Corpus Analysis and Lexical Pragmatics: An Overview

AHRC Lexical Pragmatics project team

1. Introduction

In his book *Words and Phrases*, Michael Stubbs notes that using corpus data in the study of word meaning need not involve rejecting or downplaying the value of introspective evidence. "In the areas of semantics and pragmatics," he suggests, "intuitions are strong and stable (…) and must be given the status of data." (Stubbs 2001). We share the view that the use of corpus-based evidence should be seen as complementary to more standard methods of investigation: it doesn't substitute for intuitions but can help to develop and test hypotheses and reduce the possibility of intuitive data being mere artefacts of the linguist.

A corpus-based investigation of lexical pragmatics is in many respects a pilot project. There is no established paradigm,[1] and since the interpretation of utterances in discourse may depart significantly from the linguistically encoded meaning, the analyst's intuitions about the intended interpretation must play an important role. Moreover, the practicalities of a corpus-based study of lexis (quite apart from the technical and temporal limitations of our own project) do not allow firm statistical generalizations. As Sinclair (1991) points out, with grammatical words such as *the* or *not* occurring in a sizeable corpus hundreds of thousands of times and lexical words just a few dozen, statistical generalizations about lexical meaning are relatively hard to obtain. For instance, in the Bank of English (the 56 million word corpus we used in our research), among the words or phrases whose uses we wanted to analyse, *red eyes* occurs a mere 29 times, *bulldozer* only 61 times, *painless* 89 times, *boiling* 332 times, and so on. At one point, we were interested in the metaphorical use of *bulldozer* to mean 'forceful, bullying person', which is often described in the literature as a standardized metaphor, and were surprised to find

---

[1] For an interesting attempt to establish procedures for identifying metaphorical uses in a text, see the work of the Pragglejaz group.

that it only occurred once, in a reference to Jacques Chirac being nicknamed 'the bulldozer'.

However, the corpus proved a valuable source of inspiration in our research, helping to sharpen and test our hypotheses, and raising new and intriguing questions. In this overview, we will illustrate how we used it to shed light on three main theoretical hypotheses during our first six months of corpus work:[2]

(a) The first hypothesis was that lexical narrowing is a highly flexible, creative and context-sensitive process, which cannot be easily handled in terms of default rules.

(b) The second was that there is no sharp theoretical distinction between literal, loose and metaphorical uses, but a continuum of cases with no clear cut-off point between them, which are all understood in the same way.

(c) The third was that the study of lexical pragmatic processes should shed interesting light on traditional notions of literal meaning, polysemy.

Here, we will give a brief overview of the type of results we obtained, outlining the theoretical motivations behind some of our searches and the main conclusions we draw from them. (For further analysis and discussion, see the Corpus Analysis section of the AHRC Lexical Pragmatics website hosted by the Department of Phonetics & Linguistics, UCL.)

2. Lexical narrowing

Lexical narrowing involves the use of a word or phrase to convey a more specific concept (with a narrower denotation) than the linguistically encoded 'literal' meaning. To illustrate, consider (1) and (2):

(1)   Mary is a *working mother.*
(2)   Bill *has money.*

---

[2] A version of this paper was presented by Patricia Kolaiti at the AHRC Workshop on Word Meaning, Concepts and Communication (Cumberland Lodge, September 2005).

In many circumstances, the speaker of (1) would be taken to mean that Mary is a prototypical working mother, who has young children living with her, and who works for money outside the home, while the speaker of (2) would be understood as conveying that Bill has a significant amount of money.

One standard view of lexical narrowing is that it is analysable in terms of default inference rules. For instance, Levinson (2000:37-8, 112-34) treats narrowing as involving a default inference governed by an Informativeness heuristic ("What is expressed simply is stereotypically exemplified"), itself backed by a more general I-principle instructing the hearer to

"Amplify the informational content of the speaker's utterance, by finding the most *specific* interpretation, up to what you judge to be the speaker's m-intended point …" [ibid: 114]

The I-heuristic might be seen as dealing with stereotypical lexical narrowings such as (1) below, and the I-Principle as dealing with less stereotypical cases such as (2). On this approach, hearers are seen as automatically constructing a prototypical (or otherwise enriched) interpretation and accepting it in the absence of contextual counter-indications. The alternative relevance-theoretic approach which we favour treats narrowing as a far more creative and flexible process, involving the construction of ad hoc, occasion-specific concepts influenced by a much wider range of cognitive and contextual factors than default approaches take into account (see e.g. Sperber & Wilson, 1998; Carston, 2002; Wilson & Sperber, 2002; Wilson, 2003; Wilson & Carston, forthcoming). Thus, the concept of a WORKING MOTHER, or of HAVING MONEY, might be narrowed to different degrees, and in different directions, in different context, yielding a range of occasion-specific ('ad hoc') concepts WORKING MOTHER*, WORKING MOTHER**, WORKING MOTHER***, and so on.

As a starting point for examining these hypotheses, we took a standard problem in lexical pragmatics that does not obviously favour either approach: the fact that the adjective *red* is typically narrowed in different directions in common adjective-noun combinations such as *red eyes, red apple, red hair, red stamp,* etc. (picking out a

different shade, distributed in different ways across the surface of the object, in different combinations). A 'default rule' approach might handle this by assigning *red* a different default interpretation for each common adjective-noun combination, and predict that this will be automatically preferred in the absence of contextual counter-indications. Our intuition was that, although there is probably a range of fairly standard narrowings of *red* in the context of *eyes*, *hair, apple, stamp,* etc., the interpretations would still be diverse and creative enough to raise questions about the default approach. We will illustrate using the common adjective-noun combination *red eyes*.[3]

In fact, we found considerable evidence of the creativity and flexibility of narrowing even in this common combination. In each case, the adjective *red* was plausibly understood as communicating a slightly narrower concept (e.g. RED*, RED**) appropriate to the wider discourse context, picking out a particular shade other than focal red, differently distributed over the surface of the eyes. Here are some illustrations:

(3) …*red eyes* denote strain and fatigue.

Here *red* picks out a reddish-pink shade ranging around the edges of the eye, on the bags under the eye and perhaps on part of the cornea too. The exact shade and distribution the speaker is taken to convey would vary depending on further contextual clues about the colour of the skin and the degree of strain or fatigue involved.

(4) *In a photography session*: [This flashing light is] to stop you getting *red eyes*

Here *red* picks out a luminous, rusty red on the iris only.

(5) *In a conversation about demons: …*two burning *red eyes* she recalled…

---

[3] By expanding the search to include not only 'red + eyes' but also 'eyes + red' and 'red +a number of intervening items+ eyes' etc. we managed to increase the number of occurrences to 55.

Here *red* picks out a fiery and luminous red, distributed over both the cornea and the iris or the iris alone.

There are also metaphorical uses, as in (6):

(6) …eyes *red* with resentment…

Out of a total of 54 occurrences of *red eyes* and its variants (e.g. 'eyes+red' and 'red+intervening items+ red') in the corpus, our search identified 26 different such 'discourse contexts'. The results are summarised below, along with an indication of the frequency of occurrence of the combination *red eyes* in each such context:

Context	Lines	Overall number of occurrences
Context of crying	4, 6, 18, 35, 38, 41, 42, 43, 44, 50, 57, 58, 59, 64	14
Context of hardship and/or stress and/or fatigue	21, 40, 45, 53, 55, 60	6
Context of eyes gleaming in the dark	29, 30, 31	3
Context of flu/cold	51, 65	2
Context of a gorilla mask	1	1
Context of eye damage	2	1
Context of midgets	3	1
Context of dizziness	7	1
Context of sheep-like eyes	8	1
Context of Albinos	17	1
Context of red-eye effect	19, 20	2
Context of rage*	23, 54, 68	3
Context of resentment*	39	1

Context of demons	26	1
Context of drunkenness	27	1
Context of Caymans	28	1
Context of koels /cuckoos	34	1
Context of eczema	36	1
Context of heat and sand	54	1
Context of sore eyes	70	1
Context of colour of one's eyes	11	1
inconclusive cases: (context of fiction)		
terrestrials with long ears	10	1
fictional insects	22	1
unknown entity	24, 25, 47	3
'giants' and 'heroes'	32, 33	2
unknown context	69, 9	2

Notice that 17 of the 26 discourse contexts occur only once. The significant proportion of one-off uses suggests a level of creativity that poses problems for the default account and argues for a more flexible, context-dependent approach.

These results provide some evidence for the view that a hearer interpreting the phrase *red eyes* on different occasions draws on a wide range of background or contextual assumptions in constructing an overall interpretation. Relevant contextual factors include the type of entity bearing the eyes (e.g. humans, animals, insects, demons or 'a terrifying [gorilla] mask with little red eyes that blinked') and the cause of redness: (e.g. eczema, drunkenness, crying, flu/cold, fatigue, exposure to heat, sand, light, etc.). It seems clear that the direction of narrowing varies so widely from one discourse context to another that it would be hard to come up with a single default rule that would provide a better starting point for constructing the full range of interpretations than the linguistically encoded 'literal' meaning.

The results also raise a number of questions for the default approach. For instance, should the same default rule be seen as applying in every case (e.g. would the same default rule assign an interpretation to *red* in a discourse context about a cuckoo whose eyes are gleaming in the dark and another about prolonged fatigue in humans)? To account for the flexibility in interpretation revealed by our search, there would either have to be a very large number of default rules (raising the question of how hearers choose among competing 'default' interpretations) or else the 'default' interpretation would have to be seen as overridden by contextual factors in a very wide range of cases. A simpler alternative might be to acknowledge that narrowing is directly affected by background knowledge and pragmatic principles, without passing through an initial 'default interpretation' stage.

A further question for the default approach is about how it handles cases in which the interpretation remains vague or open. In the absence of adequate contextual clues, for instance, a hearer may narrow the interpretation only to some extent (e.g. 'red in a way that would be appropriate to the eyes of an imaginary insect') or leave the interpretation open and not make the effort to narrow at all. According to the relevance-theoretic account, narrowing should not occur automatically, but is triggered by pragmatic factors (in particular, the goal of finding an interpretation that satisfies expectations of relevance). This account predicts that hearers will only narrow to a point where the utterance becomes relevant enough (i.e. to a point where the utterance yields enough cognitive effects, for a low enough processing cost, to satisfy expectations of relevance). In the absence of such triggering factors, it is predicted that narrowing will not take place.

In our search, we encountered 9 inconclusive cases in which the entities described as having 'red eyes' were either not specified in the immediate linguistic context[4] or were invented or non-existent living kinds ('fictional insects', 'terrestrials with long ears', etc). Why assume that hearers construct a concrete mental representation of the shade and distribution of redness over the eyes of a 'terrestrial with long ears' at all? It is a genuine problem for the default approach to explain what happens to the automatic application of

---

[4] We restricted the discourse context to a default of 6 lines before and after the search term. If the default context did not provide enough clues, we expanded the search to a further 10 lines before and after the search term and, if the context was still insufficient, we marked the case as open/ inconclusive.

the default rule in such cases. (See below for a more detailed look at some further cases of narrowing, involving the adjective *empty*.)

3. The continuum of literal, loose and metaphorical interpretations

Lexical broadening involves the use of a word or phrase to convey a more general concept (with a broader denotation) than the linguistically encoded 'literal' meaning. A striking feature of much research in this area is that different interpretive procedures have been proposed for a range of phenomena which could all be seen as varieties of broadening. Thus, approximation is often treated as a case of pragmatic vagueness involving different contextually-determined standards of precision (Lewis 1979; Lasersohn 1999). Metaphor and hyperbole are still widely seen as involving blatant violation of a pragmatic maxim of truthfulness, with the use of metaphor implicating a related simile or comparison and the use of hyperbole implicating a related weaker proposition (Grice 1975). Typically, these accounts do not generalise: metaphors are not analysable as rough approximations, approximations are not analysable as blatant violations of a maxim of truthfulness, and so on. Relevance theorists, by contrast, have been exploring the hypothesis that there is no clear cut-off point between literal use, approximation, hyperbole and metaphor, but merely a continuum of cases of broadening, which are all understood in the same way (Carston, 1997, 2002; Wilson & Sperber, 2002; Wilson & Carston, 2006; Sperber & Wilson, forthcoming; Wilson & Carston, forthcoming).

To illustrate, consider the utterance in (7):

(7)   The sea is *boiling*.

This might be intended and understood literally (as indicating that the sea is at or above boiling point), as an approximation (indicating that the sea is close to boiling point), a hyperbole (indicating that the sea is hotter than expected or desired) or a metaphor (indicating that the sea, while not necessarily hot, is bubbling, churning, emitting vapour, etc.). The issue is whether these are theoretically distinct interpretations involving different interpretive procedures, or whether they merely occupy different points on a

continuum, and are all understood in the same way, by broadening the linguistically-specified meaning in order to satisfy expectations of relevance.

To provide some evidence which might help to choose between these approaches, we focused on the adjectives *boiling*, *raw*, and *painless,* all of which are strictly defined but often loosely or metaphorically used. The results showed that loosening is not rare in language use. In fact, in the cases of *boiling* and *painless*, loose uses predominate:

Relative Frequency of Literal and Loose uses of *boiling*:

Type of use	Frequency in % terms
Literal	49.4%
Literal or Approximate	14.2%
Loose (i.e. non-literal)	36.1%

Relative Frequency of Literal and Loose uses of *painless*

Type of use	Frequency in % terms
Literal	20.2%
Literal or Approximate	15.7%
Loose (non-literal)	64.0%

The results for *painless* illuminate the relation between literal use and approximation in unexpected ways. Consider (8):

(8) *In a discussion of euthanasia:* I would want something clean and painless: no botch-ups. It would be the doctor or no one.

Here, the denotation of *painless* is plausibly understood as including not only cases in which the procedure was strictly and literally painless, but also those involving a small amount of physical pain, which would still be insignificant compared to the distress the patient would have to go through if allowed to die naturally. In other words, the linguistically encoded concept PAINLESS is broadened to PAINLESS*, whose extension

includes, but goes beyond, instances that are strictly and literally PAINLESS. Around 16% of all uses of *painless* fell into this category, with strictly literal uses making up around 20%.

From a theoretical point of view, 'approximations' are sometimes seen as excluding the possibility of a literal interpretation (as, for instance, describing an object as *squarish* would generally be understood as excluding the possibility that it is strictly and literally square). However, the frequency of cases such as (8), which are indeterminate between literal and 'approximate' interpretations, suggests that many loose or approximate uses of words involve a type of broadening from which the denotation of the linguistically encoded concept is not automatically excluded.[5]

Our findings for *boiling* show in more detail the form that the continuum of literal and loose uses of the same word might take. In a total of 332 occurrences, we found 164 cases which could only be understood literally (to mean 'at or above boiling point'), as in (9):

(9) Poached eggs come out well in a small dish using boiling water.

There were a further 47 cases in which either a literal or an approximate interpretation would be appropriate, as in (10):

(10) Cover the cake with the icing, smoothing with a knife dipped in boiling water.

By contrast, there were only 4 cases where an approximate interpretation would be appropriate and a literal interpretation would not, as in (11):

(11) For sauce, melt chocolate (…) over boiling water, then beat until smooth

(Those of you who have tried to melt chocolate in a bain-marie might already know that if the water in the bain-marie is literally boiling, chocolate will not melt but crumble.)

Towards the figurative end of the continuum, we found 80 cases where *boiling* was metaphorically used, as in (12):

---

[5] On this approach, cases where approximation is understood as excluding the possibility of a literal interpretation would result from a combination of narrowing and broadening – again providing evidence for a unified approach.

(12) …several small boats disappeared in boiling seas.

There were 4 clear cases of hyperbole, as in (13), and 13 cases in which metaphor and hyperbole were combined, as in (14):

(13) Bring some more ice, this whisky is boiling hot
(14) This summer is promising to be long and boiling

Finally, there were 19 cases that would be traditionally classified as synecdoche, as in (15) (we will not consider the theoretical analysis of synecdoche here):

(15) You're changing small things like boiling a kettle…

Note also that the metaphorical uses of 'boiling' were quite varied. More specifically we found metaphors indicating:

anger and emotional frustration, as in (16):

(16) The brothers, seemingly stable, are absolutely boiling inside with various frustrations…

excessive heat, as in (17):

(17) This summer is promising to be long and boiling…

tension, as in (18):

(18) Cup final, against anyone Pakistan relations almost at boiling point…

and finally, movement or appearance (typically of water or clouds), as in (19):

(19) …several small boats disappeared in boiling seas…

These results provide some evidence for our view that there is a continuum of cases of broadening, and that the degree and direction of broadening are heavily context-dependent.

3. Questions about literal meaning

In all three searches ('raw', 'painless' and 'boiling'), we were forcibly reminded of the elusiveness of 'literal' meaning. We embarked on each search with what felt to us like fairly strong intuitions about the literal meaning of the given word-set, but before long, these intuitions started to waver under the weight of the extremely diverse, thoroughly context sensitive and remarkably creative facts of language use. We will illustrate this point by considering the case of 'raw'.

We began our search with the clear intuition that the literal meaning of 'raw' is NOT COOKED. Faced with the diversity of the **308** concordances that we looked at, we had no option but to question this intuition. Utterances of 'raw' meaning NOT COOKED make up only 2% of all the examples we examined, with a striking 98% looking more like this:

Raw power, raw immediacy, raw skin, raw edges, something raw and honest, raw wood, raw adrenalin, raw noise, raw and wired experience, raw deal, raw humour, raw appeal, raw emotion, raw nerve, raw data, something raw and pure, raw big band brilliance, raw recruits, raw art, raw passions and… a sense of raw being.

The fact that the collocation 'raw materials' makes up a good proportion of all such metaphorically loosened examples made us consider the possibility that 'raw' has taken on a broader literal sense, meaning NOT PROCESSED. In (20a-c), for example, the use of 'raw materials' feels rather literal, although 'raw' does not here mean NOT COOKED:

(20a) …the swallowing up of exploitable territory, populations, raw materials and markets by commercial capital…

(20b). …explain how perfumes are constructed, show you the raw materials and invite you to experience the constituents…

(20c)   …duty free import of machinery and raw materials…

By contrast, in (21a-b) the used of 'raw materials' feels rather metaphorical:

(21a)   [it's] up to the couple to build a solid foundation out of the raw materials provided

(…) Respect, mutual value, domestic sharing…

(21b)   …all the raw material for this crisis has been around since 1988.

We decided to see if historical investigations might help. According to the OED, the etymological root of 'raw' comes from the Gr. κρέας and Skr. kravíş, meaning RAW FLESH. However, in Romance languages like French and Spanish, as well as in Greek, a large number of the metaphorical uses found in English would not be acceptable. This led us think that 'raw' in English might indeed have taken on a broader literal sense, meaning NOT PROCESSED. On this approach, the proportion 2% Literal use to 98% Loose use in our corpus sample changed into roughly 32% Literal use to 65% Loose use.

One of the most interesting aspects of this search was the way in which the contrast between intuitions and corpus evidence brought to the surface intriguing questions about the semantic representation of literal meaning. Has the formerly ad hoc concept NOT PROCESSED now become part of the permanent semantic representation of the lexical item 'raw'? If 'raw' has taken on this broader literal sense, why did our initial intuitions lead us to declare with relative conviction that 'raw' means NOT COOKED? And how should one account for the instability of these intuitions across individuals (some had stronger intuitions than others that 'raw' means NOT COOKED) and across times (we observed significant changes in intuitions about literal meaning within individuals across times or faced with different examples)?

An important advantage of the relevance-theoretic approach to word meaning used in this project is that it explains how communication can be successful even among speakers whose representations of 'literal' meaning are not homogeneous, and indeed vary considerably (see Sperber & Wilson 1998, 2006 for discussion). Suppose that for some speakers 'raw' means NOT COOKED. Given the context-dependence of lexical pragmatic processes, they should have no difficulty broadening it to mean NOT PROCESSED in appropriate circumstances. Suppose that for other speakers, 'raw' means

NOT PROCESSED. Again, given the context-dependence of lexical-pragmatic processes, they should have no difficulty narrowing it to mean NOT COOKED in appropriate circumstances. By the same token, 'painless' may have the narrower encoded meaning WITHOUT (PHYSICAL) PAIN for some speakers, and the broader encoded meaning WITHOUT (PHYSICAL OR MENTAL) PAIN for others. On the relevance-theoretic approach, such variations are to be expected, and should pose no threat to communication as long as speakers can converge on the same sense on any given occasion of use.

So far, all our searches have shown considerable context-sensitivity in the way lexical items were understood. All have confirmed our view that lexical narrowing and loosening are not incidental occurrences to be abstracted away from, but are fundamental to language use.

In the brief synopsis to follow, we will use our recent analyses of the adjective 'empty' to show how narrowing and broadening may combine in the interpretation of a single word.

4. A corpus based investigation of narrowing and broadening: 'empty'

Let's start with the plausible assumption that 'empty' means EMPTY (denoting the items that contain literally nothing). The adjective 'empty' occurs in all the subcorpora of the Bank of English in a total of 2336 concordances. To make the search manageable, we decided to focus only on the **89** relevant examples of the subcorpus Ukephem (which consists of ephemera - leaflets, adverts, etc) as we were more interested at thhat stage in colloquial/spoken language oriented samples.

Our hypotheses were, first, that 'literal' word meanings typically undergo narrowing or broadening in the course of comprehension, and second, that these departures from 'literal' meaning may take place in different directions and to different degrees. Thus, we expected to find the encoded concept EMPTY was consistently adjusted to denote a more fine-tuned type and degree of emptiness (EMPTY*, EMPTY**, etc.). The sorts of variations we expected to find were (a) variations in the entity or substance of which the item was understood as empty (e.g. EMPTY OF WINE, EMPTY OF WATER, etc.) and (b) variation in the degree to which the pertinent entity or substance is understood as lacking. Our objective was to illustrate the great diversity of use within a particular case

that would make it impossible to interpret by default, and show that narrowing and broadening are far more flexible than is normally thought.

Findings

_____

Word meanings get narrowed in different directions and to different extents

Our investigation of 'empty' illustrates all three points discussed above. In all utterances under investigation the word meaning gets narrowed in different directions and to different extents. In each case, the encoded concept is adjusted to indicate a more fine-tuned kind of emptiness. Compare, for instance the following utterances:

(23)   Later in the year, when the granaries are empty, families have to return to the market to buy grain.

(24)   But whatever you do, don't play sport on an empty stomach or after a heavy meal.

Neither (23) nor (24) involves a strict use of EMPTY. If nothing else, the empty granaries must at least contain air and the empty stomach gastric fluids. Hence, we must assume that in the first case the encoded concept is adjusted to EMPTY OF GRAIN and in the second to EMPTY OF RECENTLY RECEIVED FOOD. Such fine-tunings occur repeatedly throughout our search. In each case the audience must bring to bear different contextual assumptions in order to understand the speaker's meaning in a pragmatically plausible way.

Variation across discourse contexts

It follows from the last section that the concept communicated by 'empty' varies significantly across contexts. The sample of discourse contexts available in the Bank of English for 'empty' is very diverse. Unlike with 'red eyes', where certain contexts (e.g. fatigue, crying, etc) tended to recur rather frequently, in the case of 'empty' variation in context is much greater. With the exception of just a few recurring contexts (empty stomach, property empty of tenants, bus empty of passengers and a few others) all discourse contexts are one-off occurrences. The relatively high proportion of one-off uses

may favours a pragmatically based approach to lexical narrowing, such as the one we propose.

Variation within discourse contexts

Our investigation of the word set 'red eyes' revealed a potential problem for our hypothesis about the creativity and context-dependence of word use: there was a noticeable relative constancy in the direction of narrowing within a given type of discourse context. In that search, we observed:

'Whilst across contexts the extent and direction of narrowing regularly varies, within a given context narrowing seems to maintain rough relative constancy. In all 14 cases of crying and all 6 cases of fatigue, for example, the mentally represented shade of red (and the part of the eye involved in the loosening of the noun 'eyes') is roughly the same. The given sample does not involve even a single case in which such internal variation could be observed '.

Our intuition at that point was that the lack of variation relates to the sample available in the Bank of English for 'red eyes', rather than to some general fact about the behaviour of narrowing or the behaviour of discourse contexts themselves.

In our search on 'empty' however, we managed to find at least one case in which significant variation occurs in the direction of narrowing within the same broader discourse context (namely, a discussion of empty property). Compare (24) and (25):

(24)   …opening up to homeless people, the thousands of empty properties we know they have on their books.

(25)   She was eventually housed, but in a completely empty flat.

Although in (24) 'empty property' is understood as EMPTY OF TENANTS, in (25) the context clearly calls for the interpretation EMPTY OF FURNITURE. We take this as evidence that variation in the direction of narrowing occurs not only across but also within types of discourse context.

Constraining contexts? How empty is empty?

An interesting point about 'empty' is that context and background knowledge place constraints on the possible positions of an interpretation on the continuum between literal and loose use. From a logical point of view, we might say that first the encoded concept has to be narrowed in the contextually relevant way and only then can it be positioned at a certain point on the continuum in line with what the context permits. Thus, consider (26):

(26)   remember to take with you any empty tablet bottles or containers to show what has been…

Here, the discourse contextd in theory permits either a literal or an approximate interpretation. So we might imagine that 'empty' is first narrowed to mean EMPTY OF TABLETS. Then, if the interpretation is appropriate only if no tablets have been left in the bottle, 'empty' is understood in its strict sense, while if a tablet or two might have been left in the bottle, 'empty' is understood as an approximation.

But now consider the following examples:

(27)   … opening up to homeless people the thousands of empty properties we know they have on their books.

(28)   [clues] that your child has been sniffing include: finding empty butane, aerosol or glue cans...

(29)   …the three of us are sharing a room with pizza remnants, empty wine bottles and flagging concentration…

Again, it might seem plausible to assume that 'empty' is first narrowed to mean, say, EMPTY OF WINE, EMPTY OF AEROSOL, etc. However, by virtue of how the world is, it is hard to imagine a wine bottle or an aerosol can being completely and utterly empty of wine or aerosol. Unless some emptying process to laboratory standards takes place, both the wine bottle and the aerosol can will still have traces of the substance they contained,

which rules out a strictly literal interpretation. Only approximate interpretations seem plausible.

Example (27) raises exactly the opposite problem. For a property to be EMPTY OF TENANTS and the utterance to be true (or true enough), it is imperative that not even one tenant remains. If at least one tenant remains, then the utterance is not true. This type of context does not leave any room for lexical broadening, and consequently, approximate uses are not plausible.

These examples show that the extent to which a certain entity/substance must be lacking for something to be appropriately described as empty of it is itself heavily context-dependent. In example (26), if one or a few tablets have been left in the bottle, the description "empty tablet bottles" is true enough (or, as we would put it, relevant enough). However, in (27), if at least one tenant remains in each property, then the description "empty properties" is not true (or relevant) enough. We believe that these cases provide support for the relevance-theoretic claim that comprehension is not a strictly sequential matter (*first* you establish the explicit content, *then* you derive implicatures), but involves a process of mutually adjusting explicit content with implicatures in order to satisfy expectations of relevance.

Along these lines, our 89-line sample of the adjective 'empty' involved:

(a) 26 lines (29.2 %) in which 'empty' could be interpreted either literally or approximately,

(b) 38 lines (42.7 %) in which only a literal interpretation would be plausible,

(c) 14 lines (15.7 %) in which only an Approximate interpretation would be plausible and

(d) 11 lines (12.4 %) where 'empty' is interpreted Metaphorically.

Metaphorical uses of 'empty'

As with all other terms that we looked at in the corpus, the metaphorical uses we found in the case of 'empty' were quite varied:

(30)   a. Sugar gives you empty calories

        b. Shelter is hard to find and empty days are spent wandering the streets

c. [When] a smoker is deprived of a cigarette he or she will feel empty and restless at first

d. Otto Ritter, a German archaeologist working in the Empty Quarter

e. …at a price to stock your wardrobe and not empty your pocket

f. …photo The Empty Raincoat CHARLES HANDY

g. Sit back, empty your head of foolish thoughts and just close your eyes

h. Try to empty your mind of anxious or guilty thoughts

i. Abasio turned aside to empty his stomach, noisily and messily

j. Law, has become a target of efforts to empty it of intrinsic meaning

k. life without letters from you would be much colder and emptier than it is

The meaning of 'empty': strict uses of an under-specified term?

The data briefly presented above raise several theoretical questions about the encoded meaning of 'empty'. One possible approach would be to treat the word 'empty' as meaning EMPTY OF ANYTHING AT ALL, and thus as encoding an abstract term that cannot be strictly satisfied in nature. It would then follow that on every normal occasion of use some adjustment of the encoded concept takes place.

Another possible approach would be to treat 'empty' as an under-specified term, encoding something along the lines of EMPTY OF X, with X pragmatically supplied. In that case, at least in some of its occurrences (and more specifically in those cases where the state of the world permits an entity to be completely empty of its stereotypical contents, a classroom completely empty of pupils for instance) the under-specified term EMPTY OF X may well be used strictly. It is a genuine question for lexical semantics how to choose between these two analyses - if we need to, of course, given the remarks above about 'raw'.

5. Concluding remarks

To conclude, we will summarize some of the theoretical issues that have come up during the first six months of Corpus work in Lexical Pragmatics as starting points for further discussion. These concern:

(a) the highly diverse and highly flexible nature of lexical narrowing,

(b) the possible problems that our data may raise for default accounts of lexical interpretation,

(c) the existence of a gradient of cases between literal and loose use,

(d) the way in which lexical meaning is represented in permanent semantic memory (both within and across individuals),

(e) the particulars of the mechanisms of lexicalization: at what point does an ad hoc concept start becoming lexicalized? How can historical linguistic facts shed light on this process? What are the costs and benefits that encourage or impede lexicalization?

We would welcome any suggestions about how we might proceed in the next phase of our research.

References

Carston, R. 1997. Enrichment and loosening: complementary processes in deriving the proposition expressed? *Linguistische Berichte* 8:103-127.

Carston, R. 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.

Levinson, S. 2000. *Presumptive Meanings*. Cambridge: Cambridge University Press

Sinclair, J. 1991. *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Sperber, D. & Wilson, D. 1986.*Relevance: Communication and Cognition*. Oxford: Blackwell. (second edition, 1995)

Sperber, D. & Wilson, D. 1998. The mapping between the mental and the public lexicon. In P. Carruthers & J. Boucher (eds) *Language and Thought*, pp. 184–200. Cambridge: Cambridge University Press.

Stubbs, M. 2001. *Words and Phases: Corpus Studies of Lexical Semantics.* Oxford: Blackwell.

Wilson, D. 2003. Relevance and lexical pragmatics. Italian Journal of Linguistics/Rivista di Linguistica 15.2: 273-291.

Wilson, D. & Carston, D. 2006. Metaphor, relevance and the 'emergent property' issue. *Mind & Language* 21: 404-433.

Wilson, D. & Carston, R. forthcoming. A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. To appear in N. Burton-Roberts (ed.) *Advances in Pragmatics.* Palgrave, London.

Wilson, D. & Sperber, D. 2002.