

LINGUISTIC STRUCTURE AND INFERENTIAL COMMUNICATION

Deirdre Wilson

Department of Linguistics, University College London

Abstract: The replacement of code theories of communication by inferential theories has implications for the study of linguistic meaning. This paper surveys some of these implications and outlines an inferential theory — relevance theory — with which semantic analyses of context-dependent expressions, polysemous and ambiguous expressions, and non-truth-conditional expressions (e.g. mood indicators, discourse particles) might interact. Current analyses retain several assumptions of the code model which are unjustifiable on an inferential approach. These assumptions are dropped, and the prospects for a properly inferential treatment are explored.

Keywords: Relevance, pragmatics, context dependence, conceptual-procedural distinction, polysemy, discourse markers

1. INTRODUCTION

In the last 30 years, there has been a revolution in our views of human communication. Code theories, which treat utterances as encoding messages, have been replaced by inferential theories (inspired by the work of Paul Grice), which treat utterances merely as pieces of evidence about the communicator's intentions. This change in our views of communication has implications for the study of linguistic structure. My paper has two main aims: first, to draw out some of these implications, and second, to argue that the impact of the inferential revolution on the study of linguistic structure has yet to be fully realised. When it comes to the study of linguistic meaning, the code model still has us in its grip.

The paper is organised as follows. In section 2, I outline the code theory of communication and its implications for the analysis of linguistic meaning. In section 3, I outline the inferential

approach and show how its impact on the study of linguistic meaning has been blunted by the tendency to reduce it to an updated version of the code theory. In section 4, I consider some of the implications that a genuinely inferential theory of communication might have for the study of linguistic meaning.

2. CODE THEORIES OF COMMUNICATION

Code theories treat communication as involving (a) a set of observable signals, (b) a set of unobservable messages, and (c) a code, i.e. a set of rules or procedures pairing messages with signals. A communicator who wants to convey a certain message transmits the corresponding signal, which is received and decoded by the audience using an identical copy of the code. Successful code-based communication results in a duplication of messages: the message encoded is identical to the message received.

When applied to verbal communication, code theories yield a very standard picture of how utterances are understood. The signals are wave forms, which can be phonetically represented; the messages are thoughts, which can be conceptually represented; and the code is the grammar of a language (perhaps supplemented by pragmatic rules), which pairs phonetic representations with thoughts. Here is a quotation from Jerry Fodor which expresses something like this view:

'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 communicate 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: 106).

In much of the pragmatic and philosophical literature, the messages expressed by utterances are analysed into two components: a proposition P, which represents possible or actual states of affairs, and an illocutionary-force or propositional-attitude component F, which represents speech acts (asserting, asking) or cognitive states (believing, wondering). On this approach, utterances (1)-(3) would be analysed as potentially expressing identical propositions, but differing as to their propositional attitude or illocutionary force:

- (1) John will speak.
- (2) Will John speak?
- (3) Speak, John.

Let us assume that humans have a conceptual representation system rich enough to represent both propositions and propositional attitudes/illocutionary forces (henceforth 'attitudes' for short). Then, according to the code theory, sentences should encode two types of information, propositional and attitudinal, corresponding to the two components of messages. And indeed, the expressions in a language do seem to fall into two such groups. In the first come such regular content words as 'horse', 'sleep' and 'red'. In the second come illocutionary and attitudinal expressions such as the mood indicators in (1)-(3) above, the discourse adverbials in (4), the discourse particles in (5), the discourse connectives in (6), and perhaps the interjections in (7):

- (4) a. *Seriously*, how old are you?
b. *Unfortunately*, John is not here.
- (5) a. *Please*, you're standing on my foot.
b. Susan is, *alas*, a bore.
- (6) a. John is here, *because* I've just seen his car.
b. John must be here. *After all*, his car is here.
- (7) a. *Hey!* Get off the grass!
b. *Wow!* That looks nice!

Content words such as 'horse', 'sleep', etc. are generally treated as encoding concepts, constituents of conceptual representations. What do attitudinal expressions encode? According to what I will call the classical code theory, which claims that the code model is the only explanation of communication, the only way to communicate a concept is to encode it. On this approach, all meaningful expressions in a language, including attitudinal expressions, must encode concepts, and communicate exactly the concepts they encode. Many illustrations of this approach to attitudinal expressions are found in the generative-semantics literature of the late sixties and early seventies: for example in the performative hypothesis, which treats mood indicators as encoding higher performative clauses (e.g. Ross, 1970; Lakoff, 1972).

The classical code theory collapsed under pressure from two directions. First, Grice developed his inferential alternative to the code theory of communication, showing that messages could be communicated without being encoded (Grice, 1989, Part 1). Second, the attention of linguists became increasingly focused on the analysis of context dependence: in pragmatics, on the context dependence of utterance interpretation in general, and in semantics on the analysis of particular context-dependent expressions: indexicals, quantifiers, comparatives and so on. Context dependence is a phenomenon that cries out for inferential treatment. However, I will argue in the next section that what emerges from much of the literature on context dependence is not so much a fully inferential approach as an updated version of the code model.

3. INFERENCEAL THEORIES AND THE UPDATED CODE MODEL

Grice's work in the late sixties showed that communication is possible without the use of a code. All a communicator has to do to convey a certain thought is to get the audience to recognise her intention to convey it. In general, intentions are not decoded but inferred. When I see you take out a key and walk towards a door, I make the plausible inference that you intend to unlock the door; when you point to a closed window beside me, I infer that you intend me to open it. In both cases, alternative hypotheses are possible. Intention-recognition is a non-demonstrative inference process, not a fail-safe procedure. Grice's main contribution to pragmatics was to show that inferential intention-recognition could provide the basis for an alternative to the code theory of communication.

Perhaps the clearest cases of inferential communication are non-verbal, involving pointing, mimicry, and other types of ostension or display. Grice also showed that implicit verbal communication involves a substantial element of inference. Suppose Peter has a train to catch at 11.00, and it takes him half an hour to get to the station. Then Mary's utterance in (8) may conversationally implicate (9):

(8) It's 10.25.

(9) Peter should hurry up and get ready to leave.

In Grice's terms, (9) is a conversational implicature, because it is recoverable only on the assumption that Mary is obeying the Co-operative principle and maxims in uttering (8). It is a particularised conversational implicature, because it would not normally be carried by the same utterance in different situations. And it is best handled inferentially, because the connection between what is explicitly communicated and what is implicated is not arbitrary: anyone with access to the appropriate contextual assumptions and general pragmatic principles can work it out. The attempt to stipulate such implicatures in code-like terms would therefore be both difficult and pointless.

Explicit communication, by contrast, seems to involve a substantial element of decoding. The grammar of a language just is a code which pairs phonetic representations of sentences with semantic representations of sentences. The relation between phonetic and semantic representations is code-like, first, because phonetic representations lack logical properties and cannot be inferentially processed; and second, because an element of stipulation is involved: the meaning of a sentence cannot be computed from its phonetic properties using contextual assumptions and general pragmatic principles alone.

The overall utterance interpretation process, then, involves both decoding and inference, and an adequate account of verbal communication should show how the two interact. Perhaps the simplest hypothesis is that the borderline between decoding and inference coincides with the borderline between explicit and implicit communication, so that the proposition and attitude expressed would be recovered purely by decoding and the role of inference would be restricted to the calculation of implicatures. Grice himself seems to have held something like this view. He saw his Co-operative Principle and maxims, and hence the process of inferential intention-recognition, as applying mainly to implicit communication. He thought they might explain how the hearer of (8), having identified the proposition and attitude explicitly expressed, might go on to identify the implicature in (9). However, he seems not to have considered the possibility that his inferential approach might also help to determine the proposition and attitude explicitly expressed.

The main reason for rejecting this simple hypothesis is that utterances contain a huge variety of context-dependent expressions, which help to determine what is explicitly communicated (i.e. the proposition and attitude explicitly expressed). Context dependence seems to fit better with an inferential than a purely code-based approach. In the rest of this paper I will be mostly concerned with the relation between coding and inference in explicit communication. Here are some illustrations of the type of phenomena I have in mind (for survey and discussion, see Kempson, 1996).

Sentence (10) contains several context-dependent expressions: the pronouns 'I' and 'him', the ambiguous word 'charge', the past tense morpheme, and the scalar expression 'too much':

(10) I charged him too much.

Different propositions may be expressed by uttering (10) on different occasions: what is explicitly communicated may go well beyond what is encoded. An adequate account of explicit communication should answer three questions: what do these context-dependent

expressions encode, what do they communicate, and how is the gap between coding and communication filled?

Polysemous words like 'cut' or 'open' illustrate a further type of context dependence. Suppose Mary says (11a) or (11b) to Peter:

- (11) a. Cut the grass.
b. Cut the cake.

The word 'cut' would be differently interpreted in each example: (11a) would generally be understood as asking Peter to cut the grass with a mower; (11b) would generally be understood as asking him to cut the cake with a knife. One way of handling such cases would be to say that 'cut' encodes a single, general sense, which is narrowed down in different ways on different occasions. On this approach, the concept communicated by the use of 'cut' in (11a) or (11b) goes beyond the concept encoded. An adequate account of polysemous expressions should be able to say what polysemous words encode, what they communicate, and how the gap between coding and communication is filled.

What is explicitly communicated by attitudinal expressions such as those in (1)-(6) above is also highly context dependent. For example, (4a) is ambiguous: the illocutionary adverbial 'seriously' may be used to communicate that a serious question is being asked or that a serious answer is expected; on different occasions, the requestive particle 'please' in (5a) may be used to convey different requests, with different degrees of strength. In the speech-act literature, attitudinal expressions are generally referred to as *indicating* rather than describing. This supports the idea of an inferential approach. Indicating, like pointing, does not *determine* an interpretation, even given a context: it suggests a range of hypotheses, from which the hearer has to make a choice. An adequate account of linguistic indicators should be able to say what they encode, what they communicate, and how the gap between coding and communication is filled.

The fact that decoding and inference may combine in verbal communication suggests that several assumptions of the classical code theory should be rethought. In the first place, it is no longer necessary to assume that every expression encodes a full concept, a constituent of a message. Some expressions may encode something message-like but fragmentary: e.g. an incomplete logical form, a schematic concept which needs to be inferentially enriched. Other expressions may encode something that is not message-like at all: e.g. a procedure for supplying a concept, a constraint on inference. Analyses along these lines are already being developed: in artificial intelligence using notions such as schema, frame or script, in speech-act theory using the notion of indicating versus describing, in relevance theory using the notion of procedural versus conceptual meaning, and in formal semantics using the notion of content versus character (for discussion, see Recanati, 1997). If context dependence is to be handled in inferential terms, these analyses need to be articulated with a properly inferential account of how the gap between coding and communication is filled.

In the second place, it is no longer necessary to assume that what is encoded by an expression is part of the intended message. To the extent that communication is inferential, what is encoded may provide no more than a piece of evidence about that message. Even when a full concept is encoded, it need not be identical to the concept communicated. This applies as

much to regular content words like 'sleep' or 'red' as to polysemous words like 'cut' or 'open'. Thus, lexical semantics in general needs to be articulated with a properly inferential account.

In the third place, it is not necessary to assume that every concept communicated by an utterance *could* have been encoded. In both the semantic and the pragmatic literature, inferential communication is often seen merely as an effort-saving device, which spares the speaker the trouble of spelling out what could as well have been encoded. This assumption, which is rarely explicitly defended, needs some justification. It may be that inferential communication opens up new routes, to otherwise inaccessible end-points; this possibility needs to be explored. We should even question the fundamental assumption of the code theory that successful communication involves the exact reduplication of messages. Successful inferential communication may be better thought of as creating similarity, rather than identity, of thoughts.

The inferential approach to communication should thus lead to a radical reassessment of semantic analyses based on the assumptions of the classical code theory. So far, this reassessment has not taken place, partly because the assumptions of the code model (particularly for linguists) are very hard to drop. It is not easy to develop properly inferential accounts. Grice's proposals are extremely vague, and do not imply any explicit procedures for inferential intention-recognition. In trying to construct an explicit account of disambiguation or reference resolution, for example, it is not helpful to be told that the hearer should choose an interpretation that satisfies the Co-operative Principle and maxims. One wants to know how this is done. When code-like procedures are easy to formalise, and no one quite knows how inferential procedures should go (see e.g. Fodor, 1983), it is tempting to fall back on a code-like approach. As a result, fully inferential analyses of the context-dependent aspects of explicit communication have been slow to get under way.

Grice himself seems to have thought of explicit communication in code-like rather than inferential terms. He seems to have assumed that the notion of 'context' required for the analysis of context-dependent expressions was reducible to a finite set of parameters (for speaker, hearer, time and place of utterance, and so on), which would interact with encoded meaning to determine the correct interpretation without reference to pragmatic principles or speaker's intentions. On this approach, there would at most be an extra layer of inferential confirmation of a hypothesis recovered essentially by decoding. As I will show, this code-like conception of context-dependence is still very influential, and blocks the development of a fully inferential account.

What emerges from much of the recent semantic and pragmatic literature, then, is not a properly inferential account but an updated version of the code theory, in which the role of inference is reduced to a minimum and several assumptions of the code model are retained. I will give three examples to show how this reduction may go.

The first is from speech-act theory. Here, John Searle was quick to recognise the value of the Gricean inferential approach, but he thought he saw an objection. Grice's approach 'fails to account for the extent to which meaning can be a matter of rules and conventions' (Searle, 1969: 43): a speaker is not free to intend just any utterance to express just any thought. Searle proposes to deal with this by adding the following stipulation: in the literal performance of illocutionary acts, the speaker should intend the hearer not merely to recognise her intentions, but to recognise them 'in virtue of his knowledge of the rules for the sentence uttered' (Searle,

1969: 45). In other words, the speaker should intend her intentions to be recognised not by inference but by decoding. Thus (at least at the level of explicit communication) inference is seen as merely adding a layer of confirmation to a hypothesis recovered by decoding.

My second example involves the formal-semantic treatment of indexicals. David Kaplan (1989: 523) analyses what he calls 'pure indexicals' (e.g. the pronoun 'I') as encoding not a content but a *character*, defined as a function from contexts to contents, with the notion of context understood in traditional parametric terms. According to Kaplan, the rules which constitute the character of an indexical expression

'tell us for any possible occurrence of the indexical what the referent would be ... They *determine* the content ... for a particular occurrence of an indexical. But they are not a *part* of the content' (Kaplan, 1989: 523).

Precisely because character is seen as *determining* the content given a context, it is code-like rather than inferential; and precisely because it is code-like, it has limited application. As has often been noted, 'I' is the only pronoun for which Kaplan's approach is more than remotely plausible (and even this has been disputed): it clearly fails for pronouns like 'he', which may have several potential referents in a given context, from which the hearer has to make a choice. Kaplan's treatment of 'pure indexicals' illustrates a strategy that is very common in the formal literature: to provide a code-like treatment for a small subset of cases, and leave the rest to a theory of inferential intention-recognition or a future code-based account.

Similar points apply to my third example, which involves the treatment of polysemy. A common way of explaining the variations in the interpretation of a polysemous expression like 'cut' in (11a) and (11b) is to say that the general meaning gets specified in context by the properties of the direct object: thus, cutting grass means mowing it, cutting cake means using a knife, cutting hair means using scissors, and so on. In this way, a standard range of senses is generated on a code-like basis, taking properties of the linguistic context into account (Pustejovsky, 1995; Pustejovsky and Boguraev, 1996; for alternative treatment, see Searle, 1980, 1983). The problem with this approach is that it does not generalise: it is easy to think of situations in which the word 'cut' in (11a) would receive a non-standard interpretation, which would remain unexplained on this approach. A common response is to claim that the code-like rules for interpreting polysemy are no more than default procedures, which may be overridden by further contextual information. However, default procedures are open to the same objection made above to Kaplan's indexical rules: they bring a small subset of interpretations within the scope of the code model, and leave the rest to inferential intention-recognition or a future code-based approach. A genuinely inferential account should generalise to deal with the full range of cases in a flexible and open-ended way.

I could go on. For example, Gerald Gazdar's work on formal pragmatics (Gazdar, 1979) has inspired a variety of computational approaches to the derivation of implicatures, which are more code-like than inferential, and which it would be interesting to discuss. However, I will stop there and turn to the prospects for a fully inferential approach.

4. INFERENCE COMMUNICATION

I have criticised Grice's inferential approach for its lack of explicit interpretive procedures; I have also criticised those who do propose explicit procedures for falling back on the code model. You might ask what is left. How can explicit procedures *not* be code-like? I will try to answer that question now.

In the last fifteen years, Dan Sperber and I have been developing an inferential account of communication which aims to be more explicit than Grice's ground-breaking but very sketchy account. The basic ideas of the theory are contained in a definition of relevance and two principles (Sperber and Wilson, 1986/1995). Relevance is defined as a property of utterances and other inputs to cognitive processes. The processing of an input may yield some cognitive effects (e.g. strengthening of existing assumptions, contradiction and elimination of existing assumptions). Other things 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 mental effort. Other things being equal, the greater the effort, the lower the relevance. On the basis of this definition, two principles are proposed:

(12) *Cognitive principle of relevance:*

Human cognition tends to be geared to the maximisation of relevance.

(13) *Communicative principle of relevance:*

Every utterance (or other act of overt 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.

The search for a relevant interpretation is triggered by the automatic decoding of an utterance, which activates in the hearer's mind a conceptual representation or logical form. This initial step in comprehension involves some mental effort, and therefore creates an expectation of compensating cognitive effects. According to the communicative principle of relevance, the effects achieved should be enough to justify the effort (or at least for the speaker to have expected them to justify the effort; see Sperber, 1994). This provides the theoretical motivation for the following interpretive procedure, which we claim is automatically applied to the on-line processing of utterances. The hearer takes the conceptual structure recovered by linguistic decoding; following a path of least effort, he *enriches it 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:

(14) *Relevance-theoretic comprehension strategy:*

- a. Follow a path of least effort in computing cognitive effects;
- b. Stop when the expected level of relevance is reached.

Let me show how this procedure might apply to the interpretation of the utterance in (15):

(15) They gave Socrates the hemlock, and he drank.

In order to know what proposition was expressed by (15), the hearer must assign reference to the referential expressions 'Socrates', 'they' and 'he'; and disambiguate the ambiguous word 'gave' (does it mean 'hand over' or 'donate free of charge?') and the polysemous word 'hemlock' (does it mean the plant or the drink?). Following a path of least effort, he would probably take 'he' to refer to Socrates, and assume, moreover, that what he drank was the hemlock he had just been given. Using the encyclopaedic information that hemlock is poisonous, he could then infer the implicature in (16), which would in turn lead on to further cognitive effects:

(16) Socrates was poisoned.

Notice that this implicature would not have been recoverable if 'he' and 'drink' had been differently interpreted. There is a mutual dependency between the propositions expressed and implicated, and the relevance-theoretic comprehension strategy assumes that they will be mutually adjusted until the expected level of relevance is reached.

Consider, now, how this procedure might apply to the interpretation of 'cut' in Mary's utterance in (17):

- (17) a. Peter: What would you like me to do?
 b. Mary: Cut the grass.

Given that Peter has just asked a question, Mary's response should create a specific expectation of relevance. Peter should expect to use the decoded conceptual structure of her utterance as a starting point for inferring an answer to his question, i.e. a suggestion about what she would like him to do. Let us assume that the polysemous verb 'cut' encodes a single, general meaning, which may be inferentially narrowed down in a variety of ways, providing access to a range of further (retrieved or constructed) concepts, each involving a different amount of effort and yielding a different effect. Encyclopaedic information jointly associated with the verb and its direct object makes the concept *mow* particularly accessible, so that the inferential route in this direction is short and obvious. According to the relevance-theoretic comprehension procedure, if the resulting interpretation yields enough effects to satisfy his expectation of relevance, Peter should look no further, and other, more costly interpretations should not be considered.

In the case of (17b), the interpretation of 'cut' as meaning *mow* does not yet imply an answer to Peter's question. Mary would probably not like him to mow the grass just any old way. To satisfy his expectation of relevance, he must further narrow down the interpretation until it yields enough effects to make the utterance relevant in the expected way. The relevance-theoretic comprehension process can thus be seen as involving parallel adjustment of explicit content and implicatures (intended cognitive effects): Peter starts with an expectation of relevance which warrants the derivation of specific implicatures, for which the explicit content must be adequately enriched.

This account goes beyond the code-like treatment discussed above. It explains how 'cut' may be used to communicate more than its general encoded sense, and more than the standard sense *mow* that would result from taking the properties of the direct object into account. It also deals with more standard cases by the same procedure of parallel adjustment: standard senses are easily accessible by short and obvious routes. In this way, the full range of

examples can be dealt with by the same pragmatic process of inferring the intended sense on the basis of encoded concepts and contextual information. Defeasible rules are unnecessary. Relevance theory suggests a genuinely inferential explanation of how polysemy is resolved. (For relevance-theoretic accounts of polysemy, see Carston, 1996a, 1998; Papafragou, in preparation; Sperber and Wilson, forthcoming; Wilson and Sperber, forthcoming).

Notice, now, that the notion of cutting that Peter constructs in interpreting (17b) may be an ad hoc one, which he has never used before and may never use again. If so, it will probably not be identical to the one that Mary has in mind. She may have quite complex preferences about how the grass is to be cut on this particular occasion, and Peter is unlikely to grasp them all. This is not a failure of communication, as the code theory would predict. It is an illusion of the code theory that communication invariably aims at exact duplication of meanings; often a looser kind of understanding is intended and achieved. In the case of (17b), Peter's ad hoc notion of cutting may be close enough to Mary's to allow them to co-ordinate their behaviour. The fact that it may not correspond to anything encodable in English provides further evidence against the classical code theory and many updated versions, which retain the assumption that anything that can be communicated can be encoded. (For further discussion, see Sperber and Wilson, forthcoming.)

The relevance-theoretic comprehension strategy generalises straightforwardly to any expression, ambiguous or not, which encodes a concept, a set of concepts or a schematic concept (i.e. any descriptive expression, in speech-act terms). It allows for the possibility that the concept communicated may not be identical to the concept encoded, and explains how the gap between coding and communication is bridged.

I would now like to consider what light the relevance-theoretic framework can shed on the analysis of indicators. In section 2, I claimed that the classical code theory offers no account of the distinction between describing and indicating, and that the notion of indication seems well adapted to an inferential approach. I will now try to show how indicators might be analysed on an inferential approach. This has some implications for the treatment of indexicals, as I will also show.

Pointing is a natural device which alters saliencies of information. It may interact with inferential utterance comprehension in a variety of ways: for example, by increasing the accessibility of an intended referent or a chunk of contextual information, thus making it more available to the relevance-theoretic comprehension procedure (Breheny, to appear). Language may develop coded means for manipulating saliencies of information. Within the current relevance-theoretic framework, this is how indicators are viewed.

Here are some examples. Diane Blakemore (1987), who was the first to suggest this approach within the relevance-theoretic framework, treated discourse connectives such as 'so' and 'after all' as encoding not concepts but procedures for identifying intended contexts and/or cognitive effects. That is, she saw discourse connectives as increasing the salience of certain inferential routes which, given the fact that they had been intentionally made salient, the hearer was entitled to use.

Blakemore's discourse connectives are constraints on implicatures. Wilson and Sperber (1993) extended the procedural approach to explicit communication, analysing a variety of expressions, including pronouns, mood indicators, and illocutionary and attitudinal particles,

as encoding procedural constraints on explicatures (see also Wilson and Sperber, 1988). For example, the pronoun 'he' was seen as increasing the salience of a certain class of referents, mood indicators as increasing the salience of certain speech-act or propositional-attitude descriptions, and so on. Interesting work has been done in this framework on a variety of 'hearsay' indicators, which may be seen as increasing the salience of metarepresentational descriptions (e.g. Blass, 1989; Ifantidou, 1994; Itani, 1996; Carston, 1996b; Wilson, forthcoming; Noh, in preparation).

Analyses along these lines treat procedural ('indicating') expressions as intrinsically geared to inferential intention recognition. Different types of procedural expression may be geared to different aspects of the comprehension process: reference resolution in the case of pronouns, attitude recognition in the case of mood and illocutionary-force indicators, or the identification of implicatures in the case of discourse connectives, among others. But whatever their more specific functions, on this approach they all share the same general function of narrowing down the search space for inferential comprehension. Thus, the various types of procedural expression may receive a unified treatment.

If this approach is along the right lines, it follows that certain expressions standardly treated as indicators in the speech-act literature are not procedural at all. Consider the illocutionary and attitudinal adverbials 'seriously' and 'unfortunately' in (4) above. Although these are non-truth-conditional, and might therefore be expected to encode procedures, they clearly fall together with the conceptual rather than the procedural vocabulary. Unlike most procedural expressions, they have synonymous truth-conditional counterparts which are analysable as encoding concepts, e.g. the regular manner adverbials in (18):

- (18) a. John spoke frankly.
b. That letter was phrased unfortunately.

Unlike most procedural expressions, they can also figure in syntactically complex phrases (e.g. 'Quite frankly', 'Most unfortunately', 'More regrettably than I can say'), which appear to undergo the same compositional rules as regular conceptual expressions. Similar points apply to the 'hearsay' adverbials 'reportedly' and 'admittedly': although these are functionally quite similar to the 'hearsay' particles in other languages, there is good reason to think that, unlike 'hearsay' particles, they fall together with the conceptual rather than the procedural vocabulary (Wilson and Sperber, 1993; Ifantidou, 1994). Thus, the distinction between conceptual expressions, which encode something concept-like, and procedural expressions, which encode constraints on inference, cross-cuts the traditional distinction between truth-conditional and non-truth-conditional (describing and indicating, or propositional and attitudinal) expressions.

5. CONCLUSION

In this paper, I have surveyed some of the implications of the move to inferential theories of *comprehension for the analysis of linguistic meaning*. I have also tried to show how a particular inferential theory — relevance theory — might be articulated with semantic analyses of various types of linguistic expression. These proposals are quite tentative, and leave many questions unanswered. Here I will mention just three. First, I have suggested that some expressions encode procedures, which constrain the inferential phase of comprehension, and others encode schematic or skeletal concepts, which need to be inferentially fleshed out.

What does this distinction amount to in empirical terms, and how do we apply it to actual examples? (Pronouns might be a test case, since they arguably lend themselves to either a procedural or a schematic-conceptual approach). Second, I have suggested, in line with previous work in relevance theory, that procedural expressions do not automatically encode concepts (although some expressions, e.g. 'if', might well be analysed as encoding both a concept and a procedure). Given the wide variety of conceivable conceptual formats, it is worth considering an alternative analysis, on which certain concepts contain procedural entries (analogous to the logical entries discussed in Sperber and Wilson, 1985/1996, chapter 2) in which the procedures associated with procedural expressions might be stored. In that case, the conceptual-procedural distinction would amount to a distinction between two types of encoded concept, rather than between the encoding of a concept in one case and a procedure in the other. Finally, possible interactions between the relevance-theoretic comprehension strategy, the notion of procedural meaning and the many sophisticated semantic studies of determiners, inflectional morphology and focal stress should also be pursued (see e.g. Breheny, in preparation, in press.) What I mainly hope to have shown in this paper is something more general: that an inferential approach to communication does have implications for the study of linguistic structure, and that the prospect of exploring them is an interesting and exciting one.

ACKNOWLEDGEMENTS

I am grateful to Dan Sperber for valuable insights on the topic of this paper, and for permission to draw on some of our forthcoming work. I would also like to thank Robyn Carston, Neil Smith, François Recanati and the members of the University College London Pragmatics Reading Group for interesting comments and discussion.

REFERENCES

- Blakemore, D. (1987). *Semantic constraints on relevance*. Blackwell, Oxford.
- Blass, R. (1989). Grammaticalisation of interpretive use: the case of *re* in Sissala. *Lingua* **79**, 229-326.
- Breheny, R. (forthcoming). Interface economy and pro-active focus. To appear in *MIT Working papers in linguistics*.
- Breheny, R. (in press). A unitary approach to the weak and strong interpretations of definites. To appear in *UCL Working papers in linguistics*.
- Breheny, R. (in preparation). *Relevance and procedural semantics*. University of London PhD thesis.
- Carston, R. (1996a). Enrichment and loosening: Complementary processes in deriving the proposition expressed? *UCL Working papers in linguistics* **8**, 205-232.
- Carston, R. (1996b). Metalinguistic negation and echoic use. *Journal of pragmatics* **25**, 309-330.
- Carston, R. (1998). *Pragmatics and the explicit-implicit distinction*. University of London PhD thesis.
- Fodor, J. (1975). *The language of thought*. Crowell, New York.
- Fodor, J. (1983). *Modularity of mind*. MIT Press, Cambridge, MA.

- Gazdar, G. (1979). *Pragmatics: Implicature, presupposition and logical form*. Academic Press, New York.
- Grice, H.P. (1989). *Studies in the way of words*. Harvard UP.
- Ifantidou, E. (1994). *Evidentials and relevance*. University of London PhD thesis.
- Itani, R. (1996). *Semantics and pragmatics of hedges in English and Japanese*. Hituzi Syobo, Tokyo.
- Kaplan, D. (1989). Demonstratives. In: *Themes from Kaplan* (J. Almog, J. Perry and H. Wettstein, (Eds)), 481-563. OUP, Oxford.
- Kempson, R. (1996). Semantics, pragmatics and natural-language interpretation. In: *The handbook of contemporary semantic theory* (S. Lappin, (Ed.)), 561-98. Blackwell, Oxford.
- Lakoff, G. (1972). Linguistics and natural logic. In: *Semantics of natural language* (D. Davidson and G. Harman, (Eds)), 545-665. Reidel, Dordrecht.
- Noh, E-J. (in preparation). *The semantics and pragmatics of metarepresentation in English: A relevance-theoretic approach*. University of London PhD thesis.
- Papafragou, A. (in preparation). *Modality and the semantics-pragmatics interface*. University of London PhD thesis.
- Pustejovsky, J. (1995). *The generative lexicon*. MIT Press, Cambridge, MA.
- Pustejovsky, J. and B. Boguraev, (Eds) (1996). *Lexical semantics: The problem of polysemy*. Clarendon Press, Oxford.
- Recanati, F. (1997). *Déstabiliser le sens*. Centre de Recherche en Epistémologie Appliquée, Paris, Rapport 9714.
- Ross, J. (1970). On declarative sentences. In: *Readings in English transformational grammar* (R. Jacobs and P. Rosenbaum, (Eds)), 222-72. Ginn, Waltham, MA.
- Searle, J. (1969). *Speech acts*. CUP, Cambridge.
- Searle, J. (1980). The background of meaning. In: *Speech-act theory and pragmatics* (J. Searle, and F. Keifer and M. Bierwisch, (Eds)), 221-32. Reidel, Dordrecht.
- Searle, J. (1983). *Intentionality*. CUP, Cambridge.
- Sperber, D. (1994). Understanding verbal understanding. In: *What is intelligence?* (J. Khalfa, (Ed.)), 179-98. CUP, Cambridge.
- Sperber, D. and D. Wilson (1986/1995). *Relevance: Communication and cognition*. Blackwell, Oxford.
- Sperber, D. and D. Wilson (forthcoming). The mapping between the mental and the public lexicon. To appear in *Language and thought* (P. Carruthers and J. Boucher, (Eds)). CUP, Cambridge.
- Wilson, D. (forthcoming). Linguistic metarepresentation. To appear in *Metarepresentation* (D. Sperber, (Ed.)).
- Wilson, D. and D. Sperber (1988a). Mood and the analysis of non-declarative sentences. In: *Human agency: Language, duty and value* (J. Dancy, J. Moravcsik and C. Taylor, (Eds)), 77-101. Stanford UP, Stanford, CA.
- Wilson, D. and D. Sperber (1993). Linguistic form and relevance. *Lingua* 90: 1-25.
- Wilson, D. and D. Sperber (forthcoming). Pragmatics and time. To appear in *Relevance theory: Applications and implications* (R. Carston and S. Uchida, (Eds)). Benjamins, Amsterdam.