.

3   Inference and relevance

The alternative to a code theory of verbal communication is an inferential theory. The basic idea for this comes from the work of Paul Grice (1989); we have developed such a theory in detail in our book *Relevance: Communication and Cognition* (1986/1995). According to the inferential theory, all a communicator has to do in order to convey a thought is to give her audience appropriate evidence of her intention to convey it. More generally, a mental state may be revealed by a behaviour (or by the trace a behaviour leaves in the environment). Behaviour capable of revealing the content of a mental state may also succeed in *communicating* this content to an audience. For this to happen, it must be used ostensively: that is, it

must be displayed so as to make manifest an intention to inform the audience of this content.

Peter asks Mary if she wants to go to the cinema. Mary half-closes her eyes and mimes a yawn. This is a piece of ostensive behaviour. Peter recognises it as such and infers, non-demonstratively, that Mary is tired, that she wants to rest, and that she therefore does not want to go to the cinema. Mary has communicated a refusal to go to the cinema, and a reason for this refusal, by giving Peter some evidence of her thoughts. The evidence was her mimed yawning, which she could expect to activate in Peter's mind the idea of her being tired. The ostensive nature of her behaviour could be expected to suggest to Peter that she *intended* to activate this idea in his mind. Mary thought that the idea activated, and the manifestly intentional nature of its activation, would act as the starting point for an inferential process that would lead to the discovery of her meaning. She might have achieved roughly the same effect by saying 'I'm tired'. This would also have automatically activated the idea of her being tired (this time by linguistic decoding). It would have done so in a manifestly intentional way, thus providing Peter with strong evidence of Mary's full meaning.

In general, inferential communication involves a communicator ostensively engaging in some behaviour (e.g. a piece of miming or the production of a coded signal) likely to activate in the addressee (via recognition or decoding) some specific conceptual structure or idea. The addressee takes this deliberately induced effect, together with contextual information, as the starting point for an inferential process which should lead to the discovery of the message (in the sense of proposition plus propositional attitude) that the communicator intended to convey.

The idea activated and the message inferred are normally very different. The idea is merely a trigger for discovery of the message. Often, the triggering idea is a fragment, or an incomplete schematic version, of the message to be communicated. The inferential process then consists in complementing or fleshing out the triggering idea.

It is possible, at least in principle, for the idea activated by the communicator's behaviour to consist of a proposition and a propositional attitude (i.e. a full thought) which is just the message she intended to convey. In this limiting case, the inferential process will simply amount to realising that this is all the communicator meant. The classical code theory treats this limiting case as the only one. Every act of communication is seen as involving the production of a coded signal (e.g. a token of a sentence) which encodes exactly the intended message. No inferential process is needed. The sentence meaning (or, more generally, the signal meaning) is supposed to be identical to the speaker's meaning. The updated code theory treats this limiting case as the basic and paradigmatic one. Hearers should assume by

default that the sentence meaning is the speaker's message, but be prepared to revise this assumption on the basis of linguistic evidence (the sentence does not encode a full message) or contextual evidence (the speaker could not plausibly have meant what the sentence means).

Since the classical code theory is patently wrong, the updated code theory might seem more attractive. However, the classical theory had the advantage of offering a simple, powerful and self-contained account of how communication is possible at all. The updated theory loses this advantage by invoking an inferential mechanism to explain how more can be communicated than is actually encoded. The updated theory offers two distinct mechanisms – coding–decoding and inference – which may be singly or jointly invoked to explain how a given message is communicated. Why should the first of these be fundamental and necessary to human linguistic communication, while the second is peripheral and dispensable? The classical theory, which treats coding–decoding as the *only* explanation of communication, entails as a core theoretical claim that every communicable message is fully encodable. In the updated theory, this is a contingent empirical claim, with little empirical support and no explanatory purchase.

What is the role of inference in communication? Is it merely to provide short-cuts along the normal paths of coding–decoding (in which case any inferentially communicated message could have been fully encoded)? Or does inference open up new paths, to otherwise inaccessible end-points, making it possible to communicate meanings that were not linguistically encodable? (By 'not linguistically encodable' we mean not encodable in the public language actually being used, rather than not encodable in any possible language.) In the absence of any plausible account of the inferential processes involved in comprehension, the reasonable, conservative option might be to assume that inference does not enrich the repertoire of communicable meanings. For example, if all we had to go on was Grice's ground-breaking but very sketchy original account (in his 1967 William James lectures, reprinted in Grice, 1989), we would have very little idea of how inferential comprehension processes actually work, how powerful they are, and whether and how they might extend the range of communicable concepts.

Relevance theory (Sperber and Wilson, 1986/1995; see also references therein) offers a more explicit account of comprehension processes, which claims that what can be communicated goes well beyond what can be encoded. Here, we will give a brief, intuitive outline of relevant aspects of the theory.

The basic ideas of the theory are contained in a definition of relevance and two principles. Relevance is defined as a property of inputs to cognitive processes. The processing of an input (e.g. an utterance) may yield some

cognitive effects (e.g. revisions of beliefs). Everything else being equal, the greater the effects, the greater the relevance of the input. The processing of the input (and the derivation of these effects) involves some cognitive effort. Everything else being equal, the greater the effort, the lower the relevance. On the basis of this definition, two principles are proposed:

> (9) *Cognitive principle of relevance*: Human cognition tends to be geared to the maximisation of relevance.
> (10) *Communicative principle of relevance*: Every act of ostensive communication communicates a presumption of its own relevance.

More specifically, we claim that the speaker, by the very act of addressing someone, communicates that her utterance is the most relevant one compatible with her abilities and preferences, and is at least relevant enough to be worth his processing effort.

As noted above, ostensive behaviour automatically activates in the addressee some conceptual structure or idea: for example, the automatic decoding of an utterance leads to the construction of a logical form. This initial step in the comprehension process involves some cognitive effort. According to the communicative principle of relevance, the effort required gives some indication of the effect to expect. The effect should be enough to justify the effort (or at least enough for it to have seemed to the speaker that it would seem to the hearer to justify the effort – but we will ignore this qualification, which plays a role only when the speaker deliberately or accidentally fails to provide the hearer with sufficiently relevant information; see Sperber, 1994).

4   Relevance and meaning

The communicative principle of relevance provides the motivation for the following comprehension procedure, which we claim is automatically applied to the on-line processing of attended verbal inputs. The hearer takes the conceptual structure constructed by linguistic decoding; following a path of least effort, he enriches this at the explicit level and complements it at the implicit level, until the resulting interpretation meets his expectations of relevance; at which point, he stops.

We will illustrate this procedure by considering the interpretation of Mary's utterance in (11):

> (11) *Peter* Do you want to go to the cinema?
> *Mary* I'm tired.

Let's assume (though we will soon qualify this) that Peter decodes Mary's utterance as asserting that Mary is tired. By itself, the information that

Mary is tired does not answer Peter's question. However, he is justified in trying to use it to draw inferences that would answer his question and thus satisfy his expectations of relevance. If the first assumption to occur to him is that Mary's being tired is a good enough reason for her not to want to go to the cinema, he will assume she meant him to use this assumption as an implicit premise and derive the implicit conclusion that she doesn't want to go to the cinema because she is tired. Peter's interpretation of Mary's utterance contains the following assumptions:

(12) (a) Mary is tired.
     (b) Mary's being tired is a sufficient reason for her not to want to go to the cinema.
     (c) Mary doesn't to want to go to the cinema because she is tired.

Mary could have answered Peter's question directly by telling him she didn't want to go to the cinema. Notice, though, that the extra (inferential) effort required by her indirect reply is offset by extra effect: she conveys not just a refusal to go, but a reason for this refusal. There may, of course, be many other conclusions that Peter could derive from her utterance, for example those in (13):

(13) (a) Mary had a busy day.
     (b) Mary wouldn't want to do a series of press-ups.

But even if these conclusions were highly relevant to Peter, they would not help to satisfy the specific expectations of relevance created by Mary's utterance. The fact that she was replying to his question made it reasonable for him to expect the kind and degree of relevance that he himself had suggested he was looking for by asking this question, and no more.

   However, there is a problem. How plausible is it that the fact that Mary is tired is a good enough reason for her not to want to go to the cinema? Why should Peter accept this as an implicit premise of her utterance? Does Mary never want to go to the cinema when she is tired, even if she is just a little tired, tired enough for it not to be false to say that she is strictly speaking tired? Surely, in these or other circumstances, Peter might have been aware that Mary was somewhat tired, without treating it as evidence that she didn't want to go to the cinema.

   As noted above, a hearer using the relevance–theoretic comprehension procedure should follow a path of least effort, enriching and complementing the decoded conceptual structure until the resulting interpretation meets his expectations of relevance. We have shown how this procedure would apply to Mary's utterance in (11) to yield the implicatures (12b) and (12c). This is a case where the explicit content is complemented at the implicit level. However, for this complementation to make sense, some

enrichment must also take place at the level of what is explicitly communicated.

If comprehension is to be treated as a properly inferential process, the inferences must be sound (in a sense that applies to non-demonstrative inference). From the mere fact that Mary is tired, Peter cannot soundly infer that she doesn't want to go to the cinema. For the implicatures (12b) and (12c) to be soundly derived, Mary must be understood as saying something stronger than that she is tired *tout court*: her meaning must be enriched to the point where it warrants the intended inferences. The process is one of parallel adjustment: expectations of relevance warrant the derivation of specific implicatures, for which the explicit content must be adequately enriched.

Mary is therefore conveying something more than simply the proposition that she is tired, which would be satisfied by whatever is the minimal degree of tiredness: she is conveying that she is tired enough not to want to go to the cinema. If she were 'technically' tired, but not tired enough for it to matter, her utterance would be misleading, not just by suggesting a wrong reason for her not wanting to go to the cinema, but also by giving a wrong indication of her degree of tiredness. Suppose Peter thought that she was being disingenuous in using her tiredness as an excuse for not going to the cinema. He might answer:

> (14) Come on, you're not *that* tired!

He would not be denying that she is tired: merely that she is tired to the degree conveyed by her utterance.

How tired is that? Well, there is no absolute scale of tiredness (and if there were, no specific value would be indicated here). Mary is communicating that she is tired enough for it to be reasonable for her not to want to go to the cinema on that occasion. This is an *ad hoc*, circumstantial notion of tiredness. It is the degree of tiredness that has this consequence.

In saying (11), Mary thus communicates a notion more specific than the one encoded by the English word 'tired'. This notion is not lexicalised in English. It may be that Mary will never find another use for it, in which case it will not have the kind of stability in her mental life that we took to be a condition for mental concepthood. Alternatively, she may recognise this particular sort of tiredness, and have a permanent mental 'entry' or 'file' for it, in which case it is a proper concept. In the same way, Peter's grasp of the notion of tiredness Mary is invoking may be ephemeral, or he may recognise it as something that applies to Mary, and perhaps others, on different occasions, in which case he has the concept too.

It might be argued that the word 'tired' in Mary's utterance, when properly enriched, just means *too tired to want to go to the cinema*. This is a

meaning which is perfectly encodable in English, even though it is not lex-icalised. Suppose this were so, and that Mary has a stable concept of this kind of tiredness: her utterance would still illustrate our point that there may be many non-lexicalised mental concepts. The fact that this concept is encodable by a complex phrase would be no reason to think Mary does not have it as an elementary concept, any more than the fact that 'bachelor' can be defined is any reason to think we have no elementary mental concept of *bachelor*.

In any case, it is unlikely that Mary's answer in (11) is really synonymous with her answer in (15):

> (15) *Peter* Do you want to go to the cinema?
>   *Mary* I'm too tired to want to go to the cinema.

Mary's answer in (11) has a degree of indeterminacy that is lost in (15). Quite apart from this, the apparent paraphrasability of her answer in (11) is linked to the fact that she is answering a yes–no question, which drasti-cally narrows down the range of potential implicatures and the enrichment needed to warrant them. Yet the enrichment mechanism is itself quite general, and applies in contexts where the range of implicatures is much vaguer, as we will show with two further examples.

Imagine that Peter and Mary, on holiday in Italy, are visiting a museum. Mary says:

> (16) I'm tired!

As before, if her utterance is to be relevant to Peter, she must mean more than just that she is strictly speaking tired. This time, though, the implica-tions that might make her utterance relevant are only loosely suggested. They might include:

> (17) (a) Mary's enjoyment of this visit is diminishing.
>   (b) Mary would like to cut short their visit to the museum.
>   (c) Mary is encouraging Peter to admit that he is also tired and wants to cut short the visit.
>   (d) Mary would like them to go back to their hotel after this visit to the museum, rather than visiting the Duomo, as they had planned.

If these and other such conclusions are implicatures of her utterance, they are only weak implicatures: implications that Peter is encouraged to derive and accept, but for which he has to take some of the responsibility himself (for the notion of 'weak implicature', see Sperber and Wilson, 1986/1995, chapter 4). Whatever implicatures he ends up treating as intended (or sug-gested) by Mary, he will have to adjust his understanding of her explicit

meaning so as to warrant their derivation. Mary will be understood as having conveyed that she is tired to such a degree or in such a way as to warrant the derivation of these implicatures. This overall interpretation is itself justified by the expectation of relevance created by Mary's utterance (i.e. by this particular instantiation of the communicative principle of relevance).

That evening, at a trattoria, Mary says to Peter:

(18) I love Italian food!

She does not, of course, mean that she loves all Italian food, nor does she merely mean that there is some Italian food she loves. So what does she mean? It is often suggested that in a case like this, the expression 'Italian food' denotes a prototype, here *prototypical Italian food*. This presupposes that there is a readily available and relatively context-independent prototype. In the situation described above, it so happens that Mary is a vegetarian. Moreover, her understanding of Italian food is largely based on what she finds in an 'Italian' vegetarian restaurant in her own country where she sometimes goes with Peter, which serves several dishes such as 'tofu pizza' that are definitely not Italian. Mary's utterance achieves relevance for Peter by implicating that she is enjoying her food, and sees it as belonging to a distinct category which the expression 'Italian food' suggests but does not describe.

Even if Mary's use of the term 'Italian food' were less idiosyncratic, it would not follow that all Peter has to do to understand it is recover a prototype. Much recent research has cast doubt on the view that word meanings can be analysed in terms of context-independent prototypes, and suggests instead that *ad hoc* meanings are constructed in context (see e.g. Barsalou, 1987; Franks and Braisby, 1990; Franks, 1995; Butler, 1995). We would add that this contextual construction is a by-product of the relevance-guided comprehension process. The explicit content of an utterance, and in particular the meaning of specific expressions, is adjusted so as to warrant the derivation of implicatures which themselves justify the expectations of relevance created by the utterance act. These occasional meanings may stabilise into concepts, for the speaker, the hearer, or both.

These examples are designed to show how a word which encodes a given concept can be used to convey (as a component of a speaker's meaning) another concept that neither it nor any other expression in the language actually encodes. There is nothing exceptional about such uses: almost any word can be used in this way. Quite generally, the occurrence of a word in an utterance provides a piece of evidence, a pointer to a concept involved in the speaker's meaning. It may so happen that the intended concept is the very one encoded by the word, which is therefore used in its strictly literal

sense. However, we would argue that this is no more than a possibility, not a preferred or default interpretation. Any interpretation, whether literal or not, results from mutual adjustment of the explicit and implicit content of the utterance. This adjustment process stabilises when the hypothesised implicit content is warranted by the hypothesised explicit content together with the context, and when the overall interpretation is warranted by (the particular instantiation of) the communicative principle of relevance.

This approach sheds some light on the phenomenon of polysemy illustrated by the example of 'open' above. A verb like 'open' acts as a pointer to indefinitely many notions or concepts. In some cases, the intended concept is jointly indicated by the verb and its direct object (as with the ordinary sense of 'open the washing machine'), so that the inferential route is short and obvious. There may be cases where such routinely reachable senses become lexicalised. In general, though, polysemy is the outcome of a pragmatic process whereby intended senses are inferred on the basis of encoded concepts and contextual information. These inferred senses may be ephemeral notions or stable concepts; they may be shared by few or many speakers, or by whole communities; the inference pattern may be a first-time affair or a routine pattern – and it may be a first-time affair for one interlocutor and a routine affair for another, who, despite these differences, manage to communicate successfully. (For relevance–theoretic accounts of polysemy, see Carston, 1996, in preparation; Deane, 1988; Groefsema, 1995; Papafragou, in preparation; Wilson and Sperber, in press.)

5   Implications

Our argument so far has been that, given the inferential nature of comprehension, the words in a language can be used to convey not only the concepts they encode, but also indefinitely many other related concepts to which they might point in a given context. We see this not as a mere theoretical possibility, but as a universal practice, suggesting that there are many times more concepts in our minds than words in our language.

Despite their different theoretical perspectives, many other researchers in philosophy, psychology and linguistics have converged on the idea that new senses are constantly being constructed in context (e.g. Franks and Braisby, 1990; Goshke and Koppelberg, 1992; Barsalou, 1987; Gibbs, 1994; Franks, 1995; Recanati, 1995; Nunberg, 1996; Carston, 1996, in preparation). However, it is possible to believe that new senses can be contextually constructed, without accepting that there are more mental concepts than public words.

Someone might argue, for example, that the only stable concepts are linguistically encodable ones. Unless a new sense constructed in context is lin-

guistically encodable, it cannot be a stable concept of the speaker's, and will never stabilise as a mental concept in the hearer. When Mary says at the museum that she is tired, the understanding that she and Peter have of her kind and degree of tiredness cannot be divorced from their understanding of the whole situation. They do not construct or use an *ad hoc* concept of tiredness. Rather, they have a global representation of the situation, which gives its particular contextual import to the ordinary concept of tiredness.

   We would reply as follows. We do not deny – indeed, we insist – that most occasional representations of a property (or an object, event or state) do not stabilise into a concept. Most contextually represented properties are not recognised as having been previously encountered, and are not remembered when the situation in which they were represented is itself forgotten. However, some properties are recognised and/or remembered even when many or all of the contextual elements of their initial identification are lost. For example, you look at your friend and recognise the symptoms of a mood for which you have no word, which you might be unable to describe exactly, and whose previous occurrences you only dimly remember; but you know that mood, and you know how it is likely to affect her and you. Similarly, you look at the landscape and the sky, and you recognise the weather, you know how it will feel, but you have no word for it. Or you feel a pain, you recognise it and know what to expect, but have no word for it; and so on. You are capable not just of recognising these phenomena but also of anticipating them, imagining them, regretting or rejoicing that they are not actual. You can communicate thoughts about them to interlocutors who are capable of recognising them, if not spontaneously, at least with the help of your communication. Your ability to recognise and think about the mood, the weather, the pain, is evidence that you have a corresponding stable mental file or entry, i.e. a mental concept. The evidence is not, of course, conclusive, and there could be a better hypothesis. However the suggestion that what has been contextually grasped can only be remembered with all the relevant particulars of the initial context is not that better hypothesis.

   There is a more general reason for believing that we have many more concepts than words. The stabilisation of a word in a language is a social and historical affair. It is a slow and relatively rare process, involving co-ordination among many individuals over time. A plausible guess is that, in most relatively homogenous speech communities in human history, less than a dozen new words (including homonyms of older words and excluding proper names) would stabilise in a year. On the other hand, the addition of new concepts to an individual's mind is comparatively unconstrained. It is not a matter of co-ordinating with others, but of internal memory management. There is no question that we are capable of acquiring a huge amount of new information every day. Do we store it all in pre-existing files, or do

we sometimes – perhaps a few times a day – open a new file, i.e. stabilise a new concept? Notice that this would not involve adding extra information to long-term memory but merely organising information that we are going to add anyhow in a different, and arguably more efficient way.

Information filed together tends to be accessed together, and efficient memory management involves not only filing together what is generally best accessed together, but also filing separately what is generally best accessed separately. Thus, you may be able to recognise a certain type of food (which the public linguistic expression 'Italian food' may hint at in an appropriate context but does not describe), and this ability may play a role in your mental life: say in deciding what to eat or cook on a given occasion. Where is information about this kind of food stored in your memory? Does it have its own address, or does it have to be reassembled from information filed elsewhere every time it is used?

How and how often we open new files, and thus stabilise new mental concepts, is an empirical question, to be investigated with the methods of psychology. However, the hypothesis that we can open a new file only when we have a public word that corresponds to it is a costly one, with no obvious merit. It amounts to imposing an arbitrary and counter-productive constraint on memory management. (This is not, of course, to deny the converse point that on encountering a new word you may stabilise a new concept, and that many of our concepts originate partly or wholly from linguistic communication – a point for which there is much evidence, in particular developmental, e.g. Gelman and Markman, 1986.)

While the kind of collective co-ordination needed to stabilise a word in a speech community is an elaborate affair, the typically pairwise co-ordination involved in any given communicative act is a relatively simpler achievement – the kind of achievement that a pragmatic theory such as relevance theory aims to explain. This co-ordination may be somewhat loose. When Mary says at the museum that she is tired, her utterance gets its explicit meaning through adjustment to a set of weak implicatures: that is, implicatures whose exact content is not wholly determined by the utterance. The *ad hoc* concept of tiredness that Peter constructs (i.e. the concept of tiredness which warrants the derivation of these weak implicatures) is unlikely to be exactly the same as the one Mary had in mind (since she did not foresee or intend exactly these implicatures). This is not a failure of communication. It is an illusion of the code theory that communication aims at duplication of meanings. Sometimes it does, but quite ordinarily a looser kind of understanding is intended and achieved. The type of co-ordination aimed at in most verbal exchanges is best compared to the co-ordination between people taking a stroll together rather than to that between people marching in step.

Returning to the question of effability, we would maintain that this is a matter of degree. Some concepts are properly shared, and can be unequivocally expressed: a mathematical discussion would provide good examples. Other concepts are idiosyncratic, but as a result of common experience or communication, are close enough to the idiosyncratic concepts of others to play a role in the co-ordination of behaviour. Still other concepts may be too idiosyncratic to be even loosely communicated. The fact that a public word exists, and is successfully used in communication, does not make it safe to assume that it encodes the same concept for all successful users; and in any case, the concept communicated will only occasionally be the same as the one encoded. Communication can succeed, despite possible semantic discrepancies, as long as the word used in a given situation points the hearer in the direction intended by the speaker. Thus, Peter and Mary might differ as to the exact extension of 'tired': Peter might regard as genuine though minimal tiredness a state that Mary would not regard as tiredness at all. Mary's successful use of the term in no way depends on their meaning exactly the same thing by it. Similarly, their concepts of Italy might pick out different entities in space or time (for example, is Ancient Roman History part of Italian History? That depends on what you mean by 'Italy'). Mary's successful use of the term 'Italian' should be unaffected by these discrepancies.

More generally, it does not much matter whether or not a word linguistically encodes a full-fledged concept, and, if so, whether it encodes the same concept for both speaker and hearer. Even if it does, comprehension is not guaranteed. Even if it does not, comprehension need not be impaired. Whether they encode concepts or pro-concepts, words are used as pointers to contextually intended senses. Utterances are merely pieces of evidence of the speaker's intention, and this has far-reaching implications, a few of which we have tried to outline here.