# Pragmatics and time*

Deirdre Wilson and Dan Sperber
*University College London and CREA, Ecole Polytechnique*

## 1. Introduction

In interpreting utterances such as (1) and (2), the hearer generally treats the events described as temporally or causally related:

(1) a. I took out my key and opened the door.
b. John dropped the glass and it broke.
c. They planted an acorn and it grew.
d. Peter left and Mary got angry.

(2) a. I took out my key. I opened the door.
b. John dropped the glass. It broke.
c. They planted an acorn. It grew.
d. Peter left. Mary got angry.

Such relations are not encoded in the meanings of the sentences uttered. This paper is concerned with how they arise. We will look in particular at the following problems:

(a) *The sequencing problem.* Why does the hearer generally take the events to have happened in a certain order, so that in (1d), for example, he would assume that Peter left before Mary got angry?

(b) *The interval problem.* Why does the hearer generally take the events described to be separated by different intervals, so that in (1b), for example, he would assume that the glass broke as soon as it was dropped, whereas in (1c) he would not expect the acorn to have sprouted as soon as it touched the ground?

(c) *The cause-consequence problem.* Why does the hearer often take the events to stand in a causal or consequential relation, so that in (1b), for example, he would assume that the glass broke because it was dropped?

In the recent linguistic literature, these problems have been approached from two rather different perspectives. Within the Gricean pragmatic tradition, a sharp

line is drawn between decoding and inference, and the temporal and causal connotations of (1) and (2) are seen as purely inferential. Within this tradition, the aim is to find a few very general pragmatic principles which will interact with sentence meaning and contextual assumptions to yield the desired interpretations. In the framework of 'discourse semantics', by contrast, the dividing line between decoding and inference has become rather blurred, and a variety of special-purpose rules have been proposed to generate temporal and causal connotations.[1] In this paper, we will look mainly at issues that arise within the Gricean pragmatic framework. A fuller account would deal with the many interesting questions raised by the 'discourse semantic' approach.

The paper is organised as follows. We will argue, first, that while Grice was right to treat the temporal and causal connotations of (1) and (2) as properly pragmatic, they are best analysed not as implicatures but as pragmatically determined aspects of truth-conditional content, or 'what is said'. We will then look at some attempts to deal with the sequencing problem using principles such as Grice's maxim 'Be orderly', and suggest a more general approach. Finally, we will show how all three problems might be tackled within the framework of relevance theory, and point out some implications of this approach.

## 2. Temporal and causal connotations: implicatures or pragmatically determined aspects of what is said?

Ordinary-language philosophers (e.g. Strawson 1952) used to argue, on the basis of examples like (1), that 'and' in natural language differed in meaning from truth-functional '&' in logic. In these examples, so the argument went, natural-language 'and', was equivalent in meaning to 'and then', or 'and so'; hence, a change in the order of conjuncts would lead to a change of meaning. In logic, by contrast, 'P & Q' was invariably equivalent to 'Q & P'.

Grice, in his William James Lectures (1967/1989), defended the view that natural-language 'and' was equivalent in meaning to '&' in logic. He pointed out that the temporal and causal connotations of (1) were not best analysed as part of the meaning of 'and', since the non-conjoined utterances in (2) have the same temporal and causal connotations. On his view, these connotations were best derived via the operation of his Co-operative Principle and maxims. In other words, he rejected a decoding account of these connotations in favour of an inferential approach.

In Grice's framework, the temporal and causal connotations of (1) and (2) were analysed as conversational implicatures: that is, beliefs that had to be attributed to the speaker to preserve the assumption that she was obeying the Co-operative Principle and maxims. More generally, Grice seems to have assumed

that any aspect of utterance interpretation governed by the Co-operative Principle and maxims must be analysed as an implicature, and Gricean pragmatists have invariably followed him on this.[2] Grice drew a sharp dividing line between what was conversationally implicated and what was strictly said. Conversational implicatures made no contribution to the truth conditions of utterances, which were determined solely by what was said. It should follow from Grice's account that the temporal and causal connotations of (1) and (2), which he treats as conversational implicatures, make no contribution to the truth-conditions of utterances in which they occur. But there are problems with this approach.

On Grice's account, natural-language 'and' is semantically equivalent to truth-functional '&' in logic. As Cohen (1971) pointed out, if Grice is right, then reversing the order of the conjuncts in (1) should make no difference to truth conditions: 'P and Q' should always be truth-conditionally equivalent to 'Q and P'. But consider (3) and (4):

(3) It's always the same at parties: either I get drunk and no-one will talk to me or no-one will talk to me and I get drunk.

(4) a. What happened was not that Peter left and Mary got angry but that Mary got angry and Peter left.

b. A: So Peter left and Mary got angry?
B: No. Mary got angry and Peter left.

The disjunction in (3) is not redundant: the assumption that the events happened in a different order in each disjunct makes a genuine contribution to the truth conditions of the utterance. Similarly, the conjunction in (4a) is not a contradiction, and the two utterances in (4b) are not truth-conditionally equivalent. Such examples create a serious problem for Grice. Something which according to him is an implicature appears to be falling within the scope of logical operators and connectives. That is, it appears to be contributing to the truth conditions of the utterance as a whole – in Grice's terms, not to what was implicated but to what was said.

The causal connotations of (1) create similar problems. Consider (5):

(5) Someone left a manhole uncovered and I broke my leg.

(5) would generally be understood as communicating that the speaker broke her leg not only *after* the manhole was left open but *as a result* of the manhole being left open. That these causal connotations can contribute to truth conditions is shown by (6) and (7):

(6) If someone leaves a manhole uncovered and you break your leg, sue.

(7) a. *Peter*: If you leave that manhole uncovered, someone's going to break their leg.

b. *Mary*: No they won't.

In (6), the hearer is not being told to sue if he breaks his leg at some point after someone leaves a manhole open. In (7), Mary's denial is equivalent not to (8a) but to (8b):

(8) a. No-one's going to break their leg.
b. No-one's going to break their leg on that manhole.

These examples show that the sequencing problem and the cause-consequence problem cannot be solved at the level of implicature. Grice himself was aware of this. In a part of the William James lectures that has only recently been published, he considered a similar counterexample to his analysis of 'if': an example in which, quite clearly, something he wanted to treat as an implicature was falling under the scope of logical connectives and contributing to truth conditions. Grice took pains to emphasise that this should not happen, and added "I am afraid I do not yet see what defense, if any, can be put up against this objection" (Grice 1989: 83).

In our book *Relevance* (1986/1995), we suggested a possible defence. Grice's problem, we argued, arose from his assumption that the only way that pragmatic principles can contribute to utterance interpretation is by giving rise to implicatures. Grice assumed that the truth-conditional content of an utterance is recovered largely by decoding, and it seems not to have occurred to him that his CP and maxims could play a role in determining what is said. Robyn Carston (1988) showed how a variety of Gricean 'implicatures', including the temporal and causal connotations of (1) and (2), could be reanalysed as pragmatically determined aspects of what is said. On her account, pragmatic principles make a much greater contribution to truth-conditional content than has generally been assumed. In what follows, we will adopt her approach, and treat the temporal and causal connotations of (1) and (2) as inferentially determined aspects of what is said.[3]

## 3. The sequencing problem and the maxim 'Be orderly'

Grice's solution to the sequencing problem was based on his maxim 'Be orderly', which instructs speakers to recount events in the order in which they happened. Several of the special-purpose sequencing principles proposed by discourse semanticists look like attempts to implement this maxim in the form of code-like rules. Dowty's Temporal Discourse Interpretation Principle (1986) is a case in point:

Temporal Discourse Interpretation Principle
Given a sequence of sentences $S_1...S_n$ to be interpreted as a narrative discourse, the reference time of each sentence $S_i$ is interpreted to be:

(a) a time consistent with the definite time adverbials in $S_i$, if there are any;
(b) otherwise, a time which immediately follows the reference time of the previous sentence $S_{i-1}$.

In this section, we will consider the implications of approaches along these lines.

Notice, first, that neither Grice's maxim of orderliness nor Dowty's code-like variant yields any insight into the interval problem or the cause-consequence problem. For example, (1b) would be compatible with Grice's maxim of orderliness if ten years had elapsed between the dropping and the breaking of the glass, yet this is not normally an appropriate interval for the interpretation of (1b). The maxim of orderliness leaves the interval problem untouched.

Defenders of sequencing principles generally acknowledge that further machinery is needed to solve the interval problem. We all know that if a glass is dropped it typically breaks upon impact, whereas if an acorn is planted it takes some time to grow. The idea would be that these contextual assumptions interact with further pragmatic principles to determine the correct intervals between events in (1b) and (1c). Here, an obvious point is that if such principles are good enough to solve the interval problem, then *a fortiori* they will solve the sequencing problem, and sequencing principles are redundant.

Similar remarks apply to the cause-consequence problem. In order to make the correct predictions about the truth conditions of (6) and (7), some pragmatic principles are needed that will interact with contextual assumptions to assign the appropriate causal relations between the events described. But, since causal relations are typically sequential, such principles would *a fortiori* solve the sequencing problem and make sequencing principles redundant.

A further problem with sequencing principles of the type proposed above is that the constraints they impose on utterance interpretation seem to be too strong. Take an utterance that describes two events or states without explicitly stipulating any temporal ordering between them. Then there are four logically possible ways in which the events or states might be temporally related:

(a) the first mentioned state/event happened before the second;
(b) the two were simultaneous;
(c) the second happened before the first;
(d) no ordering, or some subtler ordering, is pragmatically understood.

Contrary to what is suggested by Grice's maxim of orderliness, or Dowty's Temporal Discourse Interpretation Principle, all four logically possible orderings are realised.

As we have seen, (1) and (2) illustrate possibility (a), with the first mentioned event happening before the second. (9) illustrates possibility (b):

(9) a. Susan is underage and can't drink.
b. The home crowd cheered and the away crowd booed.
c. Bill smiled. He smiled sadly.

As (9a) shows, the fact that two states are fully overlapping does not preclude the possibility that one is a consequence of the other. This looks like a counterexample to Levinson's (1983: 146) account of how temporal and causal connotations are derived:

*Levinson's proposal*
Given *p and q*, try interpreting it as:
(i) *p and then q*; if successful, try:
(ii) *p and therefore q*; if successful try also:
(iii) *p, and p is the cause of q.*

On Levinson's account, the temporal sequencing principle in (i) seems intended to function as a filter, saving the hearer the effort of looking for cause-consequence relations if the sequencing principle does not apply. On this account, since (9a), for example, cannot be interpreted by clause (i) as saying 'Susan is underage and then she can't drink', the hearer should never even test whether the cause-consequence clauses in (ii) and (iii) apply. Clearly, this prediction is false.

Possibility (c) is illustrated by (10):

(10) a. The glass broke. John dropped it.
b. I hit Bill. He insulted me.
c. I got caught. My best friend betrayed me.

In these examples, the speaker first states a fact and then explains it. The natural interpretation of (10a) is that the glass broke both after John dropped it and because John dropped it. In (10b) the temporal and causal order could run either way, and in many circumstances would be taken to run contrary to the predictions of Grice's maxim of orderliness. (These examples have been dealt with in the framework of discourse semantics by Lascarides, Asher & Oberlander (1992), Lascarides (forthcoming); and in the framework of relevance theory by Carston (1993).)

Notice, by the way, that this is one of the few cases where an interpretation possible for a non-conjoined utterance is generally not available for its conjoined counterpart. An adequate account of temporal and causal connotations should explain why the reverse-causal interpretations of (10) are not available for (11):

(11) a. The glass broke and John dropped it.
b. I hit Bill and he insulted me.
c. I got caught and my best friend betrayed me.

Finally, as has often been noted, no ordering is necessarily imposed in examples like (12):

(12) a. That night, our hero consumed half a bottle of whisky and wrote a letter to Lady Anne.
b. Today I signed a contract with a publisher and had tea with an old friend.

There are also some more interesting cases, first noted by Sue Schmerling (1975), where it is not clear what the various sequencing principles would predict. These are cases where, intuitively at least, there is both a temporal and a consequential relation, but the temporal relation is not adequately paraphrased by the addition of 'then'. Schmerling's example was (13); a simpler example without the quantified NP would be (14):

(13) a. We investigated all of the cases and discovered that the problem was more complex than we had thought.
b. We investigated all of the cases and then discovered that the problem was more complex than we had thought.
(14) a. I spoke to John and discovered that he was charming.
b. I spoke to John and then discovered that he was charming.

What the addition of 'then' does to (14a) is to convert it from an interpretation on which I spoke to John and *in* doing so discovered that he was charming into one on which I spoke to John and *after* doing so discovered that he was charming. In both cases there is an intuitive temporal relation, but the temporal relations are not the same. It appears that many sequencing principles would wrongly interpret (14a) as meaning (14b) (and Levinson's principles mentioned above, having failed at clause (i), would not assign a consequential relation at all).

Towards the end of his article, David Dowty lists some of the problems with sequencing principles that have been mentioned here. He suggests that his Temporal Discourse Interpretation Principle might perhaps be treated as a default rule, "to be followed when neither time adverbials nor entailments and implicatures of the discourse itself give clues to the ordering of events," and adds:

> "At this point, in fact, one is entitled to ask whether the Temporal Discourse Interpretation Principle is to be regarded as an independent principle of discourse interpretation *per se*, or merely as a description of the typical outcome of the interaction of various conversational principles and the speakers'/hearers' knowledge of typical events and typical goals of narratives, any one clause of which may be overridden in various ways in exceptional cases. But this is not a question which can be profitably addressed here." (Dowty 1986: 58-59)

Dowty's comments raise an issue of principle. Clearly, Grice's maxim 'Be orderly' is inadequate to deal with the full range of cases discussed in this section. At this point, a choice must be made. One can either go in the direction of discourse semantics, and try to develop a set of special-purpose interpretation rules which will deal with the full range of cases; or one can take seriously the distinction between decoding and inference, and go in the direction of a more general inferential account. In the rest of this paper, we will explore the second option by sketching the lines along which we think a general inferential solution might be found.

## 4. Understanding and relevance

In our book *Relevance* (1986/1995), we developed an account of inferential communication designed to explain how hearers recognise the overtly intended interpretation of an utterance: the one the speaker wants the hearer to recover, is actively helping the hearer to recover, and would acknowledge if asked. Our account was based on the following assumptions. First, every utterance has a variety of possible interpretations, all compatible with the information that is linguistically encoded. Second, not all these interpretations occur to the hearer simultaneously: for example, some disambiguations, some contextual assumptions, some implicatures require more effort to recover. Third, hearers are equipped with a single, very general criterion for evaluating interpretations as they occur to them. And, fourth, this criterion is powerful enough to enable the hearer to recognise the intended interpretation as soon as it is encountered, without having to construct and evaluate a range of alternative interpretations.

The criterion proposed in *Relevance* is based on a fundamental assumption about human cognition: that human cognition is relevance-oriented; we pay attention to information that seems relevant to us. Now every utterance starts out as a request for the hearer's attention. As a result, it creates an expectation of relevance. It is around this expectation of relevance that our criterion for evaluating possible interpretations is built.

Relevance is defined in terms of cognitive effects and processing effort. Cognitive effects are achieved when newly-presented information interacts with a context of existing assumptions by strengthening an existing assumption, by contradicting and eliminating an existing assumption, or by combining with an existing assumption to yield a contextual implication (that is, a conclusion deducible from new information and existing assumptions together, but from neither new information nor existing assumptions alone). The greater the cognitive effects, the greater the relevance will be.

Cognitive effects, however, do not come free: they cost some mental effort to derive, and the greater the effort needed to derive them, the lower the relevance will be. The processing effort required to understand an utterance depends on two main factors: the form in which it is presented (audibility, legibility, dialect, register, syntactic complexity and familiarity of constructions all affect processing effort); and the effort of memory and imagination needed to construct a suitable context. The greater the processing effort required, the lower will be the relevance, and the greater the risk of losing the hearer's attention.

Relevance theory assumes that every aspect of communication and cognition is governed by the search for relevance. Cognition is governed by the search for maximal relevance (i.e. the greatest possible effects for the smallest possible effort). This is expressed in the First, or Cognitive, Principle of Relevance (*Relevance* 1995: 260-78):

*Cognitive principle of relevance*
Human cognition tends to be geared to the maximisation of relevance.

What is unique to overt communication is that, approaching an utterance addressed to us, we are entitled to have not just hopes but steady expectations of relevance. The Second, or Communicative Principle of Relevance is the principle that every utterance (or other act of ostensive communication) creates a presumption of relevance. Relevance, we have seen, is defined in terms of cognitive effect and processing effort; but what exactly does the hearer's presumption of relevance amount to, in terms of effort and effect?

It is clear that the presumption is not one of maximal relevance. Communicators are not always expected to give the most relevant possible information, or to present it in the least effort-demanding way, as a presumption of maximal relevance would suggest. On the content side, the speaker may not have the information that the hearer would find most relevant; she may be unwilling to give it, or unable to think of it at the time. On the formal side, lack of time, lack of ability or stylistic preferences may prevent her expressing herself in the most economical way. Moreover, the most relevant utterance the speaker can think of may still not be relevant enough to be worth the hearer's attention; we need to set a lower limit on the expected degree of relevance, to explain why such an utterance would generally not be produced.

In *Relevance* (1986/1995: 260-78), we define a notion of optimal relevance, designed to take these various factors into account:

*Optimal relevance*
An utterance, on a given interpretation, is optimally relevant iff:
(a) It is relevant enough for it to be worth the addressee's effort to process it;
(b) It is the most relevant one compatible with the communicator's abilities and preferences.

Clause (a) of the definition of optimal relevance sets a lower limit on the expected degree of relevance. Clause (b) takes into account the fact that the speaker may be unwilling or unable to produce the most relevant possible utterance, or to formulate it in the least effort-demanding way. More positively, clause (b) incorporates the idea that if the speaker can see a way of increasing the relevance of the utterance at no cost to herself, it is in her interest to do so: the more relevant the utterance, the more likely the hearer will be to attend to it and understand it correctly. These ideas may be illustrated by applying them to an example.

According to clause (a) of the definition of optimal relevance, the hearer is entitled to expect the utterance to be relevant enough to be worth his attention. In general, this means that he is entitled to expect it to be more relevant than any other information that he could have been processing at the time. How relevant that is depends on what is going on elsewhere in his cognitive environment. Thus, suppose that someone walks into an important lecture and says (15):

(15) Ladies and gentlemen, the building is on fire.

'The building' is a referential expression, and different assignments of reference lead to different levels of cognitive effect. In the circumstances, the first hypothesis to come to the audience's mind would be that 'the building' means the building where the lecture is taking place. Clearly, the utterance, on this interpretation, would be relevant enough to be worth the audience's attention: their minds would be immediately filled with thoughts of how to get out. In the circumstances, it is hard to see what other interpretation would be relevant enough to justify the interruption, and the interpretation just suggested is basically the only possible one.

It might be thought that in other circumstances the intended interpretation would be harder to pin down. Surely there might be several radically different combinations of content and context, each of which would yield enough cognitive effects to make the utterance worth the audience's attention? This is where clause (b) of the definition of optimal relevance comes in. Recall that we are talking about overt communication, where the speaker is anxious to avoid misunderstanding, and is actively helping the hearer to recognise the intended interpretation. Clearly, it is in such a speaker's interest to make sure that there is no alternative line of interpretation which is both more accessible to the hearer than the intended one, and is relevant enough to be worth his attention, since such an interpretation is likely to lead him astray.

Clause (b) of the definition of optimal relevance, which excludes gratuitous calls on the hearer's processing effort, covers this type of case: that is, it excludes the possibility that the hearer will be expected to recover, process and accept the wrong interpretation before lighting on the intended one. From clause (b), it follows that a speaker aiming at optimal relevance should try to formulate her utter-

ance in such a way that the first acceptable line of interpretation to occur to the hearer is the intended one. From the hearer's point of view, this clause has an immediate practical consequence. Having found a line of interpretation which satisfies his expectation of relevance in a way the speaker might manifestly have foreseen, he need look no further. The first such line of interpretation is the only one; all alternative lines of interpretation are disallowed.

The Communicative Principle of Relevance, then, is a principle of optimal rather than maximal relevance (*Relevance* 1995: 266-78):

> *Communicative principle of relevance*
> Every act of overt communication communicates a presumption of its own optimal relevance.

It is worth noting, though, that in order to be acceptable and comprehensible, an utterance does not actually have to be optimally relevant, but merely such that the speaker might reasonably have expected it to be so. In the first place, hearers are capable of correctly interpreting cases of accidental irrelevance, where an utterance is irrelevant for reasons the speaker manifestly could not have foreseen. In the second place, hearers are capable of correctly interpreting cases of accidental relevance, where an utterance is optimally relevant in a way the speaker manifestly could not have foreseen. To allow for these two types of case, we claim in *Relevance* that interpretations are accepted or rejected according to the following criterion of *consistency with the principle of relevance*:[4]

> *Criterion of consistency with the principle of relevance*
> An utterance, on a given interpretation, is consistent with the principle of relevance if and only if the speaker might reasonably have expected it to be optimally relevant to the hearer on that interpretation.

This criterion, and the definition of optimal relevance which underlies it, provide the key to the relevance-theoretic account of comprehension. In the remainder of this paper, we will apply them to a range of examples including those in (1) and (2) above.

## 5. The interval problem

The interval problem illustrated in (1) and (2) is a special case of a much more general problem. There are many other types of case where temporal intervals are left open by the semantics and narrowed down in the pragmatics. Compare (16a) and (16b):

(16) a. I have had breakfast.
b. I have been to Tibet.

The speaker of (16a) would generally be understood as saying that she had had breakfast that morning, whereas the speaker of (16b) might be understood as saying merely that she had visited Tibet at some time in her life. An adequate treatment of the interval problem should explain why this is so.

Let us assume that what is linguistically encoded in both cases is that the event described took place at some point within an interval stretching back from the moment of utterance. Then the speaker of (16a) can be understood as saying that she has had breakfast *within the last few minutes, within the last few hours, within the last few days, weeks, or months*, and so on, and the hearer's task is to decide which interval she had in mind. Notice that the possible interpretations are logically related: the cases in which the speaker has had breakfast *within the last few minutes* are a subset of those in which she has had breakfast *within the last few hours, days, weeks*, and so on. Notice, too, that the hearer's choice will affect the truth conditions of the utterance: the claim that I have not had breakfast may be true if the chosen interval is the last few hours but false if it is the last few weeks.

There are several variants of the interval problem, each with a similar logical structure. Suppose we meet in the University library, and I say to you:

(17) I've been here all day.

The interpretation of 'here' involves spatial rather than temporal intervals: I might mean that I have been *in this room, in this library, in this building, in this town, in this country*, etc., all day. Again, the possible interpretations are logically related: the cases in which I have been *in this room* all day are a subset of those in which I have been *in this library, this college, this town*, etc., all day. Again, the hearer's choice of interval will affect the truth conditions of the utterance: (17) has different truth conditions depending on whether it is taken to mean *in this room, this town, this country*, etc.

The interpretation of many comparative adjectives has a similar subset structure. Thus, consider (18):

(18) a. John is rich.
b. Susan is tall.
c. We are happy to see you.
d. Bill has a fast car.

(18d), for example, might mean *fast enough to overtake some/many/most/all other cars; fast enough to cause envy among some/many/most/all of Bill's friends*, etc., with the set of cars that are fast enough to overtake *some* other cars being a subset of those that are fast enough to overtake *many/most/all other cars*, and so on.

Or consider (19):

(19) Mary is a working mother.

As Lakoff (1987: 80-2) points out, (19) would generally be understood as communicating more about Mary than that she is a female parent who works: it might suggest, for example, that Mary's children are not grown up and living away from home, that she did not give them up for adoption at birth, and so on; that she not only works but works for money, for more than an hour or two a week, and so on. In other words, (19) is typically understood as applying to some subset of the set of people who satisfy the definition *female parent who works*. As the literature on prototype effects shows, similar cases of concept narrowing occur in the interpretation of virtually every utterance.

The interval problem raised by examples (1) and (2), then, is a special case of a much more general problem, which seems to demand a general solution. We will argue that relevance theory is particularly well suited to resolving indeterminacies with a subset structure of the type just illustrated.

Notice first that the hearer's choice of interval in (16) will affect not only the truth-conditional content of the utterance but also its cognitive effects. Narrower intervals are associated with greater cognitive effects: thus, if you tell me that you have had breakfast *within the last few minutes*, I will be able to derive all the cognitive effects I could have derived from knowing that you have had breakfast *within the last few days*, and more besides.

We claim that these differences in logical structure explain the differences in interpretation of (16a) and (16b). According to clause (a) of the definition of optimal relevance, the hearer should look for an interpretation which is relevant enough to be worth his attention. The difference between (16a) and (16b) is that in (16a), the interval must be narrowed much more drastically to achieve even minimally adequate effects.

In normal circumstances, unless the speaker of (16a) is saying that she has had breakfast within the last few hours, her utterance will have no cognitive effects at all. Most of us would take for granted that she had had breakfast at some point in her life; moreover, it is hard to see what effects she might have hoped to achieve by telling us that she had had breakfast, say, three weeks ago. One can, of course, imagine special circumstances in which the fact that someone has become a breakfast-eater would be highly relevant; but in normal circumstances, the only way for (16a) to achieve adequate effects is by conveying that the speaker has had breakfast on the very day of utterance. With (16b), by contrast, the information that the speaker has visited Tibet at some point in her life would generally be quite relevant enough.

To generalise: the semantics of (16a) and (16b) tells us that the event described happened at some point in an interval stretching back from the time of utterance to the beginning of the universe; the pragmatics tells us that it happened recently enough for the fact to be worth mentioning. In the search for optimal relevance, we will narrow the interval to the point where we have an interpretation

consistent with the principle of relevance. Similar remarks apply to the other examples mentioned above. Thus, when I say that Bill has a fast car, I must be understood as meaning that it is fast enough for the fact to be worth mentioning: how fast that is will vary from occasion to occasion. When I say that Mary is a working mother, I must be understood as meaning that she belongs to some subset of working mothers whose properties are such that the fact that Mary belongs to it is worth mentioning; and so on.

Relevance theory, then, suggests the following general strategy for solving the interval problem: look for an interval narrow enough to yield an interpretation consistent with the principle of relevance. Which raises a further general question: how is the appropriate interval to be found? Here, it is reasonable to assume that hearers have a range of more specific strategies for finding the appropriate interval. It is to these that we now turn.

## 6. The role of contextual assumptions

Let us return to our original examples and look more carefully at the role of contextual assumptions in the interpretation of (1) and (2). Different utterances make different assumptions accessible. Differences in the accessibility of contextual assumptions will affect not only the order in which interpretations are tested, but the acceptability of the results. In this section, we will try to show how the assumptions of relevance theory interact with some widely accepted views on accessibility of contextual information to yield some explanatory insight into the sequencing problem, the interval problem and the cause-consequence problem.

We assume, as do most other people working in the area, that the hearer of (1a), for example, is given immediate access to his encyclopaedic assumptions about keys and doors. We also assume that encyclopaedic entries may contain ready-made chunks or schemas describing often-encountered sequences of actions or events. If such schemas exist, it is clear that we all have one for taking out a key and using it to unlock a door; moreover, by virtue of frequent use, such a schema would be highly accessible for use in interpreting (1a). Using this schema, it should be possible to infer that the interval between the events described is very small – a few seconds, a few minutes at most.

So far, so obvious. Everyone who has ever looked at the interval problem says that somehow it is solved by an interaction of contextual assumptions and pragmatic principles. But no-one, so far as we know, has a pragmatic principle which would explain why the hearer is *entitled* to use normal assumptions in interpreting (1a). After all, there would be nothing to stop someone taking out her key, falling asleep on the doorstep and opening the door next morning; or indeed being arrested and serving a life sentence between taking out her key and finally opening

the door. What entitles the hearer to assume that the speaker did not have some such non-standard situation in mind?

As far as we can see, on most other approaches, an unargued appeal is made to statistical likelihood of events. The speaker *could* have meant something else, but because the event described is statistically unlikely, the associated interpretation is statistically unlikely too. No such appeal is needed in relevance theory: the appropriate interpretation falls out automatically from the criterion of consistency with the principle of relevance.

Because, at least in normal circumstances, normal assumptions are the easiest to access, the hearer is entitled to use them as long as they give rise to an interpretation consistent with the principle of relevance. And as long as they do, by clause (b) of the definition of optimal relevance, other, less accessible lines of interpretation are disallowed. From which there follows a further prediction: in the circumstances just described, a speaker who wants to communicate that some non-standard interval elapsed between taking out the key and opening the door would be unable to communicate it by means of (1a).

By the same token, the interpretation of (1b) and (1c) will depend on accessible contextual schemas. The standard assumption is that a dropped glass breaks on impact, whereas a planted acorn takes days or weeks to grow. Encyclopaedic knowledge about these events makes certain assumptions highly accessible; if these lead on to interpretations consistent with the principle of relevance, all other lines of interpretation are disallowed.

Consider now another aspect of the interpretation of (1a). Though the speaker does not explicitly say so, in normal circumstances she would be taken to communicate that she opened the door with the key, and that she did so in the normal way, by inserting the key in the lock. These facts are explained by the relevance-theoretic account. By saying that she took out her key, the speaker causes her hearer some processing effort; if this is not to be gratuitous, it must make some contribution to cognitive effects. Again, the existence of an encyclopaedic schema points to an obvious hypothesis: that she used the key in the normal way to open the door. In normal circumstances, the resulting interpretation would be consistent with the principle of relevance, and all other lines of interpretation would be disallowed.

What happens when there is no ready-made schema to guide the interpretation process? Consider (20):

(20) John took out his handkerchief and opened the door.

This utterance does not describe a regular and frequently encountered sequence of events, and it is unlikely that the hearer has an appropriate schema waiting to be used. In fact, there are at least two lines of interpretation that he might pursue. In the context of a detective story, one way to exploit the information that John took

out his handkerchief might be to assume that he used it to cover the doorhandle as he opened it, to avoid leaving fingerprints. This interpretation would yield both a sequencing effect and a definite hypothesis about the interval between the two events. In other circumstances, the best hypothesis might simply be that these are two unrelated events that happened at about the same time. Notice, though, that by the arguments given above, the fact that John took out his handkerchief must contribute in some way to overall relevance, and the reader of a planned text will be left with a strong expectation that the fact that John took out his handkerchief will prove relevant later on.

Talk of instruments leads naturally to talk of causes. Consider (1d):

(1) d. Peter left and Mary got angry.

Why would this be naturally understood as communicating not only that Mary got angry after Peter left, but that she got angry because he left? Here we must speculate a little further about the type of mental schemas that humans are likely to construct.[5] A large part of our cognitive life is taken up with consideration of causes and effects. On the assumptions of relevance theory, this is no accident: causal stories are highly relevant, because they enable us to predict the consequences of our own actions and those of others. It is not surprising, then, that causal schemas come readily to mind for the interpretation of utterances such as (1d). As we have seen, if (1d) is to yield an interpretation consistent with the prin ciple of relevance, the information that Peter left must make some contribution to overall cognitive effects. Just as it is easy to see the key as an instrument in (1a), so it is easy to see Peter's departure as a cause in (1d). If this highly accessible line of interpretation proves consistent with the principle of relevance, all other interpretations will be disallowed.

As Posner (1980) has shown, there is a huge variety of ways in which conjoined utterances such as (1) and (2) may be enriched with instrumental, causal, locative, durative, etc. material. It should be clear how all these cases would be handled in relevance-theoretic terms. If the linguistically encoded information is too vague, or too incomplete, to yield an interpretation consistent with the principle of relevance, it will be enriched using immediately accessible contextual assumptions, to a point where it is relevant enough.

What we have tried to develop in the last few sections are general answers to two very general questions: why is it that, in interpreting an utterance, the hearer often enriches its truth-conditional content beyond what is strictly encoded; and how does the enrichment process go? We have argued that the enrichment process is triggered by the criterion of consistency with the principle of relevance, which warrants the selection of the most accessible enrichment that yields an acceptable overall result. This account makes no appeal to special-purpose sequencing principles such as Grice's maxim of orderliness or Dowty's Temporal Discourse In-

terpretation Principle; the temporal and causal connotations of (1) and (2) arise from an interaction between sentence meaning, general cognitive factors and the criterion of consistency with the principle of relevance.

## 7. Reverse-causal interpretations

In this final section, we will look at some examples that have been claimed to cast doubt on the feasibility of a general inferential account of the type just sketched. We will draw heavily on the work of Robyn Carston (see in particular Carston 1993; Carston 1997), who has discussed these issues in some detail.

Recall examples (10)-(11) above, in which the reverse-causal interpretations available for non-conjoined utterances are apparently not available for the corresponding conjoined utterances:

(10) a. The glass broke. John dropped it.
b. I hit Bill. He insulted me.
c. I got caught. My best friend betrayed me.

(11) a. The glass broke and John dropped it.
b. I hit Bill and he insulted me.
c. I got caught and my best friend betrayed me.

Bar-Lev and Palacas (1980) have argued on the basis of these examples that 'and' must encode some additional information that blocks the reverse-causal interpretations. If so, then our inferential approach must be supplemented by an additional element of decoding.

According to Bar-Lev and Palacas (1980: 41) the differences between (10) and (11) result from the fact that 'and' encodes the following constraint:

*Bar-Lev & Palacas: Semantic command constraint*
The second conjunct is not prior to the first (chronologically or causally).

This would allow for the 'forward' temporal and causal interpretations of the utterances in (1), for the overlapping interpretations of utterances such as (9) and (13), and for the non-temporal and non-causal interpretations of utterances such as (12). However, it would rule out the reverse-causal interpretations of (11), which violate the semantic command constraint.

There are several problems with this account. In the first place, there are coun terexamples. Larry Horn has pointed out that in cases such as (21b) a reverse-causal interpretation is possible with 'and':

(21) a. *Peter*: Did John break the glass?
b. *Mary*: Well, the glass broke, and John dropped it.

Here, Mary clearly implicates that the glass broke because John dropped it. Or consider (11c) above, with additional 'comma' intonation:

(11)c'. I got caught, and my best friend betrayed me.

(11c') might well imply that the speaker got caught because her best friend betrayed her. Notice that in these cases the inference is relatively indirect: the temporal and causal connotations appear to be part of what is implicated rather than what is said. In any case, (21b) and (11c') are counterexamples to the semantic command constraint, which excludes both reverse-causal and reverse-temporal interpretations.

In the second place, there are differences in the interpretation of conjoined and non-conjoined utterances which are not explained by the semantic command constraint. Compare (22) and (23):

(22)a. I met someone famous last night. I met Chomsky.
b. I met someone famous last night, and I met Chomsky.
(23)a. I ate somewhere nice yesterday. I ate at Macdonald's
b. I ate somewhere nice yesterday, and I ate at Macdonald's.

The events described in the non-conjoined utterances in (22a) can be understood as simultaneous (or identical); corresponding interpretations are not available for the conjoined utterances in (22b). Similar remarks apply to (23a) and (23b). There is nothing in Bar-Lev & Palacas's constraint to explain why this is so.

Bar-Lev & Palacas themselves cite a range of conjoined examples which are not explained by their constraint. These involve a variety of rhetorical relations which they call exemplification, as in (24), conclusivity, as in (25), and explanation, as in (26):

(24)a. Wars are breaking out all over: Champaign and Urbana have begun having border disputes.
b. Wars are breaking out all over, and Champaign and Urbana have begun having border disputes.
(25)a. There are his footprints: he's been here recently.
b. There are his footprints, and he's been here recently.
(26)a. Language is rule-governed: it follows regular patterns.
b. Language is rule governed, and it follows regular patterns.

In each case, an interpretation which is possible for the non-conjoined (a) sentence is ruled out for the conjoined (b) sentence. Bar-Lev and Palacas comment that in these non-temporal, non-causal utterances with 'and', neither 'forward' nor 'backward' relations between the conjuncts are possible. Notice that their semantic command constraint does not exclude these cases, and any constraint which excludes them would be likely to run into counterexamples with utterances such as (27):

(27) a. Wars are breaking out all over, and Champaign and Urbana in particular have begun having border disputes.
b. There are his footprints, and so he's been here recently.
c. Language is rule governed, and the reason is that it follows regular patterns.

Here, though the 'excluded' relations are explicitly encoded, the results are quite acceptable.

Rather than pursue the decoding approach to these examples, we would suggest the following inferential account. In each of (22a)-(26a), the speaker raises a question in the first part of the utterance, which is answered in the second part. In (22a), for instance, she expects the hearer to start wondering who she met; in (24a) she expects him to start wondering where wars are breaking out; in (26a) she anticipates the question 'What does it mean to say that language is rule-governed?'

Conjoined utterances with 'and' exclude interpretations along these lines: that is, interpretations on which the second conjunct achieves relevance primarily by answering a question raised by the first. Why is this so? One possible explanation (suggested by Blakemore 1987) is that a conjoined utterance is presented as a unit, encouraging the hearer to process the two conjuncts jointly and in parallel, looking for implications derivable from both. Question-answer pairs are not normally suited to joint processing, as witness the unacceptability of (28):

(28) a. ?You'll never guess what time I finished, and I finished at 6.00.
b. ?What time do you think I finished, and I finished at 6.00.
c. ?You'll be amazed when you hear what time I finished, and I finished at 6.00.

Inferential approaches along these lines have been explored by Carston (1993a) and Carston (1997). If successful, they would eliminate the need for a decoding account.[6]

## 8. Conclusion

In this paper, we have sketched an inferential account of the causal and temporal connotations of utterances (1) and (2). This account is very general, and deals with a wide range of phenomena that are neither temporal nor causal in nature. By contrast, the various special-purpose principles proposed by discourse semanticists have limited application, and are themselves in need of further explanation. We believe that these principles are best seen as implementations of a more general inferential account such as the one developed here. Unless backed by such an account, their appeal will remain more descriptive than explanatory.

## Notes

* In writing this paper (of which an earlier version appeared in *Languages* 112, December 1993), we have benefited greatly from discussions with Robyn Carston, and borrowed many examples and arguments from her published and unpublished research (see, for example, Carston 1988, 1993a, 1993b, 1997).

1. See, for example, Dowty (1986), Lascarides (forthcoming), Lascarides, Asher and Oberlander (1992), Lascarides and Oberlander (1993).

2. For Gricean solutions to the sequencing problem, see Harnish (1976), Gazdar (1979), Posner (1980), Levinson (1983), Comrie (1985), Green (1989) and Horn (1989); for critical discussion see Carston (1988, 1993, 1994, 1997), Recanati (1989, 1993, 1994).

3. Some experimental confirmation of the claims of this section is provided by Gibbs and Moise (1996). While the contribution of pragmatics to truth-conditional content is now increasingly acknowledged, there is still some debate about how the results are to be described. Sperber and Wilson (1986/1995) introduce a notion of *explicature*, which is broader than Grice's notion of *what is said* and covers the types of case discussed above. Levinson (1987) retains the term 'implicature' but distinguishes two sub-types: those that contribute to truth-conditional content and those that do not. Bach (1994a, 1994b) distinguishes *implicitures*, which contribute to truth-conditional content, from *implicatures*, which do not. For discussion, see Carston (1997).

4. For further discussion of these issues, see Sperber (1994), Wilson (forthcoming).

5. For some discussion, see Sperber, Premack and Premack (1995).

6. For general discussion of the notions of exemplification and restatement within relevance-based and coherence-based frameworks, see Blakemore (1997).

## References

Bach, K.
1994a "Conversational impliciture." *Mind & language* 9, 124-62.
1994b "Semantic slack: what is said and more." In S. Tsohatzidis (ed.) *Foundations of speech act theory*. London: Routledge, 267-91.

Bar-Lev, S. and A. Palacas
1980 "Semantic command over pragmatic priority." *Lingua* 51, 137-46.

Blakemore, D.
1987 *Semantic constraints on relevance*. Oxford: Blackwell.
1997 "Restatement and exemplification." To appear in *Pragmatics and Cognition* 5, 1-19.

Carston, R.
1988 "Implicature, explicature and truth-theoretic semantics." In R. Kempson (ed.) *Mental representations: The interface between language and reality*. Cambridge: Cambridge University Press, 155-81. Reprinted in S. Davis (ed.) 1991. *Pragmatics: A reader*. Oxford, Oxford University Press, 33-51.

1993 "Conjunction, explanation and relevance." *Lingua* 90, 27-48.
1994 "Conjunction and pragmatic effects." In *The encyclopaedia of language and linguistics*. Oxford: Pergamon Press.
1997 *Pragmatics and the explicit-implicit distinction.* PhD dissertation, University of London. To be published by Blackwell, Oxford.

Cohen, L.J.
1971 "Some remark on Grice's views about the logical particles of natural language." In Y. Bar-Hillel (ed.) *Pragmatics of natural languages*. Dordrecht: Reidel, 50-68.

Comrie, B.
1985 *Tense*. Cambridge: Cambridge University Press.

Dowty, D.
1986 "The effects of aspectual class on the temporal structure of discourse: semantics or pragmatics?" *Linguistics & Philosophy* 9, 37-61.

Gazdar, G.
1979 *Pragmatics: Presupposition, implicature and logical form*. New York: Academic Press.

Gibbs, R. and Moise, J.
1996 "Pragmatics in understanding what is said." *Cognition* 62, 51-74.

Green, G.
1989 *Pragmatics and natural language understanding.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Grice, H.P.
1967 *William James Lectures*. Reprinted in H. P. Grice, 1989. *Studies in the way of words*. Cambridge MA: Harvard UP.

Harnish, R.M.
1976 "Logical form and implicature." In T. Bever, J. Katz & T. Langendoen (eds.) *An integrated theory of linguistic ability*. New York: Crowell, 313-91. Reprinted in S. Davis (ed.) 1991. *Pragmatics: A reader.* Oxford, Oxford University Press, 316-64.

Horn, L.
1989 *A natural history of negation*. Chicago: University of Chicago Press.

Lakoff, G.
1987 *Women, fire and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Lascarides, A.
(forthcoming) "Knowledge, causality and temporal representation." To appear in *Linguistics* 30.

Lascarides, A., Asher, N. and Oberlander, J.
1992 "Inferring discourse relations in context." Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics, Delaware, June 1992.

Lascarides, A. and J. Oberlander
1993 "Temporal coherence and defeasible knowledge." *Theoretical Linguistics* 19, 1: 1-37.

Levinson, S.
1983 *Pragmatics*. Cambridge: Cambridge University Press.
1987 "Minimization and conversational inference." In J. Verschueren and M. Bertucelli-Papi (eds.) *The pragmatic perspective*. Amsterdam: John Benjamins, 61-129.

Posner, R.
1980 "Semantics and pragmatics of sentence connectives in natural language." In J. Searle, F. Kiefer and M. Bierwisch (eds.) *Speech act theory and pragmatics*. Dordrecht: Reidel, 168-203.

Recanati, F.
1989 "The pragmatics of what is said." *Mind & Language* 4, 295-329. Reprinted in S. Davis (ed.) 1991. *Pragmatics: A reader*. Oxford, Oxford University Press, 97-129.
1993 *Direct reference: From language to thought*. Oxford: Blackwell.
1994 "Contextualism and anti-contextualism in the study of language." In S. Tsohatzidis (ed.) *Foundations of speech act theory: Philosophical and linguistic perspectives*. London: Routledge, 156-65.

Schmerling, S.
1975 "Some remarks on symmetric and asymmetric conjunctions." In P. Cole and J. Morgan (eds.) *Syntax & semantics 3: Speech acts*. New York: Academic Press, 211-32.

Sperber, D.
1994 "Understanding verbal understanding." In J. Khalfa (ed.) *What is intelligence?* Cambridge: Cambridge University Press, 179-98.

Sperber, D., Premack, D. and Premack, A. (eds.)
1995 *Causal cognition: A multidisciplinary debate*. Oxford: Clarendon Press.

Sperber, D. and D. Wilson
1986 *Relevance: Communication and cognition*. Oxford: Blackwell. Second edition 1995.

Strawson, P.
1952 *Introduction to logical theory*. London: Methuen.

Wilson, D.
1994 "Relevance and understanding." In G. Brown, K. Malmkjaer, A. Pollitt and J. Williams (eds.) *Language and understanding*. Oxford: Oxford University Press, 35-58.
(forthcoming) "Discourse, coherence and relevance: a reply to Rachel Giora." To appear in *Journal of Pragmatics*.