

16 Pragmatics: Language and Communication

RUTH KEMPSON

1 The Puzzle of Language Use: How Do We Ever Understand Each Other?

How language is used may not seem to warrant a topic on its own. “When I use a word, it means just what I choose it to mean – neither more, nor less.” This is the view of language expressed by Humpty-Dumpty in Lewis Carroll’s *Alice through the Looking Glass*. It may seem mere commonsense that describing a person’s knowledge of a language involves simply describing how they use the language. But this can’t be *all* there is to say about the relation between our knowledge of language and the way we use it, for words invariably convey much more than they mean. The question is: What does this apparent gap between language use and meaning tell us about language? As Alice quite reasonably objected: “The question is whether you *can* make a word mean so many different things.”

The starting point for looking at language use is to consider why Humpty Dumpty might have been right. Consider the conversation in (1):

- (1) A: Can you cook?
B: I know how to put a kettle on.

Why are such conversations possible – why, for example, did B not just say “No”? After all, if B had understood the question, and knows what such a question “means,” she should know that this type of question is a request for the answer *yes* or the answer *no*. So what did she “mean” by choosing that indirect mode of reply in apparent violation of this rule? In what sense could B have “meant” that she never did anything in the kitchen other than putting a kettle on, so the answer is that she cannot cook, or has no interest in cooking, etc.

Then there is the way we can describe things by using words in ways which only have a very indirect relation to what a dictionary would indicate is their meaning:¹

(2) You're a real race-horse.

How can the words *real race-horse* be predicated of a single individual, someone transparently not a horse? What could a speaker mean by saying something apparently blatantly far from the truth? One has sympathy with Humpty and yet also with Alice – What is it that we do with language that makes statements such as (2) meaningful and effective?

What these two examples show is that words can be used to convey both more than what they conventionally mean and also something quite different. But, if this is true, how do we manage to sort out which is which? How do we know when an expression is to be taken at its face value, when it is to be taken as conveying rather more than what it actually presents, and when it has to be interpreted in some other, metaphorical, way?

This is only the beginning of the problem of understanding meaning in language use, for even setting aside supposedly special rhetorical effects as in (2), expressions in language are notoriously prone to ambiguity. Yet, by and large, we have no difficulty in sorting out what the speaker is intending to say to us. How do we manage this? To take an extreme example (Sperber and Wilson 1982), consider (3):

(3) A: How is your son?
B: He's grown another foot.

Why is it that A is most unlikely to respond with horror, suggesting that it should be amputated? And, with a much less extreme case, how does the hearer decide whether the phrase "in March" describes the time of the exam, the time of discovering the results of the exam, or the time at which Sue made her report:

(4) Sue reported to the Committee that Joan learnt that she had failed the exam in March.

Even in (1) itself, B's reply could have been intended as a deliberate understatement indicating her cooking talent, which she may be implying A ought to know about. How does A decide which B meant?

To put the problem in its most general form, when we probe the obvious truth that our knowledge of language is reflected in the way we use it, we seem to be faced with a perplexing mystery. How is it that using language is in general so effortless, when there seems to be no limit to what words can convey or what ambiguities they can give rise to? *Pragmatics* seeks to provide

an answer to this question. Pragmatics is the study of communication – the study of how language is used. This study is based on the assumption of a division between knowledge of language and the way it is used;² and the goal of pragmatics is taken to be that of providing a set of principles which dictate how knowledge of language and general reasoning interact in the process of language understanding, to give rise to the various different kinds of effects which can be achieved in communication. In this chapter, we shall look at different approaches to pragmatics. We will look first at the question of the assumed interaction between language-particular (= linguistic) and general (= nonlinguistic) types of information; and I will argue in section 2 that what we need is a model which allows integration of linguistic and nonlinguistic forms of interpretation at all stages of the interpretation process. In section 3 we will take up the question of how the hearer chooses an appropriate interpretation, and I will introduce the two major alternative views of how interpretations are selected. In section 4, I will set out the view that all words should be defined in terms of procedures for building up structures representing interpretation in context (*propositional structure*). I will give one illustration to show how pragmatic processes can feed into linguistic processes, and the chapter ends with a discussion of the general significance of this “procedural” approach for establishing what it means to “know a language.”

2 Pragmatics as the Application of Conversational Principles to Sentence Meanings

The starting point for studies in pragmatics³ is the mismatch, often a big one, between what words “mean,” which is encoded in rules of the language, and what speakers “mean” by using them, which may be much richer. In (1), the words in the sentence B utters convey the information that the speaker has the knowledge of how to put a kettle on. What speaker B means by using these words (on the interpretation indicated first) is that she cannot cook, opening up the possibility of further inferences such as that she has no interest in cooking. One way in which one might seek to generalize from this particular instance is to take (1) as evidence for the two aspects of language use being quite separate. There is the knowledge of language, on the one hand, which dictates the meanings of words and the ways in which they can combine to form sentence-meanings (to be studied under the label *semantics* as part of the grammar of a language). This is called the encoded meaning. On the other hand, there are general pragmatic principles (which I shall initially call “commonsense reasoning” principles) which enable a hearer to establish some rather different and richer interpretation – the nonencoded part of meaning. On this view, we would say that a hearer parsing B’s utterance in (1) above first uses

rules of the language to work out that B was conveying that she knew how to put a kettle on, and then, only subsequently, uses principles of commonsense reasoning to work out, say, that because B did not reply more directly, it must be that she is expecting A to access the knowledge that people who confess to knowing only how to put on a kettle are no good at cooking in order to work out the intended import of her answer – that she cannot cook. Once such general forms of reasoning are invoked, one might also anticipate that such indirect forms of answer convey more information than a simple negative answer would provide because they trigger such general reasoning processes – for example, communicating from B's implication that she has no interest in cooking the further suggestion that any attempt to extend the conversation with topics associated with food is likely to prove unsuccessful. This approach to pragmatics can be summed up as the view that a grammar of a language provides a characterization of meaning for each individual sentence as articulated in some semantic theory, and that pragmatic principles apply to the output of the grammar-internal characterization to yield its full import in context. This view has been justifiably influential (the Gricean view – Grice 1967, 1975, 1989). It keeps knowledge of language and general reasoning capacities quite separate – even in language use, the latter is seen as being brought into play only after the hearer has established a complete and use-independent characterization of sentence-meaning. It is particularly appropriate for a Chomskian view of linguistic knowledge as a body of knowledge which is encapsulated and independent of other cognitive capacities we humans display (see chapter 00). Moreover, given the full array of rhetorical effects such as metaphor, irony, etc., all of which are uses of expressions in context in some sense, the proposed approach maintains a natural separation between literal uses of words, which are reflected in sentence-meanings, and the various non-literal uses to which they may be put.

2.1 Knowledge of language: sentence-meanings as partial specifications of interpretation

There is however good reason to reject this simplistic separation of rules of semantics as part of grammar, and what I have so far called commonsense reasoning principles. The problem for this “clean” view is that we use commonsense reasoning, whatever this consists in, not merely in working out why a speaker has said something, but also in establishing what she has said in using the words chosen. Consider the conversation (5), remembering the event in the late summer of 1997 in which Diana Princess of Wales died in a car crash, vast crowds gathered in mourning outside Kensington Palace in Kensington Gardens where she had lived, quantities of flowers were left immediately outside the Palace gates which were reported to give off an intense aroma, all part of a series of events which led up to a funeral at which the singer Elton John sang the song “Candle in the Wind.”

- (5) A: Elton John sang at Diana's funeral. Did you see it?
 B: I spent the whole day in Kensington Gardens. The smell was amazing.

The questions posed by this conversation are:

- how does B understand what A has referred to by using the word *it*?
- how does talking about Kensington Gardens answer A's question?
- why is *the whole day* taken to refer to the day of Princess Diana's funeral?
- how does A understand what the words *the smell* refer to?

The general problem we want to use this example to address is: how do language-internal principles interact with more general reasoning capacities? *It* presumably means the funeral service of Princess Diana. Since the funeral was televised, there were several means of seeing the funeral – either on television or by attending the event in person. B replies that she went to the area surrounding Kensington Palace for “the whole day.” Since the funeral involved a procession from Kensington Palace to Westminster Abbey, as well as the service itself, B's reply is taken to imply that she was in the area in which the funeral took place at the time of the funeral, so her reply provides a positive answer to A's question – she was at the funeral, so indeed she saw it. B follows this reply up with the words *the smell*. She relies on A's being able to understand what these words mean by recovering information about the mass of flowers left in Kensington Gardens, and hence success in referring to the smell of these flowers. Almost none of this information is knowledge about the English language. There is nothing in the meaning of the word *funeral* which specifies a relation between this type of service and Kensington Gardens. There is nothing about the past tense in English which requires that the whole sequence of sentences should be taken to be about the same event.⁴ And there is nothing in A's knowledge of the word *smell*, either, which specifies information about flowers left outside Kensington Palace in September 1997 – it is A's presumed knowledge of the objects so described which B relies on in choosing the words *the smell* as she does – just as it is A's presumed knowledge of the event which B relies on, in choosing to reply to A's question indirectly by referring to Kensington Gardens.

In this conversation, we see that the separation between knowledge of language and commonsense reasoning is much more blurred than in the conversation (1). There is no sense in which B can be said to have parsed the sentence *Did you see it* and worked out the meaning A intended to convey using language-internal principles alone – only subsequently bringing into play more general commonsense reasoning principles to work out some broader message. Similarly with A in processing B's reply. The different kinds of knowledge – one language-based, the other a much more general store of knowledge – have to be combined together in understanding what the word *it*, *the whole day*, and *the smell* mean in the particular context in which they are used. However, so the argument might go, there is certainly something separate about

our knowledge of language and general reasoning capacities, because B has to parse the sentence first in order to establish that the word *it* has been uttered, she has to know that *it* is a pronoun which is the sort of word that is used to pick out some entity in the scene described in the conversation, that in the sentence A utters, the word *it* is presented as object of the verb *see*, and so on – all strictly English-internal information without which B will not retrieve the meaning A intends to get across. In reply, B uses the words *the whole day* relying on A's knowledge of English that an expression of this type can be taken to pick out a period of time relative to which the verb is understood. And, choosing the words *the smell*, equally, relies on A's being able to establish that *the smell* is intended as the subject being talked about, and *amazing* as the adjective predicated of it. Part of this language-specific knowledge is also the knowledge that *the* is the sort of word which leads to identifying some entity in the scenario being described; but, though words such as *it* and *the* trigger a process of identifying what is being talked about, they rely on a framework of structure constructed by parsing the sentence by language-internal rules. On evidence such as this, according to one current theory, the language system projects sentence-sized structures as sentence-meanings (*logical forms*), though these are incomplete. In the case of (5) the *logical forms* corresponding to A's question and B's reply will be along the lines indicated in (6) in which some parts of the interpretation are not filled in:

- (6) A: Question: Hearer saw X at time t_1 .
 B: Speaker spent day-Y at Kensington-Gardens at time t_2 . Smell-Z was amazing at time t_3 .

X, Y, Z, Speaker, Hearer, t_1 , t_2 , t_3 are all parts that are missing and have to be filled in from context.⁵ It is these missing parts which are transformed using general pragmatic principles of reasoning to create completed structures which more directly represent the thought that the speaker intended to convey. (Such structures are called *propositions* to distinguish them from the *sentence* that expresses them.)⁶ Representations of the *propositions expressed* by A's question and B's answer might be given as follows:⁷

- (7) A: Question: B saw Princess Diana's funeral on Saturday.
 B: Spent Saturday at Kensington Gardens. The smell of flowers outside Kensington Palace on Saturday was amazing.

The pragmatic principles which dictate how these choices are made also have a much more general role to fulfill. For, on this view, it is these very same principles that determine the broad array of metaphorical, ironic, and other effects which a sentence can convey in context. The overall picture of interpretation is that grammar-internal principles articulate both syntactic and semantic structure for sentences, a semantic structure for a sentence being an incomplete specification of how it is understood. Pragmatic theory explains how

such incomplete specifications are enriched in context to yield the full communicative effect of an uttered sentence, whether metaphorical, ironical, and so on. This view is the view adopted by relevance theory (Sperber and Wilson 1995); and it is this view and the Gricean view which constitute the two major approaches to utterance interpretation.

2.2 *Knowledge of language: a set of procedures for interpreting utterances*

There is reason to think that the interaction of language-particular knowledge and general commonsense reasoning is even more pervasive than is allowed for by this modified separation of linguistic abilities and general commonsense reasoning principles. Consider a different way the conversation (5) might have gone:

- (8) A: Elton John sang at Diana's funeral. Did you see it?
 B: I spent the whole day in Kensington Gardens. I felt I had to. The smell was amazing. Incredibly moving.

What is it that B has conveyed in uttering the words *I had to*: how is it that he can rely on his hearer, A, to reconstruct from the word *to* a structure corresponding to "spend the whole day in Kensington Gardens"? In this case, the speaker is giving a fragment which relies almost in its entirety on the ability of the hearer, given the context in which the string is uttered, to reconstruct some appropriate structure corresponding to the meaning of what in that context B is trying to convey. It is not that B's words themselves project a full sentence-structure with lexical meanings defined independent of context, on the basis of which pragmatic principles provide some add-on means of identifying what is being talked about. Here the words are mere triggers for a process of reasoning which has to reconstruct not only who is being talked about and why, but also the process of building the structure which corresponds to the meaning the speaker intended to convey. And all by the word *to* with nothing following it. Similarly, in processing *incredibly moving*, the hearer has to use some form of reasoning to establish what to take as the subject of the expression – is it the flowers that are so moving, or the event in general? And what is the basis on which B can rely on A to build the structure into which the expression *incredibly moving* projects a predicate? There is no apparent subject to this sentence, and no verb – so where does the structure come from? Is there a rule internal to English which says that sentences with no subject and no verb are well formed? If there is, it is certainly not one which any grammar book has ever included. Evidence such as this suggests a third view. On this view, there is still separation between the intrinsic content of individual words, which is encoded (i.e. part of what an English speaker knows in virtue of knowing the language), and the process of reasoning with them, which is not encoded.

Nevertheless, the process of building up the structure corresponding to some conveyed interpretation involves integrating one's knowledge of language with these general processes of reasoning at every step of the interpretation process. It is not that the rules of English syntax give rise to completed structures to which pragmatic principles fill in whatever open slots are left in the structure. Rather, we need to define a concept of structure internal to language which can be used both in building up meaning for complete sentences, and to process radically incomplete sentence "fragments" (cf. section 4.2 where the question of interpreting such fragments is taken up again).

3 The Process of Reasoning: How Do Hearers ever Manage to Choose the Right Interpretation?

I have so far sketched three possible perspectives on the nature of the interaction between what we might agree was knowledge of the individual language, and more general knowledge about the individuals being described. But we have not yet begun to look at the principles which form what is arguably the center of any pragmatic theory, which explain how a hearer selects the interpretation which the speaker intended – the so-called principles of commonsense reasoning (= inference). How is it that this inferential task for the hearer manages to be successful so much of the time, given that there are many possible ways of interpreting an utterance, direct, indirect, metaphorical, ironic, etc.? What is the criterion which enables people to choose the right interpretation?

3.1 *Grice's cooperative principle and the conversational maxims*

According to Grice who was the pioneer of the inferential approach to conversation (Grice 1975), there is a general assumption underpinning all utterance interpretation that the interpretation of utterances is a collaborative enterprise guided by a "co-operative principle" in which a speaker and hearer are engaged in some shared goal. This collaborative enterprise is structured by a number of maxims, which speakers are presumed to obey, amongst which Grice isolated:

- *The maxim of quality*: do not say that for which you lack evidence; do not say what you believe to be false.
- *The maxim of relevance*: be relevant.
- *The maxim of quantity*: make your contribution as informative as is required, but not more so.

- *The maxim of manner*: be perspicuous (avoid obscurity, avoid ambiguity, be brief, be orderly).

These are not rules that dictate behavior but, rather, maxims underpinning collaborative exchange of information. Take for example marking an exam, which in England involves two examiners. I suggest to you marks for each candidate, you disagree, and we then negotiate an agreed mark. Such collaborative endeavors are said to be directed by the cooperative principle and its maxims. Of course, it is not always the case that people do tell the truth, or are relevant. And I might get fed up with agreeing exam marks and walk away. But liars and people who refuse to cooperate are in some sense the exception that proves the rule; for in order for a lie to be successful, some presumption of the maxim of quality or its equivalent has to be in force, and people who refuse to cooperate are not engaged in any act of communication. More interestingly, the maxims are sometimes openly violated; and then they provide the trigger to a chain of reasoning which the hearer will use to reach an interpretation which the speaker intended to convey indirectly and which enables the cooperative principle to be seen to be reinstated. The conversation in (1) is an example of this. In (1), the manifestly irrelevant answer by B acts as a trigger for A to construct additional premises so that she will be led to see by indirect implication that B was trying to communicate something which is in accordance with the maxims. All such additional pieces of information, whether premise or conclusion, are said to be *conversational implicatures*. So B's answer in (1), which taken on its own is an answer that is either irrelevant or manifestly too little by way of answer, implicates that people who only know how to put a kettle on do not know how to cook, and that B does not know how to cook.

All such implicatures are derived by reasoning, and they are said to be acts of what is called nondemonstrative (i.e. nontrivial) reasoning in the face of some apparent clash with one or more of the maxims. Essential to the concept of implicature is that, unlike the intrinsic meaning of an expression, these implicatures can be "canceled" – hence their status as the result of reasoning, and not as the result of a linguistic rule. So there is nothing inconsistent with B's adding "Though I don't mean to imply that I can't cook. I can, I'm just not very interested in cooking." If, to the contrary, some aspect of interpretation cannot be consistently denied, then by definition it must be part of the meaning of the expression and not an implicature.⁸

It is this method of retrieving interpretation through a process of reasoning in the face of an apparent violation of the maxims which lies at the heart of the Gricean account of conversation (see Neale 1996). Take for example, the much-treasured compliment of being metaphorically described as a "real race-horse" in (2). This too is in blatant violation of the maxims of quality and relevance, and, in like manner, was taken by me in the situation in which it was uttered to implicate the assumptions that race-horses are extremely swift, and are exciting to watch, and that as something described as a race-horse, I was

extremely swift, and exciting to watch and hence to be with. Notice that, said with a less admiring tone of voice, what was said could equally well have been taken, much less nicely, to mean that I was highly strung, bad tempered and easily upset. Like (1), (2) can be followed up by an explicit cancellation of at least some of these implicatures:

- (9) Though I don't mean to imply that you're bad tempered or anything like that – it's just that you're exciting to be with.

Such implicatures, which in the case of (2) gave rise to its interpretation as a compliment, are said to be derived through a process of reasoning which starts from the premise that the speaker is intending to obey the general tenor of the cooperative principle, but their utterance is in transparent violation of it.⁹

The cases so far considered are clearly occasion specific; and it isn't any rule which licenses their interpretation. However, there are also cases which, though they can be construed as consequences of the maxims, are so regular that it is tempting to see their interpretation as the consequence of some kind of pragmatic rule, contrary to the general Gricean spirit. These are examples such as

- (10) Some people there were miserable

(10) would normally be taken to imply "Not everyone there was miserable," but it is cancelable as in (11):

- (11) Some people there were miserable. Indeed everyone was, though some were showing it less than others.

Since it is cancelable, according to the criterion defining an implicature, it is not an encoded principle of the grammar. Grice labeled implicatures such as these *generalized conversational implicatures*, but others since then have given them a rule-based characterization, suggesting that the concept of a grammar of a language might be extended by a pragmatic component which contains a set of default rules (see Gazdar 1979, Levinson 1983, 1987, 1996, forthcoming, and Lascarides and Asher 1993 for a concept of the *default inference rule*).

One primary difficulty with these maxims of Grice's is that they are often not clear, and any single implicature can be reasoned to be a consequence of a number of maxims. Is for example, B's answer in (1) to be construed as a violation of the principle of relevance, or of quantity? Has B said too much, or not enough, or merely something irrelevant? Any one of these could be taken as triggers to the chain of reasoning that leads to the intended interpretation that B cannot cook. Then there is the problem of what it means for some utterance to be relevant, a question to which Grice provides no answer. The result is that, though suggestive, the content of the maxims remains extremely vague.

A further problem is: What should the maxims be taken to explain? Grice articulated the maxims as a means of simplifying the overall account of the

relation between the use of language in logical arguments and the conversational use of language, arguing that recognition of maxims of conversational behavior could be used to explain a well-known problem – the apparent mismatch between how words such as *and*, *or*, *not*, *if-then* are used in logic and their use in ordinary language. In logic, which is the formal study of reasoning, the focus is on arguments that are valid in virtue of structure, for example that displayed in the English sequence:

- (12) If Bill is married to Mary and Mary is a Professor of English, then Bill is married to a Professor of English. Bill is married to Mary, but Mary isn't a Professor of English. Therefore Bill isn't married to a Professor of English.

Displaying the validity of argumentation in a sequence such as (12) involves defining *and* in such cases as combining two statements to yield a further statement of the form P and Q which is true if and only if the first statement and the second statement Q are both true. *And*, that is to say, has a purely cumulative effect. This use of the word *and* is unlike its characteristic use in conversational sequences, where it is often associated with sequencing in time. (13) for example does not merely imply that Bill was sick at some time in the past and that Bill went to bed some time in the past – it implies that he went to bed after he was sick, a sequence of events reversed in (14):

- (13) Bill was sick and went to bed.

- (14) Bill went to bed and was sick.

Examples such as these were used in the 1950s to 1970s (Strawson 1952, Cohen 1971, Walker 1975) to demonstrate the difference in content between natural language expressions and elements of logical languages, with the logical concepts being defined in terms of conditions necessary for the truth of a given element (e.g. *and* as in (12)). Grice however argued (1975) that the difference between *and* as in (12) and as in (13)–(14) was merely a difference in the implicatures that can arise as a result of the cooperative enterprise underlying the process of communication. This insight of Grice's was taken as a major advance, because it enabled natural language content to be defined in terms familiar from logic, and therefore taken to be better understood. Following this methodology, the meaning of sentences was defined in terms of truth-relations, the meaning of *It's raining* being said to be given by the set of conditions which have to hold in order for the sentence to be true. This gave rise (Gazdar 1979) to the slogan:

Pragmatics = meaning – truth conditions

Pragmatics was seen as the heterogeneous remainder left over once the account of descriptive content in the form of truth conditions is articulated as the basis

of semantics – this left-over being explained in terms of maxims of behavior which provide the trigger for the “commonsense reasoning” process that adds to (or replaces) the stricter descriptive content of an uttered sentence. Notice how the slogan itself leaves open the question of whether these maxims should be characterized as constraining an inferential task (as Grice advocated – Grice 1975), or as default rules as some of his followers have assumed (see Levinson 1988, forthcoming).

As indicated earlier, the Gricean program is attractive to linguists as it buttresses the view that there is a clear separation between grammar-internal processes, which characterize sentence structures, and arguably also a specification of their meanings, and the interpretation of utterances (see Levinson 1988, Atlas 1989). It is attractive to semanticists also for the same reason: it allows concepts of truth-conditional semantics familiar from the study of logic to be extended to sentences, defining the domain of pragmatics as a form of explanation that takes as input such specification of sentence-meanings (Kamp 1979).

There are however serious difficulties with this view. The first is most obviously displayed by pronouns and other anaphoric expressions.¹⁰ In order to establish the truth-conditional content expressed by a sentence containing a pronoun, some choice as to how the pronoun is to be interpreted has to be assumed; and these choices are not given as part of the grammar of a language in *any* sense – they depend on the interpretation of the sentence as understood in a context. The conditions under which (15) is true are, for example, quite different depending on whether the pronoun *she* is taken to refer to Princess Diana’s body, or to each individual woman being talked about:

(15) [*uttered in follow-up to (5)*]: Every woman cried as she went past the gates.

If *she* is construed as picking out Princess Diana’s body, it picks out a fixed object: if *she* is construed as ranging over the same set of individuals as is picked out by *every woman*, it picks out a set of individuals, each one in turn. The two circumstances described are very different. Notice, too, the way the interpretation of the pronoun varies may be sensitive to details of the particular event described. The funeral procession that morning started from Kensington Palace, where the Princess had lived, and that morning no one came through the gates until the beginning of the procession. If the speaker and hearer can both presume on this sort of knowledge, then in (16), *she* will be construed as the Princess’ body. However, in (17) *she* will be construed as each of the women being talked about, because it was bystanders who put flowers beside the gate (hardly the dead Princess!):

(16) Every woman cried as she came through the gates.

(17) Every woman cried as she put down her flowers.

The problem that truth-conditional content depends on specifics of a context is by no means restricted to anaphoric expressions such as pronouns. The women

being talked about in an utterance of (17) as follow-up to (5) – who are assumed to be picked out by the speaker’s utterance of the expression *every woman* – would be tightly restricted to the women described outside Kensington Palace as seen by B. It could not be every woman who went that day to put flowers down on some other grave elsewhere in the country, nor every woman at home watching the event on television. Even the predicate *put down* does not describe every event of putting down flowers that took place that day. Many people at the funeral event may have put down their flowers as they adjusted buttons on their coat, wiped a child’s nose, etc. etc. – let alone women right around the world. But *put down* here refers only to the act of putting down flowers outside the gates of the Palace. Indeed the interpretation of *any* of the words may in part be due to the context in which the sentence is understood. From this we are driven to conclude that grammar-internal principles do not determine full specifications of truth-conditional content but much less complete specification; for “commonsense reasoning” principles are also dictating what is expressed by a sentence as uttered in a context. Grice himself did not see the construal of pronouns and other anaphoric expressions as a major problem, and, within the Gricean concept of utterance interpretation, this was not seen as controlled by the maxims (see Carston 1998 for discussion). However, as we shall see, the very same criterion that determines how indirect implicature effects are recovered by the hearer also determines how all such context-dependent aspects of interpretation are chosen.

Finally, as Grice himself pointed out (1975, 1989), there are some aspects of meaning projected by words that are signally left out of any program that defines the meaning of natural language expressions in terms of truth-conditions. These are aspects of meaning which do not have anything to do with properties of external objects that the word can be used to describe, but, rather, have to do with the mode of reasoning about such objects that the word triggers. Take the word *but*, used in the previous sentence. The word *but* is used to indicate some form of contrast, but this is not a contrast of content intrinsic to the entities described. We know that this is so, because of examples such as (18) and (19) where explicit identity of the predicate is asserted in the two clauses joined by *but*:

- (18) John is applying for the Liverpool job, but so am I; and I have more publications than he does.
- (19) John got 70 percent but so did his brother, so neither of them could boast about having done better.

The truth-conditions contributed by the use of *but* cannot be distinguished from that of *and*: a statement formed by joining together two statements by *but* is true if and only if the two statements are true, exactly as in the case of *and*. Any characterization of word meaning merely in these terms will miss the idiosyncratic contrastive flavor intrinsic to the meaning of *but* altogether. Yet

the phenomenon is not a conversational implicature either, as it is an invariant aspect of sentences conjoined by *but* that some form of contrast is intended to be recovered. The Gricean program has to allow for special stipulations for aspects of meaning such as these which fall outside both the truth-conditional program, and the implicature form of explanation, as Grice himself pointed out (Grice 1975). He called these phenomena “conventional implicatures” to indicate that they were not regular implications of descriptive content but nonetheless part of the conventional meaning of the word in question; but this term was little more than a classificatory label.

3.2 *Relevance theory*

These various challenges to the Gricean program were taken up in Relevance Theory (Sperber and Wilson 1995). This theory claims to characterize pragmatic phenomena in terms of a single cognitive concept, that of relevance, replacing the social underpinnings of Grice’s cooperative principle.

3.2.1 *The principle of relevance*

According to Sperber and Wilson, there is one overriding constraint on human cognition, the principle of relevance. All signals are said to be processed relative to an ever-evolving background context with the aim of retrieving more information from the signal than it itself presents; and optimal relevance is getting the right balance between size and type of context and amount of information derived for the task at hand. So in our Princess Diana example, we might say that the interpretation of the word *it* in (5) picks out the funeral because this has just been mentioned in the context and so costs no effort to recover, and that referring to Kensington Gardens and the smell of the flowers gives rise to a whole chain of impressions and so is highly informative in talking about the funeral. Hence the relevance of each succeeding sentence in the exchange in (5).

This trade-off between cognitive effort and cognitive effect is at the heart of the concept of relevance itself. Humans are said to always subconsciously balance the amount of effort a task should have against the benefits to be gained from it – interpreting incoming signals to get as much information from them as is possible relative to putting in the minimum effort necessary to achieve that effect. This balancing of effort and inferential effect is the constraint of maximizing relevance. The more information some stimulus yields, the more relevant it is said to become, but the more effort the interpretation of that stimulus requires, the less relevant it will become. And to be minimally relevant a stimulus must lead to at least one non-trivial inference being derived.¹¹

To take in more detail a nonlinguistic example first, imagine yourself trying to write a letter in reply to some job advertisement which needs to be got off

today if you are to have any hope of being treated as a serious candidate. Imagine also that it is raining. To you, the information that it is raining is not relevant because you are trying to work out how best to present yourself and your achievements. This can be characterized through considerations of amount of effort required for the inferential effects to be achieved. What you are currently focussing on – the context against the background of which you are constructing your application – are premises which concern this particular act of writing: “I must remember to mention my degree results” “I must remember not to use too many adjectives,” “If I use the word *impressive* too often, they will think I am boasting.” And so on. What the weather is like does not impinge on this activity, at least not just at the moment. The effort of retrieving and manipulating information about umbrellas, whether to go by train, car or bus, is not warranted relative to your current worries, for none of these premises will combine with premises about how best to communicate what an impressive individual you are. When, however, it comes time to go out of the house, then there will be decisions to be made, and these involve reasoning with premisses about the weather and the nature of the journey to be made; and the information that it is raining will combine with these to yield appropriate inferential benefits (for example, if it were me I would be thinking things like “It’s better to go by train and read, since the rain is likely to mean that going by car will be slow”). The cognitive effort of drawing such inferences at this later point in time is suitably rewarded, in the sense that noticing that it’s raining combines with other things that are then on your mind anyway. According to this approach to understanding, the interpretation of a stimulus is defined as the manipulation of additional information relative to which a nontrivial set of inferential effects are achieved. Interpretation of signals of all sorts on this view invariably takes place relative to a context; because context is defined as the premises selected (the extra information) which ensure the relevance of a signal. There is no concept of a null context.

This example only involved one person, working away on their own, not being distracted by the rain. However interpretation of an act of communication involves two agents – the speaker and the hearer. The constraint of balancing cognitive effect (the drawing of nontrivial inferences) with cognitive effort will also apply to what the hearer does, but here the task of interpretation is more specific because the hearer has to try and recover what the speaker intended to convey. There are two aspects to the task:

- 1 Decoding the information intrinsically associated with an uttered expression – i.e. working out what words have been said and the information that they by definition carry.
- 2 Making choices which enrich that encoded information to establish what the speaker had intended to convey using those words.

To succeed in the first task, one has to know what the words of the language mean – what information they encode. This is the starting point from

which the proposition expressed has to be recovered. To succeed in the second task means establishing (a) some proposition corresponding to the intended interpretation of the utterance, (b) additional propositions which establish the required inferential effect. These propositions are, however, not just those which happen to be maximally relevant to the hearer: they must at least include those which the speaker could have intended.

And this is where a somewhat different principle of relevance – the communicative principle of relevance – comes in. A presumption of optimal relevance, according to Sperber and Wilson, determines how the hearer succeeds in arriving not merely at some most informative interpretation, but at the interpretation which the speaker intended. The context against which these decisions are made is said to be the set of representations retrievable with least effort that establish requisite inferential effects (this is what corresponds to “optimal” relevance).¹² This set may be taken as containing just the immediately previously constructed proposition: and, indeed, direct answers to a question do combine with such a context, viz. the question itself, for which they provide an answer. However, the context selected could also be some extension of the minimal context, as long as the extra effort required is offset by additional inferential effects; and this is what is triggered in indirect replies to questions. For example in interpreting the uttered sentence *I went to Kensington Gardens* in response to *Did you see the funeral?* in (5), the proposition taken to be expressed will be “B went to Kensington Gardens.” This indirect response, despite the increased effort required to process it, would have the advantage of triggering an extension of the context to include a premise such as “Kensington Gardens was the starting point for the procession preceding the funeral” from which A will deduce that “B saw the funeral procession by going to it” and probably also that she saw the service as well on the huge screens that were made available so that everyone outside the abbey could watch it. The benefit of the indirect answer here, according to relevance theory, is that for a minimal increase in effort of processing the given input, the hearer A is recovering extra information which she would not otherwise have got – here, the much richer stock of information that B did not see the funeral by watching it on tv, that B was part of a historic event, that B has much more information about some aspects of the event than she does, that B will not have seen other parts of the event, etc. Hence B’s answer in (5), though indirect, and causing A more effort in parsing and constructing an interpretation, is optimal in guiding the hearer to the requisite range of inferential effects. The very indirectness of B’s answer indeed is intentional, allowing an open-endedness in the interpretation since the choice of context is not fully determined.

3.2.2 *Relevance and the recovery of what is “said”*

One immediate advantage of the relevance-theoretic approach over the Gricean one is that the explanation of how implicatures intended by the speaker are worked out applies equally well in explaining how the proposition the speaker

has expressed is arrived at. So, for example, (16)–(17) uttered in the context of (5) are predicted to give rise to different interpretations. Choice of *she* as the variable bound by *every woman* in processing (16) would not be possible because it would combine with the information that no woman went through the gates other than the dead woman in her coffin, giving rise to immediate inconsistency, and no further nontrivial inferences.¹³ All such interpretations are therefore ruled out as not relevant. Hence *she* can only be used in that context to refer to the individual inside the coffin. Similarly, though to reverse effect in (17), *she* cannot be construed as Princess Diana since neither dead people nor their coffins can lay flowers. In both cases, the only available choice of representation is the one that is selected – a representation which meets the criterion of giving rise to a consistent set of inferential effects without undue cognitive effort. Exactly the same constraint dictates the construal of *put down* as “put down outside the gates of Kensington Palace,” though this time for a different reason. It is perfectly possible that an individual can cry as she puts down her flowers to wipe a child’s nose or do up her coat, but in this case, the type of premise required is one that combines with an assertion about everyone there. Information about women except in relation to events concerning the activities involved in the mourning at Kensington Palace has not been made salient and so is not easily recoverable. In contrast, the selection of *put down* as “put down beside the Kensington Palace Gates” in that same context naturally triggers such easily available premises as:

Crying is an explicit gesture of mourning.
Putting down flowers by Kensington Palace was an explicit act of mourning.

Such a choice of context would give rise to the inferential effect

Every woman who put down flowers by Kensington Palace did at least two things as an explicit gesture of mourning.

Since such a set of premises is easily recoverable, the interpretation of *put down* as “put down by Kensington Palace” is the concept the speaker intended to convey.¹⁴ Similar relevance considerations dictate that *every woman* picks out women standing by as mourners rather than, for example, women watching the same event at home on tv.¹⁵

So the principle of balancing cognitive effort and inferential effect can be seen to underpin both the deduction of so-called implicatures and the fixing of context-dependent aspects of the proposition expressed. It has the advantage also of not requiring the explanation of additional pragmatic effects to be triggered only in the event of apparent violation of a conversational maxim, as does the Gricean account. Indeed it purports to explain why deduction of additional information is an unvarying consequence of interpreting an utterance, and not merely a feature of exceptional apparently anomalous conversational

exchanges. Moreover it provides a natural distinction between implications which the hearer believes the speaker intended to convey (= *implicatures*), and those which she recovers from the utterance despite knowing that the speaker could not have intended to convey them (= *contextual implications*) (Sperber and Wilson 1995, Carston 1988). The implications the hearer believes the speaker intended to convey are those dictated by the criterion of optimal relevance (with minimal cognitive effort as the overriding factor). Those which the speaker need not have intended are the result of the less restrictive criterion of maximizing relevance. These will often add very considerably to the relevance of the overall utterance for the hearer, though they cannot be taken to be part of what the speaker has intended to convey. The fact that B went to the funeral for example will provide A with information about B, that B got completely caught up in the fervor that swept the country that weekend, that therefore she is probably a traditionalist, and so on – none of which B would have explicitly intended to convey in her reply. Such implications, by the way, would not be characterized by a Gricean system as part of the utterance interpretation process at all.

3.2.3 *Relevance and speech acts*

We have so far assumed that information retrieved from an utterance is solely about the object referred to by a speaker. However humans fluently reason at both the level of what is communicated, and at the level of how / why something is communicated. So A in our first conversation (1) will not only retrieve the information that B knows how to put a kettle on, but also that B wants A to believe that she only knows how to put a kettle on. Such higher level information is partly encoded – the differences between assertions, imperatives, and questions, in particular, rely on our ability to retrieve such higher level implications. An assertion implies that the speaker, if taken sincerely, believes the proposition his utterance expresses is true. An imperative is a request by the speaker that the hearer make some proposition true. A yes–no question is a request to the hearer to indicate to the speaker whether some proposition is believed by the hearer to be true. These higher level *explicatures* (as they are labeled in relevance theory (see Carston 1988) have been studied as part of semantics under speech act theory (initiated by the philosopher J. L. Austin in Austin 1962) and, following him, Searle and Bach and Harnish (Searle 1969, Bach and Harnish 1979). In this earlier theory (which predated Gricean pragmatics and relevance theory), language use was described in terms of a range of *speech acts* in terms of which the meaning of natural language expressions was explained.¹⁶ There are many cases where we do much more with words than merely describing objects and activities around us. So, for example, when a minister holds a baby over a font, sprinkles a few drops onto his head, and says the appropriate words from the baptism service, he is not merely saying these words, he is carrying out the act of baptizing. And when I say “I promise you that I will send you a letter tomorrow,” I am not merely

saying those words – I am carrying out an act of promising. On the speech act view of language, language can best be understood in terms of acts such as these which speakers carry out in using language. The observation by speech act theorists that there is clearly more to language than just describing things is quite uncontroversial – no one working in pragmatics doubts this. Nonetheless, in relevance theory, where the type of implications that can be drawn is quite unrestricted, there is no need of any special discrete categories for such different kinds of act. All these implications would fall under the category of *explicature*, and as part of the proposition expressed would come within the general umbrella of information retrievable from an utterance, for which the hearer's task is to recover those implications that the speaker intended to convey. Like all other implications, whether or not they are retrieved depends on their relevance to the participants. It may be relevant to construct the explicature that the speaker believes that P, for some arbitrary P, though in ironical utterances, this is contra-indicated. Questions normally require an answer, but so-called rhetorical questions signally do not. And, equally, it may be relevant that someone we speak to has specifically promised that P. Even in the highly conventionalized case of baptism, the implication that a child has been baptized through some specific act in a religious service is potentially relevant in just the same way as every other act of communication – it needs no special pragmatic category to explain its communicative effect – merely an explanation of the role of the priest within a given religious ritual, and the significance of baptism within a set of religious practices. I shall not have anything more to say about such “speech acts” in the remainder of this chapter, but they nonetheless have an integral role in establishing the relevance of an utterance, (see Wilson and Sperber 1988a, Carston 1988).

3.2.4 *Procedural aspects of interpretation*

The assumption that interpretation involves constructing both some distinguished proposition and some context set of propositions to combine with it provides a natural basis for explaining the conventional implicature phenomena problematic for Gricean approaches to meaning. Given the two-fold nature of the inferential task triggered by natural language input, it is entirely natural that the content of some words might be directed more towards constraining the context set of premises to be selected, rather than in establishing the proposition expressed (see Blakemore 1987, 1992). In this light, we can view connectives such as *but* as constraining the context relative to which the sentences it connects are to be construed, establishing both one form of context for the first conjunct, and a guaranteed modification of it in adding the second conjunct. Seen in these terms, *but* has to be defined as a procedure for context-construction, imposing a choice of context for the first conjunct which must lead to a conclusion, relative to which the context for the second conjunct (which will automatically contain the first conjunct) must yield a contradiction (Blakemore 1989).

4 The Interaction between Linguistic Processing and General Processing

This concept of procedures for interpretation has much more general application (see Wilson and Sperber 1993, Kempson 1996). In some sense words – all of them – are the procedures we have to start the process of interpretation. Consider first pronouns. Pronouns do not fix the way the hearer understands them; they merely guide the utterance process. This guidance takes the form of a constraint. Consider (20)–(21):

(20) The Queen frightened her.

(21) The Queen frightened herself.

(21) must mean that the Queen frightened herself, rather than anybody else. (20) however cannot mean this – it must mean that it is someone other than herself that the Queen frightened. More formally the pronoun provides a place-holding device for which the hearer has to choose some representation, relative to the constraint that the individual represented be female *plus* the constraint that the representation the hearer selects must not be one that is already set up within the propositional structure being constructed. It is for this reason that (20) and (21) cannot be taken to mean the same. This “locality” restriction is generally seen as a syntactic restriction defined over syntactic structure, separating some aspects of pronoun interpretation off from their use in discourse as grammar-internal and subject to syntactic explication (see chapter 11, and Chomsky 1981); but, looked at from a more procedural perspective, we can see the pronoun as encoding a constraint on the building up of a propositional structure (see Kempson 1988a, 1988b, Kempson et al. in preparation) – a procedure which guides the hearer but leaves open what it is that the pronoun actually picks out on any occasion of use.

Notice how, with this concept of word meaning as a set of procedures for interpretation, it is essential that interpretation be defined in terms of structured representations of content, and the progressive building up of such representations: locality conditions have to be defined over a level of representation, and this can be construed as an encoded constraint on pragmatic interpretation *only* if the process of interpretation is also taken to be defined in terms of structure. So far, however, the concept of structure over which anaphoric interpretation is defined could be a structurally complete configuration of a propositional structure, missing only some values to place-holding variables (as presumed in relevance theory – see section 1). However, remember how the interpretations of *the smell*, and *the whole day* in (5), *woman* and *put down* in (17) also depended on context, displaying a similar gap between the characterization of the word independent of context and its particular interpretation in a given utterance. This shows that it is not just pronouns whose interpretation

depends on context, but all words. The very generality of this phenomenon of context-dependence suggests that all specifications of words should be viewed as constraints, underspecifying interpretation, on the basis of which a hearer builds some propositional structure.¹⁷

4.1 Metaphor as an enrichment process?

There are two ways in which this concept of underspecification might be extended. On the one hand, current work on the construal of metaphor in relevance-theoretic terms is suggesting that an enrichment approach to interpretation should be extended to metaphorical uses of language, the word being but a trigger for the online construction of the concept specific to that interpretation (Carston 1998). There is an important background debate behind this suggestion. There has been disagreement within pragmatics to date over whether there should be any independent maxim of quality constraining people to tell the truth (Wilson 1995). The maxim of quality contains the heart of the social principle of cooperation intrinsic to Grice's theory of conversation. On the Gricean account, remember, the interpretation of metaphor involves transparent violation of the maxim of quality with the false proposition expressed by the sentence as its literal meaning having to be replaced by some quite different proposition (see Sperber and Wilson 1981 for a critical evaluation of the Gricean account). On the Sperber and Wilson account, which is *not* a social theory of communication, metaphorical interpretations involve a different relation between a proposition and the thought which the hearer recovers from the sentence as uttered in context. Just as drawings, such as cartoons, may "resemble" a person without depicting them at all accurately, so propositions can be used to resemble thoughts they convey, a use of language which is called "interpretive use." Metaphorical uses of language, like other rhetorical effects, are said to constitute an interpretive use, with the sentences being used to convey a relationship of resemblance between the proposition expressed by the utterance and the thought(s) it is intended to convey Sperber and Wilson (1995), Wilson and Sperber (1988b). On this view no maxim of quality is required – it is merely the criterion of presumed optimal relevance which, as elsewhere, constrains the interpretation process.

The view that all word meaning is but a set of procedures provides a new shift in this debate; for it suggests that any word is but the input to the construction of some novel "ad hoc" concept specific to that utterance. On this view interpretation of a sentence such as (22):

(22) She cornered him.

involves constructing a concept on the basis of the presentation of the word *corner* in such a way as to yield extra information about the individuals picked out by the pronouns in that context, in ways which make the whole utterance

relevant to the hearer. So *corner* has to be understood as a relation between individuals, something someone can do to someone else. The construction of the new concept is direct and not via a process of rejecting a literal interpretation. No maxim of quality needs to be invoked in this account of metaphor, but, equally, no indirect concept of resemblance is invoked either: it is, rather, that all that a word provides is a set of procedures relative to which a hearer establishes a relevant interpretation, constructing new concepts online from the presented word. The consequence is that it is not merely the pronouns whose encoded specification underdetermines their interpretation, but words that express concepts also.

4.2 *Syntax as the building of propositional structure*

The idea of expressing the content of a word as a procedure can apply in syntax too, for the projection of structure from the lexicon can also be defined as sets of procedures for building up propositional structure (Kempson 1996, Kempson and Gabbay 1998).¹⁸ To take relatively straightforward cases first, names in a language can be defined as the projection of a representation uniquely picking out some individual. Intransitive verbs which involve the assertion of a property attributed to (“predicated of”) some individual, can be defined as introducing a structure in which the property in question is represented at one node in a structure with another node also introduced to represent the subject of whom the property is predicated. Transitive verbs, such as *hit*, might be said to introduce a structure containing a position waiting to be filled by an object and a subject – notice how as soon as the hearer processes the word *hit*, she knows that there must be some object of the action of “hitting.” Auxiliary verbs, such as *will*, *did*, can be defined as adding information about time to some propositional structure. And so on. In all cases, words can be seen not in terms of the structure of sentences in which the words occur as a string, but, rather, in terms of progressively building up a propositional structure (Kempson et al. forthcoming, Kempson et al. in preparation).

Though highly unorthodox as a basis for describing syntactic properties, there is reason to think this is the right direction in which to go. First, there are elliptical constructions. These are radically incomplete strings, which have to be enriched in context by being assigned a structure corresponding to that of some antecedent structure:

- (23) A: Have you handed in your assignment yet?
 B: Nobody has asked me to.

The puzzle presented by examples such as (23) is that fragment reconstruction involves syntactic and discourse properties simultaneously.¹⁹ Though the elliptical fragment is reconstructed as a process of interpretation, the output of such a building process displays structural properties exactly as does a sentence in

which the structure of this interpretation is explicitly introduced. But, nevertheless, the structure that is reconstructed is not made up of the words of the previous sentence, despite the fact that it relies on them to provide the interpretation of the incomplete expression *to* in B's reply: what is reconstructed is the structure of the proposition expressed by the words; and in (23) this is a structure corresponding to "Has B handed in B's assignment yet?" (*you / your* are pronouns which serve to pick out the hearer). We know this because if the speaker B uses words to mimic the effect which the fragment has more economically expressed in (23), she must choose different words to reflect that content directly, and not the words actually in the previous sentence – hence the shift from *your* to *my* in (24):²⁰

- (24) A: Have you handed in your assignment yet?
 B: Nobody has asked me to hand in my assignment yet.

By using a fragment, no such shift in the choice of words is necessary. In reconstructing interpretation as a structural process, the hearer will be building up a representation of the content expressed by the words (i.e. the proposition expressed), *not* a representation of structure defined over the sequence of words themselves. But this suggests that the syntax in terms of which this phenomenon has to be explained is the structure that underpins interpretation – not some structure displayed by the words themselves.

Second, it turns out that phenomena taken to be central to syntax can be revealingly characterized in terms of some rather weak starting point and its resolution in context, analogous to the way in which pronouns contribute to interpretation. Incompleteness of information is of course endemic to language processing, because, at least in speech, we process the input sequentially, progressively building up structure throughout the interpretation process.²¹ Imagine now the parsing of (25) – what syntacticians call a *topicalization* structure:

- (25) John she admired.

Suppose we decide to represent the proposition expressed by a sentence as a tree structure (as in syntax – see chapter 11, Syntax), so that we can display the individual parts of the interpretation, and how they combine together. This means that the output structure of parsing this sentence, when it is finally established, will take the form shown in figure 16.1.

The first question in wondering how a string such as (25) is parsed to achieve this result is: What information has the hearer got when she has processed the word *John* in (25) on the assumption that she decides that this occurrence of *John* is not to be understood as the subject? We take it she knows which individual named *John* is being talked about – this is what the speaker assumes in using this name – but she doesn't yet know how to construct the resulting proposition the speaker is trying to convey. In particular, she doesn't know where John is to fit in any such resulting structure. All she will know at this

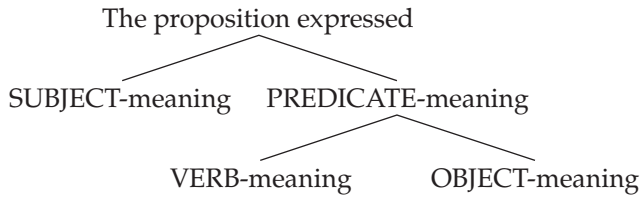


Figure 16.1

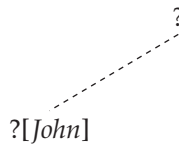


Figure 16.2

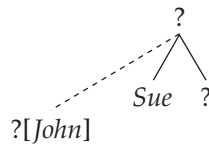


Figure 16.3

first stage is as in figure 16.2 (I put question marks above the parts of the structure which she knows must be there, but doesn't yet know how to fill out).

As the hearer then gets more information she is able to gradually build up the structure, at the same time identifying any items (such as entities referred to by pronouns). For example, in processing *she* (let us say identifying that it is Sue that is being talked about), the hearer may add to her initial tree description, to establish the structure as in figure 16.3.

At this second stage she knows *she* is the subject picking out Sue and that when she has established some property she can attribute to Sue she will have a completed proposition about Sue. Hence the second node with a ?.

With the word *admired*, she gets yet further information that a two-place relation of "admire" is asserted between Sue and some other individual, a relation said to hold in the past. So she can add to the structure so far built up and update it into the structure, figure 16.4.

With this as the full sequence of words, she can at last identify the position in the propositional structure into which the representation of "John" should fit – viz. as object of the relation "admire" asserted to hold between Sue and John (figure 16.5). Each word has by this stage led to a concept being entered at a fixed position in the tree: even the representation "John" has got its contribution in the propositional structure established.

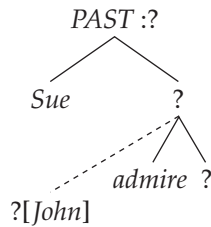


Figure 16.4

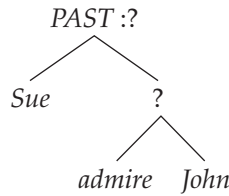


Figure 16.5

With the tree structure corresponding to the proposition expressed now completed, it follows that the hearer can establish what property it is that is asserted of Sue – it is “admire(John).” So now the hearer has established the full content of the proposition expressed, which represented as a tree structure, is figure 16.6:²²

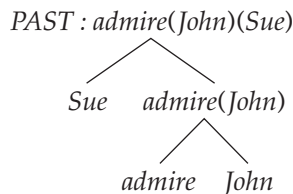


Figure 16.6

Notice what form the information presented by the words has to take in this step-by-step modeling of the interpretation process. What the hearer does is, simply, a progressive updating of an initially very partial tree description.²³ The words all provide sequences of actions, procedures, which enable the structure to be progressively built up. Structures like these, by the way – topicalization structures – have been the focus of much study as part of what linguists call syntax, where the metaphor, over many years, has been that the representation of *John* has moved from some original position to the position in the front of the string, leaving what is called a “gap” (see ch. 12). Looked at from a parsing perspective, the hearer, as we’ve just seen, does the reverse of this. She starts from very little information and gradually builds up a completed structure which this initial expression can be slotted into.

As confirmation of this processing perspective on syntax itself, we can now see that the way these partial trees are built up by a hearer may be affected by other processes of interpretation which we know to be pragmatic, an interaction which we would not expect if the projection of structure had to be characterized as quite separate from all pragmatic processes. The evidence comes from the interpretation of so-called relative clauses, and the way pronouns inside them can contribute to the building up of their structure.

First, by way of introduction, notice that languages have more than one kind of pronoun. There are not only the ordinary pronouns, *he*, *her*, *them* etc., but there are what traditional grammarians have called *relative pronouns* that introduce these relative clauses.²⁴ The function of these, rather crudely put, is to carry information from one piece of structure to another, enabling the hearer to build up two pieces of information at once. So in (26), through the use of the word *who*, the hearer is able to build up two pieces of information about *John* at the same time – that Sue says Tom admires him, and that he’s likely to get the job:

(26) John, who Sue says Tom admires, is likely to get the job.

Specifically what *who* does within this structure is to shift the hearer temporarily away from the task of building up the first structure to start the building up of the second. With this shift, *who* also conveys the information that a second occurrence of the expression “John” is to play a part in building up this added structure. In this respect *who* is like a pronoun, for it is interpreted by information got from elsewhere in the context in which it is contained. However, there are two respects in which *who* is not like a pronoun. First, unlike pronouns, *who* cannot be understood freely; on the contrary, *who* in (26) has to be understood as picking out *John*: this is fixed by a rule of the language. Second, the word *who* has to come at the beginning of the structure in which it occurs, and, like in topicalization structures, it doesn’t itself give any indication as to where in the tree this occurrence of “John” in the new structure should be placed.²⁵ In this respect, *who*, in this secondary structure, is behaving exactly like the word *John* in (25) – the hearer does not initially know where to fix it in the structure she is building up as interpretation, and the information where to slot it in comes later – after *admires*. In short, on the one hand, we have the word *who* in relatives behaving like a pronoun, and, on the other hand, we see that it behaves as though in a topicalization structure.

Now we are ready to see the interaction with what we know to be a pragmatic process – the process of establishing what a pronoun means in context. We need to focus specifically on the sequence of steps involved in the process of interpreting the remainder of the relative clause, but this time with a pronoun inside the relative clause itself:

(27) John, who Sue says she’s so worried about him that she’s taking him to the hospital, is begging her not to make such a fuss.

Sentences such as these occur on a daily basis, and we none of us have any difficulty processing them, though we may judge them less than perfect, because the pronoun isn't strictly necessary. Consider what happens in processing, as the hearer continues on from the point at which *who* is processed, this fixed as picking out "John." The task that lies ahead of her is how to decide what contribution this occurrence of "John" obtained from *who* makes in the interpretation of the relative. The question that arises is: can we assume that the pragmatic process of establishing an interpretation for a pronoun in such a context may feed into the process of working out how "John" is to contribute to the building up of this second structure? And the answer is "Yes, of course – what could possibly prevent it?" In (27), the pronoun occurs in a position where, *if* it is understood as picking out "John," it will establish how the representation of "John" is to be used in this second structure, which otherwise up to that point in the sequence of parsing actions the hearer can only identify as contributing somehow in the second structure, without knowing exactly how. Once the pronoun in the processing of (27) is construed as picking out John, then the hearer knows that John is the object of Sue's worry. But this result yields the pairing between the relative pronoun and the place in the structure to which it is to contribute which is definitive of topicalization structures.

This so-called resumptive use of the pronoun in (27) isn't strictly necessary in working out how *John* is to contribute to this second structure – after all (28) is just as acceptable, and its only difference from (27) is that it lacks this pronoun:

- (28) John, who Sue says she's so worried about that she's taking him to the hospital, is begging her not to make such a fuss.

In sentences such as (27), the speaker has slipped in a pronoun, presuming that the hearer will understand that it picks out the individual just described,²⁶ in so doing, providing the hearer an extra clue with which to work out how the relative is to be understood. In the case we've just seen, the use of the pronoun is optional, indeed possibly dispreferred (hence some linguists judge that such resumptive uses of pronouns are substandard, however widespread). But this is by no means always so. In (29), where the speaker is using the word *even* which forces her to emphasize the following word, the presence of this resumptive pronoun becomes obligatory:

- (29) My son, who sometimes even *he* gets fed up with me, is staying out tonight.

There is no other way to express this information within a relative clause like this, because you can't emphasize silence. Silence won't be said any louder. And without the pronoun, as in (30), the sentence means something quite different – it is the predicate "gets fed up with me" that is emphasized, not the expression "my son":²⁷

(30) My son, who sometimes even gets fed up with me, is staying out tonight.

The significance of these data is that they provide us with cases where a pragmatic process – that of working out what a pronoun is picking out – can be seen to contribute towards establishing a perfectly regular grammar-internal relation – which is the pairing between a fronted expression and the position from which it contributes to the interpretation as a whole.

One could of course argue in rejoinder that this shows nothing more than that in language use, structural and pragmatic phenomena may reasonably interact; and that this doesn't show that the characterization of internal properties of the language needs to be in these dynamic use-directed terms. However, as we shall now see, it does indeed provide evidence of this stronger conclusion, for it is only by making this move that we retain a unitary account of the pronouns.

What these data unquestionably show is that the availability of pronouns in the processing of relative clauses presents the speaker with a choice: either to use a pronoun, or not. The choice that then confronts the analyst is what consequences this should have for the articulation of the underlying structural properties of language in the form of a grammar. The first option is to define two alternative rules of grammar for projecting relative clause interpretations, with a characterization of the extent to which they differ. On this alternative, there is some grammar-internal process that dictates the pairing between a fronted expression and a pronoun in relative-clause and topicalization structures, and this process is by definition quite distinct from the general pragmatic process of anaphora. The alternative, on the other hand, is to define a single process of relative clause construal, that of pairing an unfixed node with a fixed position in the structure; and allow that the pragmatic process of working out what a pronoun means may contribute to the process of establishing this pairing.

How should we decide which of these two approaches is correct? Evidence that would tend to favor adopting the second alternative would be evidence showing that some general principle of communication was sufficient to explain the interaction between the syntactic and pragmatic processes. And this is what we find. The puzzle is that, most of the time, these two processes are not freely available. When it is acceptable not to use a pronoun, it is generally much less acceptable to use it. For example, when the structure is less complex than in (27) it becomes quite unacceptable to use a pronoun resumptively:

(31) ??John, who Sue's worried about him, is begging her not to make a fuss.

And when a pronoun is used fully acceptably, then it's often because leaving it out would not convey the same information – as in (29)–(30).

The presumption of optimal relevance here explains this distribution very straightforwardly. If the speaker has two equivalent means of expressing what is characterized by a single grammar-internal process, then the presumption of

optimal relevance should dictate that any choice available to the speaker which involves the hearer in processing additional lexical material for *no* additional inferential effect will always be less than optimal, unless this less economical strategy helped in what otherwise might be a potentially over-complex structure which risks causing the hearer undue processing difficulty. It is this processing account that explains the difference in acceptability between (27) and (31). (27) is, perhaps marginally, acceptable because the structure is somewhat complex: the pronoun helps in identifying the position from which the unfixed node is to be understood. The additional cost of processing the words is in this case just about offset by the contribution that this information makes to greater ease in retrieving the structure intended. In contrast, in (31), which is so much simpler, a pronoun cannot be used resumptively. Here there is no reason whatever to buttress the identification of the position into which the occurrence of "John" projected by the relative pronoun has to be fitted. Hence the unacceptability of the pronoun in this case – no speaker would judge the extra cost to the hearer warranted with such a simple example, no matter how little parsing effort would seem to be involved in processing just a pronoun.

Relevance theory can also explain the other type of case, where the resumptive use of a pronoun succeeds in conveying a particular kind of effect. For if the additional words in the more complex form of wording lead to a difference in pragmatic effect, then the additional effort for the hearer in processing the additional word(s) would be offset by the extra communicative effect achieved. And it is this property that explains why (29) is acceptable. In (29) the pronoun occurs with what is called a *focus* particle, *even*, which has a particular form of interpretation that demands the presence of the following pronoun – the interpretation is simply not derivable without the pronoun. Right across languages, indeed, resumptive pronouns are always acceptable where otherwise they would be debarred if the particular position is being emphasized by a word such as *even*, or even merely if the particular position is being stressed for some particular contrastive purpose (as in (32) drawing attention to the fact that the speaker believes that other people might well let her down):

- (32) John, who I'm certain *he* won't let me down, has said he'll come to my party.

This suggests that the correct conclusion to draw is that the grammar makes available a single principle underpinning such relative structures (shared with the process of topicalization), and that while the pragmatic process of anaphora itself remains free and controlled only pragmatically, it, nevertheless, feeds into the encoded process that drives the building up of interpretations for relative clauses. And though the system itself will freely allow use of the pronoun in such contexts, it will only be acceptable if either it leads to a reduction in processing costs or it enables additional communicative effects to be achieved. Now, the final step is that we can only characterize the problem in this way if the characterization of relative clauses has been defined in terms of the dynamics

of how interpretation is built up. So these data do indeed provide evidence that pragmatic principles may contribute to establishing those structures which are characterized as part of our knowledge of language, hence as part of what we call the grammar of the language. The grammar on this view defines how structures may be developed in the parsing process, and, within this, the words define procedures that drive this process.

Of course, this one indication of the fruitfulness of looking at the way structures are processed is far from conclusive. At best, it cannot be more than a taste of results to be established in detail,²⁸ but it is indicative of how, if we take the dynamics of building up interpretation as integral to describing both natural language content *and* its structure, then not only can we model the interpretation process in an intuitive and natural way, but we have the bonus of providing a revealing perspective on structural properties of natural language.

5 Summary

In this chapter, we started by surveying the apparently acute problems facing any general account of communication. We set out the different types of separation that have been imposed on the distinction between linguistic knowledge and language use; and I argued that though there is a distinction between encoded and nonencoded information, nevertheless, there needs to be full integration of the two types of processing. I then sketched the two primary different sets of criteria for choosing the intended interpretation – the Gricean route through conformity to maxims governing conversational behavior, and the relevance theory route through a choice driven by balancing effort and effect. I then presented some of the evidence that words should be seen as procedures for interpretation, rather than having some fixed descriptive content. And finally, I sketched out one way in which the interpretation of topicalization structures and pronouns can both be explained in terms of the incremental building up of tree structure corresponding to interpretation.

On the view that is emerging from this much more dynamic, use-oriented perspective on linguistic description, all words provide a set of procedures relative to which a hearer progressively constructs a structure corresponding to the proposition expressed. The encoded specifications intrinsic to language are defined explicitly as the driving force in this incremental process of building up interpretations from a natural language sequence of words. Linguistic and nonlinguistic processes of interpretation, nevertheless, freely interact in determining what proposition or set of propositions is expressed by a given input. The only externally imposed restriction is that linguistic input, being an encoded set of instructions on structure building, cannot be set aside. The linguistic knowledge that we have as users of the language is the encoded input which the individual language provides to enable the structural dynamics of

the interpretation process to take place. Pragmatic principles are the general cognitive principles that enable us to enrich information by general reasoning strategies, and to make choices between alternative structures as the interpretation is progressively established. This strongly suggests that modeling the process of communication itself provides a basis for explaining what it means to know a language (= competence – see chapter 12).

This view is a departure from the view that linguistic ability should be a body of knowledge which is quite independent of whatever principles determine how language is used (= performance). On this view, natural language ability is, rather, a capacity for natural language parsing. The current focus of debate in deciding on these views is barely begun, but at the center, as it always has been, is the status of our capacity for language. Is it a static store of knowledge relative to which pragmatic principles of use determine the apparently conflicting uses to which it can be put? Or is it the possession of a capacity for dynamically projecting structures which correspond to interpretation for a given piece of language uttered? At the present time, these remain questions over which researchers struggle to reach the most revealing answer; and much of what I have introduced in the last section of this chapter would be fiercely contested by many. One thing we can be sure of, however, is that study of the way people use language is a central preoccupation of linguistic study.²⁹

NOTES

- 1 This was once said to me as a fervent (and much appreciated) compliment!
- 2 What a user of a language knows about her language is called a *grammar*. In a grammar there are facts about the sounds used to build words (*phonology*), facts about how words are arranged to form sentences (*syntax*), and facts about those aspects of the meaning of words / sentences that are integral properties of those words / sentences (*semantics*) (see chapter 8, Phonology; chapter 11, Syntax; chapter 15, Semantics).
- 3 A very useful collection which provides readings in all the major topics in pragmatics is Davis 1991.
- 4 Sentences in sequence are often used, contrarily, to imply events in sequence:
 - (i) I finished my Ph.D. I got it published. And I became pregnant.
- 5 It has long been recognized by both linguists and philosophers that pronouns such as *I* and *you* can only be interpreted relative to entities in the discourse context. In this connection, cf. the papers of Roman Jakobson writing in the 1950s, who referred to them as *shifters* (in Waugh and Halle (eds) 1984). In the philosophical literature, such pronouns are called *indexical pronouns*, cf. Perry 1979, 1993. However the recognition of the extent to which interpretation is dependent on context was not widely recognized until the early 1980s with the work of Kamp 1984, Barwise and Perry 1983, Sperber and Wilson 1986.

- 6 An important assumption here is that all cognitive activity involves the construction or manipulation of internal cognitive representations. We can, for example, only see that rose bush in the garden in virtue of setting up some internal representation of that rose bush (we don't have rose bushes in our heads – not even those of us who are obsessed with rose gardening). This view is known as representationalism and in its current form is largely due to Fodor, who has argued for a so-called language of thought (cf. Fodor 1975, 1981, 1983, 1998). What is controversial for linguistics is the relation of such representations to representations in linguistic description. There are substantial differences of opinion for example over whether any level of representation is required in modeling interpretation in language other than that articulated within syntax. This is a debate which has rumbled on in different forms for at least thirty years, cf. Katz 1972, Lewis 1972, Kamp 1984, Groenendijk and Stokhof 1991, Sperber and Wilson 1995, Kamp and Reyle 1993, Dekker 1996, Kamp 1996, Kempson et al. 1997, Carston 1998.
- 7 A proposition needs to be complete only to the extent that it is one for which inference can be defined. So for example, absolute precision as to time is generally not required (I would bet that anyone reading this chapter has forgotten the exact date of Princess Diana's funeral). Given the nature of the conversation, suggesting a recent event, I have recorded the time variable as being "Saturday." The date was in fact Saturday 6 September 1997.
- 8 Contrast (1) with:
- (i) I never cook anything but I make an omelette at the weekends.
- 9 The characterization of metaphor through the supposed violations of the maxim of quality is not entirely unproblematic. See Sperber and Wilson 1982 for a critique of Grice's theory of conversation and Wilson 1995 for a critique of the problems imposed by the maxim of quality characterization of metaphor. See also section 4.1.
- 10 An anaphoric expression is one which can only be understood by reference to some other representation made available during the interpretation process. Pronouns are the central type of case, but words as *the*, *this*, *that* are also anaphoric.
- 11 The qualification of "nontrivial" is important as from a purely formal perspective "P" implies "P and P," "P and P and P . . ."; P also implies "P or Q" for arbitrary Q, etc. A nontrivial inference is one which cannot be drawn from considerations of P alone.
- 12 The modification of maximal relevance to optimal relevance for the particular activity of utterance interpretation is because of the imposed task of recovering the speaker's intentions. Some utterance may be extremely relevant in terms of the hearer's own privately held assumptions but nonetheless not be the intended interpretation because the speaker couldn't possibly have had access to the assumptions that led to these implications. Paranoia is a good example of this, where one's private fears and anxieties are so easily retrievable that they constantly lead to inferential effects of one sort or another, but fortunately most of us, at least some of the time, realize that these private worries are not a reliable basis for recovering what speakers have intended to convey to us.
- 13 In logic, one is taught that from an inconsistency any proposition can be derived, so there is a trivial inference from "P and not-P" to Q for arbitrary

- Q. As with the earlier trivial inferences, these have to be explicitly debarred.
- 14 There is no fixed choice of premisses which must constitute the context, so any choice of premisses will do as long as it licenses inferential effects associated with being at Princess Diana's funeral.
 - 15 There are several alternative accounts of the way in which construal of language depends on the context in which it is uttered. In some of these, there is no commitment to any form of representation. Amongst these is Searle, who argues for a concept of *Background* relative to which language is interpreted, without any commitment to mental representations. Cf. Searle 1983, 1995.
 - 16 The Speech Act theory of meaning, which was originally articulated as a theory of meaning for natural language under the slogan "Meaning is Use," played an important part in the development of pragmatic theory. Language was explained in terms of different kinds of actions that can be carried out by the use of language. Primary distinctions were made between *locutionary acts* (what the speaker said to the hearer), *illocutionary acts* (what a speaker does by performing such locutionary acts – e.g. baptizing, promising, threatening), and *perlocutionary acts* (the effect a speaker has on the hearer – e.g. persuading, frightening).
 - 17 This view has been consistently championed by Atlas in connection with negation (cf. Atlas 1977, 1989).
 - 18 The *lexicon* is the part of a grammar where all information about individual words is stored. Entries in the *lexicon* are generally referred to as *lexical items* but the simpler notion of *word* is sufficient for our present purposes.
 - 19 The issue of ellipsis remains controversial. There are analyses of ellipsis which are purely syntactic (Fiengo and May 1994 and others), analyses which are purely semantic (Hardt 1993, Crouch 1995) or some mixture of the two (Kehler 1995, Lappin 1996). Cf. Kempson 1995, Kempson et al. 1999, for an account of ellipsis in terms of how interpretation is inferentially built up as a propositional structure. For a representative selection of views, cf. Lappin and Benmamoun 1999.
 - 20 This type of example is problematic for purely syntactic accounts of ellipsis, requiring a concept of "vehicle change" (cf. Fiengo and May 1994, Lappin 1996).
 - 21 Even in reading, the parsing process is broadly sequential, though some effects arising from effects of linearity are noticeably weaker in reading. For example, it is not possible to open a conversation with (i) and be confident that your hearer will interpret *his* as picking out the same individual as the following phrase *a friend of mine*:
 - (i) His mother attacked a friend of mine.

It's more likely to pick out, say, your partner, particularly if your hearer already knows you are worried about his mother's uncontrolled behavior. (Many people talk about people they live with through the use of pronouns without introducing them into the conversation by name first.) However a newspaper headline might well report this event in the form

 - (ii) His mother attacks young boy.

relying on the fact that short sentences can be visually processed with much less reliance on strict left-to-right processing.
 - 22 The order of expressions in the propositional structure, with that projected by the object preceding that projected by the subject, is to reflect the way the interpretation is built

up, the two-place predicate “admire” applying first to “John,” and then to “Sue.” This is a standard notation in formal semantics. See Heim and Kratzer 1998, for discussion of the concept of functional application underlying this notation.

23 This is a very simple sketch, and to give it proper formal characterization we shall need a tree-description language, a matter which I shall here simply presume on (cf. Kempson et al. 1999b, in preparation).

24 See Jespersen 1927.

25 *Who* is no longer restricted to being construed as nominative and picking out subjects, as (i) indicates:

(i) I dislike the man who John is thinking of living with.

26 Notice that this is just what the presumption of optimal relevance would

dictate – for the previous mention of “John” by using the name *John* is by far the most salient representation in this discourse context, hence recoverable with the least possible effort.

27 Indeed this is quite an awkward sentence, because it implies that it is somehow surprising or exceptional that my son should get fed up with me – contrary to most people’s experience!

28 See Kempson et al. forthcoming, Kempson et al. in preparation.

29 I have been helped in the writing of this chapter by detailed comments on earlier drafts from David Swinburne, Johannes Flieger, and Lutz Marten, and the two editors Mark Aronoff and Janie Rees-Miller. Without their help, this chapter might have been even more partial in its coverage than it no doubt remains.