5 Discourse and Relevance Theory

DIANE BLAKEMORE

0 Introduction

It is generally agreed that the study of discourse takes us beyond the study of the sentence. However, as this book demonstrates, we are not always taken to the same place. In some cases, it seems, we are not taken that very far at all: thus according to the tradition set by Zellig Harris (1951), discourse is a structural unit which can be studied by analogy with the sentence. For example, Salkie (1995) suggests that while grammar is "basically about how words combine to form sentences, text and discourse analysis is about how sentences combine to form texts." And Hovy and Maier's (1994) work in artificial intelligence is based on the claim that "one of the first observations that one makes in analysing discourse is that it exhibits internal structure" (1994: 2).

In other cases, we are taken beyond and away from the notion of structure altogether to the notion of discourse as social behavior which must be studied in terms of its function. Thus Fasold (1990) defines the study of discourse as the study of any aspect of language use (1990: 65). And of course, one of the most influential books on linguistic aspects of discourse, Halliday and Hasan's (1976) *Cohesion in English*, is based on the view that a text is a "unit of language in use" (1976: 2) which must be studied in terms of its function in communication.

Notice that the analogy between discourse and language that is assumed by Zellig Harris is an analogy between discourse and what Chomsky has called externalized language (or E-language) (Chomsky 1986). This means that according to this view, a theory of discourse, like a grammar, is a collection of descriptive statements external to the human mind. Similarly, the functional view of discourse, in leading us from the study of the structural properties of discourse to the study of discourse as communicative behavior, has taken us to a phenomenon that is defined independently of the human mind.

If discourse is defined from either of these perspectives, then relevance theorists do not study discourse at all. For the object of study is not discourse, whether this be defined in terms of a structural phenomenon or a social phenomenon, but rather discourse understanding, or more particularly, the mental representations and computations underlying utterance understanding. In other words, the concern in relevance theory is with something internal to the human mind.

In drawing this analogy between relevance theoretic approaches to discourse and Chomskyan linguistics, I do not mean to suggest that there is an analogy between a theory of utterance understanding and grammar, or that a theory of discourse understanding is to be somehow accommodated within a theory of generative grammar. On the contrary, it is argued that Chomsky's modular view of the mind allows us to draw a principled distinction between a theory of grammar and a theory of utterance understanding. As we shall see in this chapter, while grammar plays a role in communication, this role is to deliver not representations of the thoughts that speakers communicate, but semantic representations which fall short of the complete interpretation intended. The contextual assumptions required for a complete interpretation of the speaker's intentions and the computations that are used in deriving this interpretation are *outside* the language module (grammar). As Deirdre Wilson (1995) has said, "there is no more reason to expect discourse to have the same structure as language than there is to expect it to have the same structure as vision." In particular, there is no reason to expect discourse to be analyzed in terms of a code or set of rules or conventions (see also Wilson and Sperber 1986).

1 Coherence and Discourse

The claim that a theory of discourse involves the search for the rules or conventions which govern it has dominated both structural and functional approaches to discourse. In structural approaches, the aim is to discover the rules which, if followed, result in an acceptable or well-formed text. In approaches which view discourse in terms of communicative behavior, the aim is to discover the social conventions which determine which utterances may occur and what they may be combined with. In other words, the main concern is with the *acceptability* of discourse.

According to one example of this approach, discourse is acceptable to the extent that it exhibits coherence relations between its segments. Thus for example Mann and Thompson (1987, 1988) argue that the reason why only the first of the sequences in (1) "works" is that our contextual assumptions about cars do not allow us to derive an interpretation of (1b) which is consistent with our assumption that the text is coherent:

(1) a. I love to collect classic automobiles. My favourite car is my 1899 Duryea.
  b. I love to collect classic automobiles. My favourite car is my 1977 Toyota. (Mann and Thompson 1987: 57)

(1a) succeeds as a text because the contextual assumption that a 1899 Duryea is a member of the set of classic automobiles enables the hearer to establish that the two segments satisfy the relation of *elaboration*.[1]

This is not the only approach to coherence.[2] I focus on this approach here because some theorists who have taken it have also claimed that it provides the key to a theory of discourse *comprehension* in the sense that it is the search for coherence that leads to the successful comprehension of utterances. For example, Mann and Thompson (1987, 1988) have claimed that the search for coherence plays an essential role for the recovery of the implicatures recovered from an utterance. Hobbs (1979) has claimed that reference assignment is a consequence of the hearer's search for coherence. And Asher and Lascarides (1995) have argued that disambiguation can be seen as a consequence of the hearer's search for discourse coherence. My aim in this chapter is to outline the arguments which suggest that a theory of discourse comprehension should not be regarded as a by-product of a theory of discourse acceptability (or coherence), but is actually the key to the explanation of our intuitions about coherence. In other words, it is the notion of coherence that is derivative. More specifically, it can be seen as a consequence of the hearer's search for an interpretation that is consistent with Sperber and Wilson's (1986) Principle of Relevance.

Clearly, the success of a theory based on the assumption that the acceptability of discourse depends on coherence relations must be based on a complete taxonomy of coherence relations. However, this is the focus of considerable controversy. Mann and Thompson (1988) themselves propose a taxonomy based on 15 relations – rather fewer than the 70 relations proposed by Hovy and Maier (1994) but rather more than the four basic relations proposed by Sanders et al. (1993). Moreover, there are disagreements over how relations should be subclassified. For example, while Hovy and Maier (1994) suggest that both exemplification and restatement are subtypes of elaboration, Mann and Thompson (1988) include only exemplification as a subtype of elaboration and define restatement as a separate relation.

In fact, as we shall see, it is not clear that either restatement or elaboration provides an adequate theoretical basis for the analysis of reformulation sequences or for utterances intended as examples. More fundamentally, it is not clear that the assumptions underlying any taxonomy of coherence relations can be justified. Work by Blass (1990), Deirdre Wilson (1998), and myself (Blakemore 1988b, 1996, 1997) shows that coherence relations are necessary or sufficient for the acceptability of discourse or for its successful comprehension.

Coherence relations are structural relations which hold in virtue of formal properties of utterances. However, as Blass (1990) points out, intuitions about pragmatic acceptability are affected not only by the form of utterances, but also by their content. This means that it is possible to construct texts which are unacceptable even though they satsify formal coherence relations. Consider, for example, elaboration, which, for some writers, includes not only examples like (2), but also repetitions like (3):

(2) Go down Washington Street. Just follow Washington Street three blocks to Adams Street. (from Hobbs 1979)

(3) There's a mouse, a mouse.

As Blass's examples in (4) and (5) show, not every utterance recognizable as an elaboration or repetition is appropriate:

(4) Go down Washington Street. Just pick up your left foot, place it down in front of your right foot, transfer your weight from right to left foot, lift your right foot . . .

(5) [speaker goes into a shop] A box of cornflakes please. A box of cornflakes please. A box of cornflakes please . . .

In the same way, not every utterance satisfying a restatement relation would be appropriate. For example, a speaker who has just seen a mouse running across the bedroom floor is unlikely to produce the sequence in (6):

(6) There's a mouse, a small grey furry rodent.

According to Mann and Thompson (1988), the intended effect of restatement is simply that the hearer recognize that a restatement is being made. However, it is difficult to see how this provides a means for distinguishing acceptable restatments from unacceptable ones or for distinguishing the effects of a restatment such as (7) from those derived from the sequence in (8):

(7) a. At the beginning of this piece there is an example of an anacrusis.
b. That is, it begins with an unaccented note which is not part of the first full bar. (Blakemore 1997a)

(8) a. A well-groomed car reflects its owner.
b. The car you drive says a lot about you. (Mann and Thompson 1988)

In the following section, we shall see how the interpretation of restatement sequences can be explained in terms of the notion of optimal relevance and the criterion of consistency with the principle of relevance.

It might be argued at this point that while the hearer's recognition of coherence relations is not enough to provide a full account of how these sequences are interpreted, the recognition of coherence relations is nevertheless necessary for comprehension. In other words, it could be claimed that in order to understand the utterance *U1* in the sequence *U1. U2* it is necessary to recover what Mann and Thompson (1987) call "relational proposition," which expresses a particular structural relation.

However, as Blass (1990) has pointed out, everyday discourse is full of acceptable utterances which cannot be understood in isolation from the context, but which cannot be said to be part of a coherent text. For example, travellers on the London Underground are able to recognize that the utterance displayed at the foot of escalators is not intended to be interpreted as a requirement that everyone using the escalator must carry a dog, but only that travellers who are travelling with dogs on the escalator must carry them:

(9) Dogs must be carried.

It is not clear why the psychological processes involved in accessing and using contextual assumptions for the interpretation of isolated utterances like (9) and the principles governing those processes should be different from the ones involved in the interpretation of utterances which are part of a text.

As we have seen, repetitions are analyzed in a coherence approach in terms of a structural relation between adjacent segments. However, as Wilson (1995) points out, it is not clear how this approach would analyze repetitions which are not adjacent or repetitions in one-clause utterances like (10):

(10) That was a really really stupid thing to do.

Wilson argues that since there is no obvious intonation break in (10), it could only be treated as a two-clause utterance because it is a repetition. Clearly, it would be more satisfactory to have an analysis which covers *all* cases of repetition whether or not they occur in adjacent segments.[3] However, as Blass (1990) has pointed out, an utterance may be part of an incoherent discourse, but still be understood by a hearer. For example, B's response in (11) has both a coherent interpretation in which it reports what she said, and an incoherent interpretation in which it describes what B has just seen:

(11) *A*: What did she say?
B: That man has a gun.

Some writers (for example, Tsui 1991) analyze interruptions as violations of a "coherence rule" which, unless they are justified, result in antisocial and impolite behavior. This raises the question of what justifies such violations. Giora (1996), who, unlike many coherence theorists, does not believe that coherence is analyzable in terms of a set of local coherence relations, regards the incoherent interpretation of B in (11) as unacceptable because it violates a "Relevance Requirement" (not to be confused with Sperber and Wilson's Principle of Relevance) which requires that all the propositions of a well-formed discourse be related to a discourse-topic proposition. However, she also suggests that violations from this requirement are acceptable provided that they are explicitly marked by an expression such *as by the way*, *incidentally*. However, as Wilson (1998) has shown, this would rule out B in (11) (which is not explicitly marked) but allow something like (12):

(12) *A*: What's the time?
B: By the way, Tutankhamen ate my dog. (example from Wilson 1998)

As we shall see in the following section, the interpretation of interruptions can be explained in terms of the notion of optimal relevance and the criterion of consistency with the principle of relevance. More generally, as Wilson (1998) points out, Giora's notion of topic relevance can be shown to be derivative in a relevance theoretic account. It is generally agreed that the function of the discourse topic is to provide access to contextual information required for comprehension. However, as Sperber and Wilson (1986) show, it is contextual information rather than the discourse topic that is essential for comprehension: on the one hand, a text may be comprehensible even where there is no explicitly stated topic, and, on the other hand, it may remain incomprehensible even where there is an explicitly stated topic.[4]

Although Giora (1996) does not analyze coherence in terms of local coherence relations between the segments of a text, she does recognize that we have intuitions about the way in which adjacent segments are related. And, indeed, there is no question that we are capable of recognizing coherence relations like restatement, elaboration, sequence. The question is whether these relations actually are computed in the course of utterance comprehension. In this section, we have seen that a coherence-based approach lacks the generality required for an account of comprehension which covers all utterances. In the following section I shall show that in a relevance theoretic framework the computation of coherence relations is not only unnecessary, since they can be derived as a consequence of the hearer's search for relevance, but may also be inconsistent with the Principle of Relevance.

2 Relevance and Coherence

The assumption that an utterance is consistent with the Principle of Relevance is based on the hearer's recognition that it is an act of ostensive communication – that is, an act of deliberate, overt communication in which the speaker not only intends to convey a particular message but is also actively helping the hearer recognize this. From the speaker's point of view, it is simply not worth engaging in such an act unless the audience pays attention to it. But equally, from the hearer's point of view it is not worth paying attention to an act of communication unless there is information worth processing – or in other words, unless it is relevant. This means that a speaker who requests the hearer's attention, for example by producing an utterance, communicates his or her assumption that his or her utterance is relevant.[5]

Relevance is defined in terms of *contextual effect* and *processing effort*. Contextual effects are simply the ways in which a new piece of information may interact with contextual assumptions to yield an improvement to the hearer's overall representation of the world. These are not confined to new assumptions derived from combining the new information with contextual assumptions, but may also include increased evidence for existing assumptions or even the elimination of existing assumptions. Processing effort is a function not only of the linguistic complexity of the utterance itself, but also of the cost of accessing and using contextual assumptions in the derivation of contextual effects.

Sperber and Wilson (1986) argue that the presumption of relevance carried by every act of ostensive communication has two aspects: first, it creates a presumption that the information it communicates interacts with the context for derivation of adequate *contextual effects*; and second, it creates a presumption that no gratuitous processing effort is required for the recovery of effects. Taken together, these presumptions define a level of *optimal relevance*. And the principle of relevance is simply the thesis that every act of ostensive communication communicates a presumption of its own optimal relevance.[6]

This is not to say that every act of ostensive communication is in fact optimally relevant. Suppose you grab my arm and point to the clock, which is now showing 3 o'clock. If I have seen the clock, then the presumption of optimal relevance communicated by your behavior is false. However, your behavior is still consistent with the principle of relevance inasmuch as it is not difficult for me to see how you *thought* it was optimally relevant.

Nor is it to say that the intended interpretation is always recovered. The Principle of Relevance does not guarantee that communication will succeed. Suppose, for example, the interpretation of A's utterance in (13) provides an immediately accessible context which, taken together with B's answer, yields adequate contextual effects. The resulting coherent interpretation, in which B's utterance is a report of what Jane said, will then be justified under the principle of relevance. However, it may not have been the one intended, and communication may fail:

(13) *A*: What did Jane say?
B: It's 3 o'clock.

On the other hand, the coherent interpretation is not the only interpretation which might be justified under the principle of relevance. Obviously, the answer to A's question is relevant to him. Why otherwise would he have asked it? However, it is not difficult to imagine how B's utterance might trigger an immediately accessible context in which the information that it is 3 o'clock yields contextual effects. For example, A and B may have been planning to catch a train which leaves just after 3. They can always continue their discussion of Jane on the train. But the train will not wait.

In this case, B's utterance is processed for relevance in a context which is distinct from the one in which A's utterance is interpreted. There are no contextual assumptions used in the interpretation of B's utterance that are used in the interpretation of A's. Moreover, the contextual assumptions used in establishing the relevance of B's utterance do not include the content of A's utterance or any contextual effects derived from it. This, argues Blass, is the source of the incoherence. Putting this the other way round, if a discourse *is* coherent, then it is because there is continuity of context in the sense that assumptions made accessible by the interpretation of one segment are used in establishing the relevance of the next. Since the interpretation of information which has just been processed will provide a highly accessible context for the interpretation of an utterance, coherence can be regarded as a consequence of the hearer's search for optimal relevance.

If this is right, then it ought to be possible to show how particular coherence relations can be reanalyzed in terms of a consequence of the way relevance is established. This is the aim of the following section, where I shall outline a relevance theoretic reassessment of so-called sequential relations, and, then, some of the subtypes of elaboration. However, first, let us see how the claim that computation of coherence relations are necessary for comprehension would have to be justified in a framework which assumes that comprehension is constrained by the Principle of Relevance.

On the assumption that understanding an utterance is a matter of recovering its *explicatures* (or intended explicit content) and the *contextual effects* that the hearer is intended to derive from those explicatures, the claim that the computation of coherence relations is necessary for comprehension amounts to the claim that their identification is necessary for the recovery of an utterance's explicatures and intended contextual effects. If the identification of this relation is not necessary for the recovery of adequate contextual effects, then the effort required for its identification would be gratuitous, and would be ruled out by the second clause of the definition of optimal relevance (above). In other words, in a relevance theoretic framework a coherence relation should never be computed unless its identification contributes to adequate contextual effects.[7]

3 The Reassessment of Coherence Relations

3.1 Sequence

In a coherence-based framework the interpretation of sequences like the ones in (14) and (15) involve the identification of relations of temporal and causal sequence respectively:

(14) a. A number 16 bus finally arrived.
b. I asked the driver whether he was going to the university.

(15) a. The number 16 bus was half an hour late.
b. I missed most of the syntax lecture.

However, if Carston's (1993) analysis of these sequences is correct, these relations are a consequence of the way in which the hearer of these utterances uses contextual information to develop the linguistically determined semantic representation into a proposition which can achieve optimal relevance.[8] Her argument is that the linguistically determined semantic representation of these utterances underdetermines their propositional content in just the same way as the linguistic meaning underdetermines the explicit content of utterances like those in (16):

(16) a. It's too hot.
b. Too hot.

In order to recover a proposition which can achieve optimal relevance, the hearer must use contextual information to recover the reference of whatever is too hot, the intended sense of *hot* and the identity of what it is too hot for. Similarly, the search for optimal relevance will lead the hearer of (14) and (15) to use contextual assumptions in the recovery of the enriched propositional forms in (17) and (18) respectively:

(17) a. A number 16 bus finally arrived at time t_n.
b. At time t_{n+1} I asked the driver whether he was going to the university.

(18) a. [The number 16 bus was half an hour late]$_i$.
b. As a result of that$_i$ I missed most of the syntax lecture.

As Carston (1993) points out, this interpretation can be explained in terms of the fact that ready-made scripted knowledge makes the contextual assumptions that give rise to it highly accessible. However, she also points out that this cannot be the whole story, since there is a range of other cases in which sequential or causal enrichment cannot be a result of ready scripted knowledge – for example, (19) – and, moreover, a range of cases in which a sequential interpretation is not necessarily recovered at all – for example (20):

(19) John broke his leg and skied over a precipice.

(20) a. John broke his leg.
b. He skied over a precipice. (examples from Smith 1990)[9]

Carston's suggestion that the causal interpretation of sequences like (19) can be seen as "the product of some quite general cognitive predisposition to forge certain connections and relations between states and events whenever it is reasonable to do so" (1993: 33) is not to be construed as a suggestion that hearers have a cognitive tendency to compute coherence relations in the course of comprehension, but rather that their information-processing capacities and their tendency to optimize relevance leads the hearer to enrich the linguistically determined semantic representation so that the proposition expressed has the sort of form given in (18). Once the hearer has recovered this proposition there is no justification (under the Principle of Relevance) for recovering a further proposition that a particular coherence relation holds (cf. Mann and Thompson 1987).

3.2 Explanation

However, as Carston (1993) recognizes, it still has to be explained why this causal interpretation is not necessarily recovered in nonconjoined utterances like (20). Carston's explanation of the difference between (19) and (20) hinges, first, on the claim that since we are "question-asking, explanation-seeking creatures" (1993: 38), our search for optimal relevance in a sequence in which the speaker has presented a fact involves asking "Why?," and, second, on the fact that a conjunction is a syntactic unit and hence a unit of relevance. If the first segment of (20) raises the question "Why?," then the second will achieve optimal relevance in virtue of answering that question. Once again, since the hearer has recognized that this is how the utterance achieves relevance, there is no justification for recovering the information that it stands in a particular coherence relation.

As Carston points out, this explanation is not restricted to the interpretation of utterances which follow an utterance that raises an implicit "Why?" question. In the following examples, which in a coherence framework would be analyzed in terms of elaboration, the (b) segments seem to answer implicit "Where?" and "Who?" questions:

(21) a. I ate at a good restaurant last week.
b. It was McDonald's.

(22) a. I met a great actress at the party.
b. It was Vanessa Redgrave. (examples due to Deirdre Wilson)

Questions and answers are by their very nature planned as separate utterances each carrying the presumption of relevance individually. However, as Carston argues, the fact that a conjoined utterance like (19) is a single syntactic unit means that it is a single processing unit which is interpreted for relevance as a whole. For Carston, this follows from syntactic considerations, in particular, the assumption that an utterance unit is in correspondence with a grammatical unit. However, I have argued (Blakemore 1987: 120) that this follows from relevance theoretic considerations: the processing effort that follows from the extra lexical and syntactic structure involved in conjoining can only be offset if the conjoined proposition carries the presumption of relevance. This suggests that both syntactic and pragmatic considerations could support an explanation of why the second conjunct of a conjoined utterance can never be interpreted as an answer to an implicit question raised by the first conjunct.

3.3 Exemplification

As Carston points out, these sorts of considerations also explain the interpretive difference between examples like the ones in (23a) and (b):

(23) a. The buses never arrive on time these days. Yesterday I waited 20 minutes for the number 16.
b. The buses never arrive on time and yesterday I waited 20 minutes for the number 16.

On the assumption that "exemplification is a common way of providing evidence for a claim or, equivalently, giving a reason for believing something" (Carston 1992: 11), then it is not surprising that only the juxtaposed sequence in (23a) can be interpreted as a claim and exemplification. For to present a claim and then to present evidence for it is to present two utterances each of which carries the presumption of relevance individually. But why should exemplification be a means of providing evidence for a claim?

In Blakemore (1997) I argue that the answer to this question lies in the fact that once the state of affairs described by the speaker is recognized as an example, there is an expectation that it is typical in some respect and hence that there are other states of affairs which the speaker could have cited. For to say that there are a number of buses which are like the speaker's bus in virtue of their lateness is to provide support for the generalization that buses never arrive on time these days. It is possible for the speaker to strengthen her evidence by citing more examples. However, if these are recognized as examples, then no matter how many cases are cited, it will always be understood that there are others. In other words, it is the suggestion that there are other cases which could have been cited which makes exemplification such a good means of providing evidence for the claim exemplified.

This argument would seem to suggest that the hearer must recognize that an utterance is intended as an exemplification before he or she can understand it, for the assumption that the state of affairs is an example plays a central role in the recovery of its contextual effects. And indeed, it seems that a speaker who questions or denies the assumption that an utterance is an exemplification also questions or denies its intended contextual effects. For example, in the following, which is based on a radio interview (Radio 4, 12 August 1997), B is denying that the second segment of A's utterance provides support for the first by denying that it is an example.

(24) *A:* There seems to be something really wrong with the army. I assume you know about those soldiers who smashed up their hotel room in Uruguay?
B: Yes, it was disgraceful, but it was just one isolated and very atypical incident.

However, it does not follow from this that the computation of coherence relations is essential for the comprehension of discourse. It only shows that the assumption that the utterance is intended as an exemplification is recovered because it contributes to the recovery of adequate contextual effects. Moreover, the crucial assumption in the inferential processes involved in establishing the relevance of the utterance is not so much the assumption that it is connected to the preceding text in a particular way as the assumption that the state of affairs it represents is typical in a particular respect. The role played by the interpretation of the preceding utterance is to give the hearer access to contextual assumptions which enable him or her to identify this respect.

3.4 Restatement

In my recent work on reformulations and reformulation markers (Blakemore 1993, 1994, 1997) I have argued that reformulations are on example of the way that utterances may be relevant as representations of utterances which they resemble. As Sperber and Wilson (1986) point out, all sorts of phenomena can be used as representations in this way; for example, pictorial representations and mimes. Of course, no two phenomena are exactly alike, and a communicator expects the hearer to identify the respects in which the resemblance holds. In the case of an utterance which is used to represent another, the resemblance may hold in virtue of resemblances in phonetic and phonological form, or resemblances in lexical and syntactic form, or resemblances in logical properties. For example, all the utterances in (27) could be produced as answers to (26) in a situation in which the director had produced the utterance in (25):

(25) We will have to let her go.

(26) What did the director say?

(27) a. We will have to let her go.
b. They'll have to let her go.
c. She's fired.

(27a) is a direct quotation and represents the director's utterance in virtue of resemblances in linguistic and semantic structure. (27b) has a different semantic structure (since it uses the third person pronoun instead of the original first person pronoun), but the two utterances share a common propositional form. (27c) has neither the same linguistic structure nor the same propositional form as the original. However, its propositional form may still be said to *resemble* the propositional form of the original in the sense that it is not difficult to imagine a context in which it gives rise to the same contextual implications. In such cases where the resemblance involves the sharing of logical and contextual implications, Sperber and Wilson say that the utterance can be said to be relevant as an *interpretation* of a propositional form or thought.

A speaker who produces an utterance which is relevant as a representation of another utterance cannot be taken to be creating expectations of truthfulness since she or he is not using that utterance *descriptively*. She or he can only be taken to be creating expectations of *faithfulness*. Faithfulness is a matter of degree, the degree of faithfulness being determined by the extent to which the two propositional forms share logical and contextual implications, and the degree of faithfulness attempted will vary from situation to situation. Thus in (7), repeated here as (28), the second segment achieves the same contextual effects as the first:

(28) a. At the beginning of this piece there is an example of an anacrusis.
b. That is, it begins with a unaccented note which is not part of the first full bar. (Blakemore 1997)

Since the speaker is restating his own utterance, he is as committed to the factuality of the reformulation as he is to the original. However, the main point of utterance lies in the fact that it is a faithful interpretation of the preceding segment.

In an unplanned discourse, an utterance like (28b) would be justified under the Principle of Relevance if it followed the speaker's recognition that he had made a miscalculation of the hearer's contextual and processing resources, and that the original did not in fact achieve optimal relevance. However, sequences like (28) may also be part of a planned discourse. Why would a speaker aiming at consistency with the Principle of Relevance deliberately produce *both* the original and the reformulation if the second segment alone would have achieved the same contextual effects for less processing effort?

The use of a term with which the hearer is assumed to be unfamiliar and then its reformulation is characteristic of what might be called a pedagogical style, which itself can be justified in terms of the Principle of Relevance. For the speaker can be taken to communicate not just the information about the beginning of the piece of music, but also information about the term *anacrusis*. The assumption that it is relevant to teach the hearer what the term means by reformulating it is based on an assumption about the hearer's processing resources, and clearly a miscalculation here would result in a patronizing style.

A rather different effect is achieved in Mann and Thompson's example, repeated here as (29), which I assume is an advertisement for car polish:

(29) a. A WELL-GROOMED CAR REFLECTS ITS OWNER.
b. The car you drive says a lot about you. (Mann and Thompson 1988)

The pun in the first segment captures the hearer's attention by presenting her with a sort of puzzle: the speaker could mean either that one's reflection shows on a well-groomed, shiny car or that owning a well-groomed car is evidence for being a well-groomed, smart kind of person. The second segment is an interpretation of only the second proposition and in this sense could be regarded as providing a solution to the puzzle posed by the pun, or, in other words, a means of constraining the hearer's interpretation of the first segment. However, the second segment alone would not have captured the hearer's attention in the way that the first segment does. Nor would it have yielded contextual effects about the shiny qualities of well-groomed cars. This means that although the interpretation of the first segment entails processing costs not entailed by the second segment, this effort is offset by, first, the way it captures the attention of the hearer and, second, contextual effects which would not have been achieved by the second segment alone.

These analyses have described (28b) and (29b) as reformulations. However, I have argued that this description must itself be analyzed in terms of the notion of interpretive representation. The question of whether an utterance is relevant as an interpretation (rather than a description) is not a question about how it is connected to the preceding text, but a question about the relationship between the proposition it expresses and the thought it represents. As Sperber and Wilson (1986, 1985/6) and Wilson and Sperber (1992) have shown, the notion of interpretive representation is involved in the analysis of a range of phenomena; for example, reported speech, free indirect speech, interrogatives, irony, and metaphor. In some cases, an utterance may be relevant as an interpretation of a thought that has been communicated by an utterance that is not part of a continuous text, and in other cases it may be relevant as an interpretation of a thought that has not been communicated at all. Indeed, according to relevance theory, the identification of an utterance as a reformulation follows from an aspect of interpretation which is fundamental to the way in which the relevance of all utterances is established, and will not itself contribute to the identification of contextual effects. This is not to say that a hearer, or, indeed, an analyst, will not describe the utterance as a reformulation. The point is that such a description is a consequence of the recognition that the utterance is an instance of interpretive rather than descriptive language use.

4 Implications for Discourse Understanding

In this chapter I have focused on an approach to discourse which assumes that discourse coherence provides the key to a theory of discourse comprehension, and have shown how in a relevance theoretic framework hearers' intuitions about coherence can be explained as a consequence of the hearer's search for an interpretation that is consistent with the Principle of Relevance. However, work in relevance theory is not just concerned with the reassessment of coherence relations. It has also shown how the notion of optimal relevance can be used to explain those aspects of comprehension which are claimed to be a consequence of the search for discourse coherence.

For example, recently Wilson and Matsui (1998) have compared the predictions made by Asher and Lascarides's (1995) coherence-based heuristics for disambiguation in discourse with those made by relevance theory. Whereas relevance theory claims that the same criterion of consistency with the Principle of Relevance explains disambiguation in both isolated utterances and extended texts, Asher and Lascarides's heuristics are designed to *supplement* the word-association heuristics given in the artificial intelligence literature for disambiguating isolated utterances. Wilson shows that neither the heuristics for isolated utterances nor the heuristics for discourse make the correct predictions and argues that disambiguation phenomena are more satisfactorily explained in terms of the notion of optimal relevance.

The criterion of consistency with the Principle of Relevance also provides a unitary explanation for the assignment of reference in isolated utterances and discourse sequences such as (30) (from Wilson 1992):

(30) Sean Penn attacked a photographer. *The man* was quite badly hurt.

While it is often claimed that reference resolution is affected by the relative accessibility of the candidate referents, it is also agreed that an account based on accessibility alone would make the wrong predictions. For example, Herb Clark (1977) proposes that reference assignment in examples like (30) is affected by the number and plausibility of the assumptions needed to introduce the intended referent; but as Wilson (1992) and Matsui (1993, 1995) show, this proposal does not deal with all examples. Candidate referents must also be evaluated in terms of a pragmatic criterion that the overall interpretation is supposed to meet.

However, Wilson and Matsui (1998) have shown that neither the attempts to define such a criterion in terms of truth (cf. Lewis 1979; Sidner 1983) nor the attempts to develop a coherence-based criterion (cf. Hobbs 1979; Fox 1987) explain reference resolution in all cases. Moreover, a criterion which is powerful to choose among the various interpretations of an utterance on either of these grounds could do so only by considering them all. As Wilson (1995) says, this "would create a combinatorial explosion of gigantic proportions, and be quite unlike what hearers actually do." Her relevance theoretic analyses of examples that are problematic for both truth- and coherence-based accounts show that what hearers actually do is to accept the first interpretation that is consistent with the Principle of Relevance and that the speaker could have manifestly foreseen.

Within coherence-based approaches to discourse, expressions like utterance-initial *so*, *well*, *still*, *after all* are classified as *discourse markers*, a term which is intended to reflect the role that these expressions play in *marking*, *signaling*, or *indicating* how one unit of discourse is connected to another (cf. Levinson 1983: 87–8; Fraser 1990; Mann and Thompson 1987; Sanders et al. 1993; Knott and Dale 1994).[10] Since relevance-based approaches are concerned with processes of utterance understanding rather than the structure of discourse, and appeal to contextual effects rather than coherence relations, it is not surprising that relevance theoretic analyses of these expressions are significantly different from coherence-based ones.

For example, whereas Sanders et al. analyze *but* as an explicit guide to a range of coherence relations (namely, Contrast, Antithesis, Contrastive Cause–Consequence), my 1987 analysis treats *but* as an expression which constrains the interpretation process by narrowing down the search for the intended contextual effects. Thus while this analysis, like Sanders et al.'s, treats *but* as expressing either contrast or denial of expectation (cf. Lakoff 1971), it does this not by analyzing it in terms of a marker of coherence relations, but by analyzing it as an instruction for the recovery of contextual effects.[11]

The analysis of a discourse marker as an expression which links units of discourse would seem to imply that it cannot be used discourse initially. However, as the examples in (31–2) show, this is clearly not the case:

(31) (speaker sees hearer come in laden with shopping) So you've spent all your money.

(32) (speaker takes an enormous slice of cake) After all, it is my birthday.

If these expressions connect an utterance and a *context*, as my 1987 analysis suggests, this is not surprising. For while contextual assumptions may be derived from the preceding discourse, they may also be derived from the hearer's perception of the environment from memory. Not all discourse markers can be used discourse-initially, of course. However, as Blakemore (1998) shows, this can be explained in terms of the particular constraint that the expression imposes.

Within coherence-based approaches, discourse markers are said to have a *pragmatic meaning* on the grounds that they do not contribute to the truth conditions of the utterance that contains them. The relevance theoretic analysis I have described was an attempt to provide an explanation of the distinction between truth-conditional and nontruth-conditional meaning in terms of the cognitively motivated distinction between conceptual and procedural meaning. However, recent work within relevance theory has shown that the conceptual–procedural distinction is not coextensive with the truth-conditional–nontruth-conditional distinction, and that in particular there are discourse connectives which, although they do not contribute to truth conditions, nevertheless encode concepts (cf. Wilson and Sperber 1993; Blakemore 1996, 1997; Ifantidou-Trouki 1993). For example, in contrast with expressions like *but* and *well*, the so-called apposition marker *in other words* is both nontruth-conditional and conceptual.

Sperber and Wilson's (1995) speculation that the conceptual–procedural distinction will shed more light on linguistic semantics than the traditional distinction between truth-conditional and nontruth-conditional meaning provides an exciting agenda for future semantics research. Since expressions classified as discourse markers may encode either conceptual or procedural meaning, it seems that they will have an important part in this research.[12] At the same time, a relevance theoretic analysis of these expressions will play a significant role in showing how the approach I have outlined in this chapter can offer more insight into the psychological processes underlying discourse understanding than can an approach which analyses them as expressions which link units of discourse.

5 Conclusion

In this chapter, I have shown that according to relevance theory, discourse understanding is not a by-product of discourse acceptability or coherence, and that our intuitions about the coherence of discourse are a consequence of our search for relevance. However, neither the relevance theoretic reassessment of coherence relations in section 3 nor the reanalysis of discourse phenomena in section 4 should be taken as an argument that we should simply replace talk of coherence relations by talk of "relevance relations." Coherence is a property of an object external to the human mind and is defined in terms of structural relations between subunits of that object. Relevance is a property of a mentally represented interpretation of the evidence a communicator provides for the thought(s) she or he intends to communicate, and is defined in terms of a function of the effects this interpretation has on the hearer's overall representation of the world and the effort that is needed for its derivation.

NOTES

1 For definitions of elaboration, see Hobbs (1979, 1983), Mann and Thompson (1987), and Hovy and Maier (1994).
2 For example, Samet and Schank (1984) propose that although local coherence must be defined in terms of coherence relations, global coherence must be analyzed in terms of stereotypic scripts and goals. Others, for example, Reinhart (1980), Giora (1996), and Sidner (1983), adopt a more functional approach and propose that coherence should be defined in terms of relevance to a discourse topic.
3 For a relevance theoretic analysis of repetitions, see Sperber and Wilson (1986). If the recognition of coherence relations is necessary for comprehension, then it would seem to follow that only coherent discourses are comprehensible.
4 For further discussion, see Sperber and Wilson (1987: 742).
5 Expository articles on Relevance Theory include Blakemore (1988b, 1995); Carston (1988, 1993); Smith and Wilson (1992); Wilson (1994); Wilson and Sperber (1986). For a précis of *Relevance*, see Sperber and Wilson (1987). For a book-length introduction, see Blakemore (1992).
6 This principle is what Sperber and Wilson (1995: 260–72) call *the communicative principle of relevance* and must be distinguished from *the cognitive principle of relevance*, which states that human cognition tends to be geared to the maximization of relevance. As Wilson (1998) points out, the confusion between these two principles has lead to misunderstandings about how relevance theory works (see for example, Giora 1996).
7 Unger (1986) makes a similar point.
8 See also Wilson and Sperber (1998).
9 Smith (1990) uses similar arguments against the view that a notion of narrative tense is necessary to account for the interpretation of narrative sequences.
10 Schiffrin's (1987) analysis of discourse markers is grounded in a more functional approach to discourse which assumes that language is designed for communication and attempts to show how their use is a consequence of structural, semantic, and pragmatic factors. In contrast with the approaches mentioned here, she argues that they play a role in establishing discourse coherence not just at a local level, but also from a global level. However, it should be noted that in contrast with relevance theoretic analyses, her analysis treats a marker like *so* as linking either ideas, premise, and conclusion in inference *or* acts of communication.
11 For other relevance theoretic analyses of discourse markers, see Blakemore (1988a); Blass (1990); Higashimori (1994); Itani (1993); Jucker (1993); Moeschler (1989, 1993); Rouchota (1998); Unger (1996). Ducrot (1984) has also developed a procedural approach to the analysis of discourse markers, but not from within a relevance theoretic framework.
12 For further discussion of this issue, see Blakemore (1997).

REFERENCES

Asher, N. and Lascarides, A. (1995) Lexical disambiguation in a discourse context. *Journal of Semantics*, 12. 69–108.

Blakemore, D. L. (1987) *Semantic Constraints on Relevance*. Oxford: Blackwell.

Blakemore, D. L. (1988a) *So* as a constraint on relevance. In R. Kempson (ed.) *Mental Representations: The Interface between Language and Reality*. Cambridge: Cambridge University Press. 183–96.

Blakemore, D. L. (1988b) The organization of discourse. In F. Newmeyer (ed.) *Linguistics: The Cambridge Survey* vol. 4. Cambridge: Cambridge University Press. 229–50.

Blakemore, D. L. (1992) *Understanding Utterances*. Oxford: Blackwell.

Blakemore, D. L. (1993) The relevance of reformulations. *Language and Literature*, 2(2). 101–20.

Blakemore, D. L. (1994) Relevance, poetic effects and social goals: a reply to Culpeper. *Language and Literature*, 3(1). 49–59.

Blakemore, D. L. (1995) Relevance theory. In J. Verschuerne, J-O. Ostman, and J. Blommaert (eds), *Handbook of Pragmatics*. Amsterdam: John Benjamins. 443–52.

Blakemore, D. L. (1996) Are apposition markers discourse markers? *Journal of Linguistics*, 32. 325–47.

Blakemore, D. L. (1997a) Restatement and exemplification: a relevance theoretic re-assessment of elaboration. *Pragmatics and Cognition*, 5(1). 1–19.

Blakemore, D. L. (1997b) Non-truth conditional meaning. *Linguistische Berichte*, 8. 92–102.

Blakemore, D. L. (1998) On the context for so-called "discourse markers." In K. Malmkjaer and J. Williams (eds) *Context in Language Learning Understanding*. Cambridge: Cambridge University Press. 44–59.

Blass, R. (1990) *Relevance Relations in Discourse*. Cambridge: Cambridge University Press.

Carston, R. (1988) Language and cognition. In F. Newmeyer (ed.) *Linguistics: The Cambridge Survey* vol. III. Cambridge: Cambridge University Press. 38–68.

Carston, R. (1992) Conjunction, explanation and relevance. Unpublished ms. Revised (1993) version in *Lingua*, 90(1/2).

Carston, R. (1993) Conjunction, explanation and relevance. *Lingua* 90(1/2). 23–48.

Chomsky, N. (1986) *Knowledge of Language*. New York: Praeger.

Clark, H. (1977) Bridging. In P. Johnson-Laird and P. Wason (eds) *Thinking: Readings in Cognitive Science*. Cambridge: Cambridge University Press. 411–20.

Ducrot, O. (1984) *Le dire et le dit*. Paris: Minuit.

Fasold, R. (1990) *Sociolinguistics of Language*. Oxford: Blackwell.

Fox, B. (1987) *Discourse Structure and Anaphora*. Cambridge: Cambridge University Press.

Fraser, B. (1990). An approach to discourse markers. *Journal of Pragmatics*, 14. 383–95.

Giora, R. (1996) Discourse coherence and theory of relevance: stumbling blocks in search a unified theory. *Journal of Pragmatics*, 27. 17–34.

Halliday, M. A. K. and Hasan, R. (1976) *Cohesion in English*. London: Longman.

Harris, Z. (1951) *Methods in Structural Linguistics*. Chicago: University of Chicago Press.

Higashimori, I. (1994) A relevance theoretic analysis of *even*, *sae/sura/mo/*

temo/ddemo/datte/made. *English Literature Review* (Kyoto Women's University) 38. 51–80.

Hobbs, J. (1979) Coherence and coreference. *Cognitive Science*, 3. 67–90.

Hobbs, J. (1983) Why is discourse coherent? In F. Neubauer (ed.), *Coherence in Natural Language Texts*. Hamburg: Buske. 29–70.

Hovy, E. and Maier, E. (1994) Parsimonious or profligate: how many and which discourse structure relations? Unpublished.

Ifantidou-Trouki, E. (1993) Sentential adverbs and relevance. *Lingua*, 90(1/2). 65–90.

Itani, R. (1993) The Japanese sentence-final particle *ka*: a relevance theoretic approach. *Lingua*, 90(1/2). 129–47.

Jucker, A. (1993) The discourse marker *well*: a relevance theoretic account. *Journal of Pragmatics*, 19. 435–52.

Knott, A. and Dale, R. (1994) Using a set of linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, 18(1). 35–62.

Lakoff, R. (1971) Ifs ands and buts about conjunction. In C. J. Fillmore and D. T. Langendoen (eds) *Studies in Linguistic Semantics*. New York: Holt Rinehart and Winston. 115–50.

Levinson, S. (1983) *Pragmatics*. Cambridge: Cambridge University Press.

Lewis, D. (1979) Scorekeeping in a language game. In R. Bauerle et al. (eds) *Semantics from Different Points of View*. Berlin: Springer. 172–8.

Mann, W. C. and Thompson, S. (1987) Relational propositions in discourse. *Discourse Processes*, 9. 57–90.

Mann, W. C. and Thompson, S. (1988) Rhetorical structure theory: towards a functional theory of text organization. *Text*, 8(3). 243–81.

Matsui, T. (1993) Bridging reference and the notions of topic and focus. *Lingua*, 9(1/2). 49–68.

Matsui, T. (1995) Bridging and Reference. University of London, PhD thesis.

Moeschler, J. (1989) Pragmatic connectives, argumentative coherence and relevance. *Argumentation*, 3.3. 321–39.

Moeschler, J. (1993) Relevance and conversation. *Lingua* 90. 1/2. 149–71.

Reinhart, T. (1980) Conditions for text coherence. *Poetics Today*, 1(4). 161–80.

Rouchota, V. (1998). Procedural meaning and parenthetical discourse markers. In A. Jucker and Y. Ziv (eds) *Discourse Markers: Description and Theory*. Amsterdam: John Benjamins. 97–126.

Salkie, R. (1995) *Text and Discourse Analysis*. London: Routledge.

Samet, J. and Schank, R. (1984) Coherence and connectivity. *Linguistics and Philosophy*, 7(1). 57–82.

Sanders, T., Spooren, W., and Noordman, L. (1993) Towards a taxonomy of coherence relations. *Discourse Processes*, 15(1). 1–36.

Schiffrin, D. (1987). *Discourse Markers*. Cambridge: Cambridge University Press.

Sidner, C. (1983) Focusing and discourse. *Discourse Processes*, 6. 107–30.

Smith, N. (1990) Observations on the pragmatics of tense. *UCL Working Papers in Linguistics*, 2. 113–46.

Smith, N. and Wilson, D. (1992) Introduction to the special issue on relevance theory. *Lingua*, 87(1/2). 1–10.

Sperber, D. and Wilson, D. (1985/6) Loose talk. *Proceedings of the Aristotelian Society*, LXXXVI. 153–71.

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

Sperber, D. and Wilson, D. (1987) Presumptions of relevance. *Behavioural and Brain Sciences*, 13(1). 736–54.

Sperber, D. and Wilson, D. (1995) *Relevance* (2nd edition). Oxford: Blackwell.

Tsui, A. (1991) Sequencing rules and coherence in discourse. *Journal of Pragmatics*, 15. 111–29.

Unger, C. (1996) The scope of discourse connectives: implications for discourse organization. *Journal of Linguistics*, 32(2). 403–38.
Wilson, D. (1992) Reference and relevance. *UCL Working Papers in Linguistics*, 4. 165–91.
Wilson, D. (1994) Relevance and understanding. In G. Brown, K. Malkmjaer, A. Pollit, and J. Williams (eds), *Language and Understanding*. Oxford: Oxford University Press. 35–58.
Wilson, D. (1995) Issues in pragmatics. Unpublished lectures.
Wilson, D. (1998) Discourse, coherence and relevance: a reply to Rachel Giora. *Journal of Pragmatics*, 29. 57–74.
Wilson, D. and Matsui, T. (1998) Recent approaches to bridging: truth coherence and relevance. *University of London Working Papers in Linguistics*, 10. 173–200.
Wilson, D. and Sperber, D. (1986) Pragmatics and modularity. Reprinted in S. Davis (ed.) 1991, *Pragmatics: A Reader*. Oxford: Oxford University Press. 377–93.
Wilson, D. and Sperber, D. (1992) On verbal irony. *Lingua*, 87(1/2). 53–76.
Wilson, D. and Sperber, D. (1993) Linguistic form and relevance. *Lingua*, 90. 5–25.
Wilson, D. and Sperber, D. (1998) Pragmatics and time. In R. Carston and S. Uchida (eds) *Relevance Theory:Implications and Applications*. Amsterdam: John Benjamins. 1–22.