Understanding Causal Coherence Relations
Understanding Causal Coherence Relations

Het begrijpen van causale coherentierelaties
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. J.C. Stoof,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen
op vrijdag 1 februari 2008 des middags te 12.45 uur.

door

Gerben Mulder

geboren op 11 maart 1975 te Ede
Promotoren: prof. dr. T.J.M. Sanders
prof. dr. H. van den Bergh
# Table of Contents

Acknowledgements ................................................................................................................... 7

Chapter 1: The Processing and Representation of Causal Coherence Relations ........... 9
  1.1 Introduction and research questions ............................................................................. 9
  1.1.1 A cognitive account of coherence relations ......................................................... 9
  1.1.2 Research questions ............................................................................................. 12
  1.2 Causal coherence relations and the discourse representation .................................... 14
  1.3 Cognitive processes of discourse interpretation ....................................................... 21
    1.3.1 Discourse inference processes ........................................................................ 21
    1.3.2 The role of causal relations in discourse inference theories ................................. 24
    1.3.3 Empirical evidence for causal inferences ....................................................... 29
    1.3.4 Top-down versus bottom-up processing ....................................................... 36
  1.4 The present study ...................................................................................................... 38

Chapter 2: Multilevel Modeling of Experimental Data in Discourse Processing Studies 43
  2.1 Introduction ................................................................................................................ 43
  2.2 Analysis of Variance ................................................................................................. 44
    2.2.1 Calculating quasi F-ratios ................................................................................ 46
  2.3 Multilevel modeling ................................................................................................. 53
  2.4 Multilevel modeling of random cross classifications .............................................. 57
  2.5 Additional differences between MLM and ANOVA ............................................. 59
  2.6 Conclusion .............................................................................................................. 61

Chapter 3: The Role of Text Connecting and Extra-textual Inferences in the Processing of Causal Coherence Relations .................................................. 63
  3.1 The processing of implicit and explicit causal relations ............................................ 63
  3.2 Processing evidence for the integration and inference effect of connectives ............ 66
  3.3 The experiment ....................................................................................................... 71
    3.3.1 Method .......................................................................................................... 72
    3.3.2 Results ......................................................................................................... 75
    3.3.3 Conclusion .................................................................................................... 77
  3.4 General Discussion ................................................................................................. 78

Chapter 4: The Representation of Causal Coherence Relations in Expository Text ....... 83
  4.1 Introduction ............................................................................................................. 83
  4.2 Levels of representation of causal coherence relations .......................................... 84
  4.3 The experiment ...................................................................................................... 87
    4.3.1 Method ....................................................................................................... 87
    4.3.2 Results ....................................................................................................... 93
    4.3.3 Conclusion .................................................................................................. 94
  4.4 General Discussion ............................................................................................... 95
Acknowledgements

A few weeks ago, I saw a rerun of “de TV Show op reis”. I think Ivo Niehe was interviewing Jan des Bouvrie who told a story about a very old Chinese artist.

Once upon a time, there was an 88-year old Chinese painter who was one of the most popular artists of his days. He spent his whole life perfecting the technique of painting roosters (male chickens). And not without success either. Indeed, his rooster paintings were so popular that every rich man needed to have one of his paintings, and was willing to pay a lot of money for it too. Not having one of his paintings was seriously frowned upon, and for this reason, a certain Chinese nobleman also wanted to buy one of the artist’s paintings. The old painter agreed, took some pencils, ink, and paper and within two or three minutes he finished another painting of a rooster.

‘What,’ the nobleman cried out, ‘do you really expect me to pay this much money for only three minutes work?’

‘What do you mean, three minutes?’ the artist replied. ‘This painting took me 88 years to finish.’

In the same way, it took me almost 33 years to finish this dissertation. During these 33 years, I have met a lot of people (some of them more inspiring than others) and I would like to thank each and every one of them, because I am sure you all had your influence on me and therefore on this dissertation.

I would also like to take the opportunity to thank a few people without whom this work would never have been finished.

First of all, I would like to thank Ted Sanders. Frankly, without Ted I would never have started this research in the first place. Thank you for all your support (I admire your patience) and I am really looking forward to working with you, if you know what I mean.

I would also like to thank Huub van den Bergh for his support throughout the years and our many interesting conversations about statistics.

Also, many thanks go to Ted, Nina, Gijsbert, Pieter and Arjen for assisting me with the construction of experimental texts and the analysis of the continuation protocols in the experiment reported in chapter 5. I would also like to express my gratitude to Wilbert for helping me correct some of the writing errors in this dissertation and to Aleth and Rosie for their comments on my English.

I would also like to thank Iris and Nina for their moral support the last couple of years. Thanks for listening to me rambling on about bad research and statistics.

And of course, many thanks go to Ilse. (I do not know where to start). We will have a nice vacation this year, I promise.
Chapter 1: The Processing and Representation of Causal Coherence Relations

1.1 Introduction and research questions

1.1.1 A cognitive account of coherence relations

One of the most widely used metaphors in text comprehension research is that understanding a text means constructing a coherent representation of the information in the text (Graesser & Britton, 1996; Lorch & Van den Broek, 1997; Noordman & Vonk, 1992, 1997, 1998). Several theories claim that the coherence of a text originates from the coherence relations between the information units in the text (Hobbs, 1979; Mann & Thompson, 1986, 1988; Sanders, Spooren, & Noordman, 1993). Coherence relations are the meaning relations between the information units in the text, such as Cause-Consequence or Argument-Claim. The presence of these coherence relations distinguishes a text from a random set of sentences.

One of the starting points in this dissertation is that coherence should not be considered as a property of the text per se, but as a characteristic of the cognitive representation readers construct on the basis of the text (Sanders et al., 1993). From this point of view, constructing a coherent representation of a text means establishing the coherence relations between the information units in the text. If readers fail to combine the information units in a text by means of coherence relations, they will not be able to fully understand the text. The term information unit refers to the representation of chapters, paragraphs, sentences or clauses. Here, the information units are considered to be minimally clauses.

A distinction can be made between the coherence relations underlying text and the cohesive devices the author uses to make the relations explicit. One crucial assumption in a cognitive account of text coherence is that a text may still be perceived as coherent by the reader in absence of overt linguistic indicators of text connectedness. Although these surface realizations of coherence relations may be expected to influence the interpretation of text, they are neither a sufficient nor a necessary condition for text coherence. The following example serves to illustrate this point.

(1) Ice cream is one of the most favorite snacks in the summer. People like to keep themselves cool when temperatures are high.

This text does not contain any explicit cohesive devices. Indeed, elements of this text are only connected by means of implicit links between the concepts. For instance, “high temperatures” and “summer” are related and so are “ice cream” and “cool”. Thus, part of the connectedness of this text stems from repeated reference by means of part-whole relations. But there is more to it than this. The information in the second sentence can be interpreted as an explanation for the fact that ice cream is one of the most popular snacks in the summer. In other words, the two sentences of the text are consistent with an interpretation in terms of a Consequence-
Cause relation. However, this relation is not overtly indicated by a linguistic device. In this case, the connective *because* could have been used to make this relation explicit, as is done in example (2).

(2) Ice cream is one of the most favorite snacks in the summer, because people like to keep themselves cool when temperatures are high.

A coherence account of text structure goes beyond (earlier) theories of text connectedness that focus entirely on the surface realizations of coherence relations. For instance, Halliday and Hasan (1976) describe text connectedness by referring to explicit linguistic indicators of text structure such as coreference and conjunction. Such an account can be called a *cohesion* account of text structure (Sanders & Pander Maat, 2006). As we shall see below, this distinction between the surface code of the text (cohesion) and the underlying meaning representation (coherence) is also a crucial distinction in the field of discourse processing. Sanders et al. (1993) propose a cognitive account of coherence relations. From a cognitive perspective, their account is attractive because the set of possible coherence relations is based on only a few cognitively basic primitives. In contrast, coherence accounts such as Mann and Thompson’s (1988) Rhetorical Structure Theory, which also assumes cognitive plausibility (Mann & Thompson, 1986; Taboada, 2006) require the assumption that all relations are cognitively basic (Sanders, 1992; Sanders et al., 1992, 1993).

Sanders et al. (1993) distinguish four cognitive primitives. The combination of these four primitives determines the type of coherence relation between the information units in the text. The first primitive is the Basic Operation, which distinguishes between causal and additive operations. The Basic Operation underlying the relation is causal if an implication relation (P → Q) can be deduced between the information units. The Basic Operation is additive if at most a conjunction (P & Q) can be deduced.

The second primitive is the Source of Coherence. The Source of Coherence is semantic if the information units are related because of their propositional content. The Source of Coherence is pragmatic when a relation exists between the illocutionary meaning of one or both of the information units. In that case, coherence arises from the speech act status of the information units. Examples (3) and (4) illustrate the difference between the two sources of coherence.

(3) The cat ran away, because it was chased by a dog.
(4) The cat must have been chased by a dog, because the cat suddenly ran away.

In (3) the Source of Coherence may be considered to be semantic, because the content of the first clause describes a situation that is the consequence of the cause described in the second clause. In (4) on the other hand, the Source of Coherence may be interpreted as pragmatic, but only if (4) is taken to mean that the claim that the cat must have been chased by the dog, is supported by the argument that the cat ran away. Note that in both (3) and (4) the Source of Coherence may be analyzed as semantic as well as pragmatic. More recently, the distinction between connectives that express semantic and pragmatic relations has also been analyzed in
terms of subjectivity (e.g. objective (semantic) versus subjective (pragmatic) connectives) (cf. Pit, 2003; Pander Maat & Sanders, 2001).

The third cognitive primitive is the Order of the information units. Basic Order refers to those cases in which the order of the information units in the text follows the order of the Basic Operation. Thus, the Basic Order applies when the implication relation \( P \rightarrow Q \) is linguistically realized in two consecutive information units, where the first unit describes \( P \) and the second describes \( Q \). The order is Non-Basic when the order in the text does not match the implication relation. That is, Non-Basic order applies when the first unit denotes \( Q \) and the second describes \( P \). For example, in (3) the Order is Non-Basic, since the implication relation \( P \rightarrow Q \) can be paraphrased as \( P \) ‘the cat was chased by the dog’ \( \rightarrow \) \( Q \) ‘he ran away’. \( P \) is described in the second clause, whereas \( Q \) is described in the first. Thus, the order of events in the text does not match the order of the antecedent and consequence in the implication relation. Since additive relations are symmetric, the primitive Order does not apply to non-causal relations.

The fourth and final primitive is the Polarity of the relation. A relation is said to be positive if the two information units function in the basic operation; the polarity is negative if the negation of (one of the) information units function(s) in the basic operation. The relations underlying examples (5) and (6) respectively have positive and negative polarity.

(5) The dog chased the cat. The cat got scared.

(6) Although the dog chased the cat, the cat did not get scared.

The Basic operation underlying both (5) and (6) is causal. The implication relation is: “the dog chased the cat \( \rightarrow \) the cat got scared.” In (5) the information in the respective clauses functions in this implication relation, whereas in (6) the negation of the second clause is part of the basic operation. Therefore, Polarity is positive in the case of (5) and negative in the case of (6).

The combination of the four cognitive principles determines the type of coherence relation between the information units. For instance, in example (5) the primitives take the values +causal, +semantic, +basic order, and +polarity, a combination which constitutes a Cause-Consequence relation. In example (4) the primitives have the values +causal, -semantic, -basic order, and +polarity, a combination that can be analyzed as a Claim-Argument relation.

Although the proposed taxonomy is an attractive starting point for a cognitive theory of coherence relations, the account has, as it stands, only sketched the details of the actual cognitive processes taking place during the interpretation of coherence relations. A fully developed cognitive account of coherence relations should include both descriptive adequacy and psychological validity (Sanders, 1992). The requirement of descriptive adequacy is fulfilled when the account enables the description of the structure of all naturally occurring texts. Psychological validity, on the other hand, requires the theory to lead to clear predictions about the cognitive processes and representations involved in the comprehension of connected discourse (Sanders, 1992; Sanders et al., 1992). One of the aims of this dissertation is to contribute to the development of a cognitive theory of coherence relations by investigating the psychological status of the crucial assumptions underlying cognitive accounts of coherence relations. Thus, this dissertation primarily seeks to contribute to the psychological validity of
a cognitive account of coherence relations. The separate but related issue of descriptive adequacy will not be pursued.

A logical starting point for the further development of a psychologically valid coherence relation theory is providing explanations of comprehension phenomena in terms of coherence relations. This amounts to showing that the cognitive notion of coherence relations accounts for empirical data obtained in processing studies. To illustrate this further, I will give two straightforward examples. As was just explained, the cognitive account of coherence relations claims that the relations should be distinguished from their surface indicators. A prediction that can be derived from this claim is that the interpretation of coherence relations by the reader should not depend on the presence of an overt lexical marker of the relation. As a case in point, the presence of a connective like because should not lead to a qualitatively different interpretation of the relation: the reader will still interpret the causal relation between the events (all other things being equal) regardless of the presence of a connective. On the other hand, it may be expected that the connective speeds up the interpretation process, because identifying the coherence relation between sentences should be easier when a text contains an explicit indicator of the relation.

A second prediction than can be derived from the claim that coherence relations are cognitive entities is that the nature of the relation (for instance, causal versus additive) should affect the processing and representation of text. For example, it could be hypothesized that the inherent complexity of the relation influences processing time. For instance, Sanders et al. (1992, 1993) have argued that their primitives differ in cognitive complexity (Sanders, 2005b); for example, negative relations are considered more complex relations than positive ones, because they require an extra negative operation. Therefore, negative relations can be expected to be processed slower than positive ones. This is indeed a classical finding in psycholinguistics (Clark & Clark, 1977).

In a similar way it can be argued that memory phenomena that reflect the representation readers construct on the basis of the text can be explained in terms of coherence relations. For example, it could be hypothesized that the probability of specific textual information being remembered by a reader depends on the nature of the coherence relation between that information and the rest of the text. Of course, as we shall see below, actual processing and representation data are complex and, therefore, an account of these data in terms of coherence relations is much more challenging than these simple examples would suggest.

1.1.2 Research questions

In this thesis, a cognitive account of coherence relations is investigated by focusing on the processing and representation of causal coherence relations. The main reason for focusing on this type of relation is that numerous discourse processing studies indicate that causal relations play an important role in discourse understanding. For instance, causality accounts for certain memory phenomena: if information is connected by means of a causal relation it is remembered better than when the same information is connected by a non-causal relation (Black & Bern, 1981; Sanders & Noordman, 2000; Trabasso, & Sperry, 1985; Trabasso, &
Van den Broek, 1985; Trabasso, Secco, & Van den Broek, 1984). Also, causally related information is processed faster than non-causally related information (Haberlandt, & Bingham, 1982; Halldorson & Singer, 2002; Singer, Halldorson, Lear, & Andrusiak, 1992; Keenan, Baillet, & Brown, 1984; Myers, Shinjo, & Duffy, 1987; Sanders & Noordman, 2000). Therefore, discourse processing theories and the experimental results that corroborate these theories provide an excellent starting-point as to how to investigate the way in which the cognitive notion of coherence relations relates to processing and representation phenomena.

In this thesis, the distinction between coherence relations and their linguistic markers plays a crucial role. The first set of research questions focuses on the representation of coherence relations and their linguistic markers. As explained above, a crucial assumption is that coherence relations are part of the mental representation readers construct on the basis of the text. Although considering coherence relations as part of the cognitive representation may seem straightforward at first sight, recent developments in the field of discourse processing show that it is, in any case, in need for a additional specification. Ever since Van Dijk & Kintsch (1983) there seems to be a general agreement in discourse psychology that the mental text representation should in fact be regarded as a mental structure that consists of different levels of representation (see among many others, Fletcher, 1994; Fletcher & Chrysler, 1990; Graesser, Millis, & Zwaan, 1997; Kintsch, 1998). In general, the levels surface code, textbase, and situation model are distinguished (see section 1.2). These levels differ with both respect to the cognitive processes involved in their construction, and their memory characteristics. This widely shared belief in three distinct levels of representation implies a challenge for a cognitive theory of coherence relations to the extent that such a theory should be precise about the level of representation at which coherence relations are represented. The statement: ‘Coherence relations are part of the text representation’ is simply too general.

The first two research questions asked in this dissertation are:

RQ1: What is the representational status of causal coherence relations?
RQ2: What is the representational status of connectives as linguistic indicators of causal coherence relations?

Both RQ1 and RQ2 focus on the cognitive representation readers construct on the basis of the text. The second set of research questions focuses on the processes involved in the interpretation of coherence relations and the way in which these processes are influenced by linguistic marking of coherence relations. With respect to the interpretation process of causal relations, the discourse processing theories generally assume that the interpretation of these relations require, or are the result of, inferential processing (see among many others: Kintsch, 1998; Singer, 1994; Van den Broek, 1994). In other words, this suggests that coherence relations can be psycholinguistically characterized as (the product of) inferences. This idea will be investigated further in section 1.3. With respect to the linguistic indicators of coherence relations, it has been proposed that these guide the processing of the relations by providing the reader with processing instructions (Britton, 1994). For instance, with respect to the causal connective because, Noordman and Vonk (1997) propose that this connective has three different functions during processing. The first function is the segmentation function:
because signals to the reader how the information being processed is linguistically structured, for example, that because instructs the reader to process the clause it introduces as a subordinate clause. The second function is the integration function: because specifies to the reader that a causal coherence relation holds between the clauses and instructs the reader to integrate the clauses by means of that relation. The third function is the inference function: because instructs the reader to access the general knowledge that underlies the causal connection. The experimental evidence that supports this view will be reviewed in section 1.4. In short, both coherence relations and their linguistic markers seem to play a role during discourse processing. It seems worthwhile to investigate this role further.

The second set of research questions is as follows:

RQ3: How can the process of causal coherence relation interpretation be characterized in psycholinguistic terms?
RQ4: How can the process of the interpretation of the linguistic markers of causal coherence relations be characterized in psycholinguistic terms?

The first four research questions define the main domain of interest of this dissertation. The fifth and final question concerns a statistical issue. In section 1.4 and chapter 2 it will be argued that the current statistical practice in discourse processing studies - using separate F1- and F2-analyses in the analysis of experimental data - suffers from serious methodological and statistical flaws. This state of affairs gives rise to the fifth research question:

RQ5: What is the preferred statistical technique for the analysis of experimental data from common designs used in discourse processing studies?

In the remainder of this chapter the theoretical background central to the five research questions will be addressed. Section 1.2 concerns the first two research questions. In this section the differences between the levels of representation will be discussed and the extent to which a cognitive notion of coherence relations fits in with these differences will be investigated. Section 1.3 addresses the cognitive processes involved in the interpretation of causal coherence relations and describes the theoretical background pertinent to research questions three and four. In section 1.4 the scope of the present study will be specified further by giving a more detailed interpretation of the research questions than this introductory section allows.

1.2 Causal coherence relations and the discourse representation

In the discussion above, repeated reference was made to the discourse representation. In this section, properties of the cognitive representation of discourse will be described. Special attention is given to the question in what way causal coherence relation may be represented by the reader.
It is generally assumed that the text representation as a unified, monolithic concept, does not exist. Rather, it is conceived as a cognitive structure that consists of several levels (Graesser et al., 1997; Fletcher, 1994; Kintsch, 1998; Lorch & Van den Broek, 1997). Most researchers assume that the representation consists of three different levels. Van Dijk and Kintsch (1983) were the first to propose a three way distinction between the surface code, the propositional textbase and the situation model. The surface code is a representation of the literal wording of the text. The propositional textbase is a representation of the meaning of the text per se (Fletcher, 1994). This representational level can be seen as a network in which the propositional representations of the individual sentences are interrelated by means of referential links, and, maybe, coherence relations. The situation model is a representation of the situation that is described in the text and is sometimes also referred to as mental model (Johnson-Laird, 1983). The situation model is constructed by integrating textual information and background knowledge (Kintsch, 1988; 1998). For instance, the statements in the text may be integrated in knowledge structures like scripts or schemas (Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998) that represent the typical structure of narratives, expository texts or common events. Thus, a situation model of a text describing two characters going to a restaurant can be constructed by the reader by integrating the explicit statements in the text in an existing restaurant script. This will form a coherent representation of the text at the situational level. The propositional textbase and the situation model form the meaning representation of the text.

Given the topic of this dissertation, the question now is: at what level of representation are (causal) coherence relations represented? In this study it is assumed (in line with Sanders et al., 1993, see section 1.1) that coherence relations are part of the meaning representation of the text. Coherence relations are therefore not part of the surface code (representation), but are either represented at the textbase, the situation model or both. With respect to the question of the representational status of causal relations, no direct evidence exists that exactly pin points the level at which they are represented. However, the literature on discourse processing does offer some clues as to how to answer this question. This issue is taken up more elaboratively in chapter 5, but the crucial issues will already be introduced here.

A distinction will be made between explicit and implicit causal relations. In the case of an explicit causal relation, it seems plausible to assume that the relation itself is represented at both the textbase and the situational level (since the relation is part of the meaning representation), whereas the linguistic indicator is represented at the level of the surface code representation. Note that for a full comprehension of a text it is necessary that the reader constructs both a textbase and a situational representation. Of course, this will only happen if the explicit text and the reader’s background knowledge permit the reader to do so. The textbase is usually seen as a level of representation in which explicit propositions in the text are represented in a connected form. Propositions can be further decomposed into a predicate and one or more arguments. Connectives are one of the elements that may function as a predicate (Graesser et al., 1997; Kintsch, 1998). Thus, when the causal relation is made explicit in the text it may be represented at the level of the textbase. For example, consider the following simple sentence:
Chapter 1

(7) John fell, because Bill had pushed him.

In a Kintsch-style propositional analysis (cf. Kintsch, 1998), this sentence may be represented at the textbase level as \( \text{CAUSE}(\text{PUSH, BILL, JOHN}), (\text{FALL, JOHN}) \), where the predicate \( \text{CAUSE} \) conveyed by the connective takes the propositional representations of the clauses it connects as its arguments.

At the same time however, situation model theorists see causality as one of the major dimensions of situation models, and readers are assumed to monitor the causal continuity in the model (Graesser et al., 1997; Zwaan & Radvansky, 1998). Clearly, the situation model is constructed on the basis of the explicit text and background knowledge, so it seems counterintuitive to assume that if the text contains an explicit indicator of the causal link, the causal relation between the events is not represented at the level of the situation model. In other words, if the text contains a causal connective, it is highly plausible that the causal relation is represented at both the textbase level and the situation model.

This idea is highly compatible with Noordman & Vonk’s (1998) account of the function of the causal connective \( \text{because} \) during reading. This approach was already discussed briefly in section 1.2., but here we take it one step further because the levels of representations may be connected to the functions of \( \text{because} \). As was explained above, Noordman and Vonk hypothesize that \( \text{because} \) has three different functions that are related to the three levels of representation described above. The first function is the segmentation function. \( \text{Because} \) as a segmentation device signals to the reader how the incoming sentence is structured. For instance, \( \text{because} \) signals that the clause it introduces is a subordinate clause. This function of \( \text{because} \) is related to the surface code representation. The second function of \( \text{because} \) is the integration function. In this function the connective signals that the clauses have to be integrated by means of a causal coherence relation. The integration function of \( \text{because} \) may be related to the textbase as it specifies the text internal link between the clauses. The third function of \( \text{because} \) is the inference function. \( \text{Because} \) signals to the reader to activate background knowledge that underlies the causal relation explicated by \( \text{because} \). Here, the explicit text is related to background knowledge. As we have already seen, during the construction of the situation model, explicit text is integrated with background knowledge. For this reason, the inference function of \( \text{because} \) may be related to the situation model. The processing results obtained by Cozijn (2000) and Millis & Just (1994) in their investigations of the processing influence of \( \text{because} \) can be taken as support for Noordman and Vonk’s hypothesis. These results will be reviewed in section 1.3.3. For the present purposes, the main point is that causal coherence relations may be represented at both the textbase and the situational levels of representation, whereas the representation of the connective itself is represented at the level of the surface code.

But what about implicit causal relations? If there is no connective, is it still plausible to assume that a textbase and a situational representation of the relation are being constructed? There are reasons to believe this is indeed the case, though other considerations may lead to the conclusion that implicit causal relations are more properly described as being represented at the level of the situation model only. Let us start with the first possibility. Besides explicit text propositions, the textbase may also include inferences that establish local coherence...
Local coherence can be seen as a characteristic of the textbase (Singer, 1990). As will be explained below, local coherence often arises from the construction of causal inferences that connect sentences (Singer, 1994; Van den Broek, 1994). Thus, it may be the case that even when the causal relation between the sentences remains implicit, this relation is represented cognitively at the level of the textbase. Consider the following example.

(8) John fell. Bill had pushed him.

If the reader infers the causal link between the sentences, the propositional textbase may look like CAUSE((PUSH, BILL, JOHN), (FALL, JOHN)). This is exactly the same propositional representation that underlies sentence (7). The difference is that the causal link is inferred by the reader on the basis of his background knowledge and not explicitly indicated by because. This is an example of a text connecting inference (Graesser, Bertus, & Magliano, 1995; see section 1.3). In short, causal links between sentences may be represented cognitively at the level of the textbase even in the absence of an overt linguistic indicator.

On the other hand, it can also be argued that in the case of implicit causal relations, the relation is only represented at the situational level. If the relation is not overtly marked, the connected representation of the two sentences is a product of explicit text and the reader’s background knowledge. As was explained above, this is one of the characteristics of the situation model. From this point of view, the textbase contains the representation of the explicit statements in the text, but the causal coherence relation that connects the statements is located at the level of the situation model. The causal link between the sentences is not represented in a propositional format if the text does not contain an overt marker that expresses the causal link. The fact that the text is perceived as causally coherent can then be attributed to the situation model.

Fletcher (1994) gives additional reasons as to why causal relations may be best considered to be represented only at the level of the situation model. The first reason is that causal relations often link clusters of propositions. However, this reason does not seem to be a particularly strong argument. In principle, the textbase also includes relations between propositions that take other propositions as arguments, thus one proposition might as well be a whole cluster of propositions. Indeed, the propositional representation of (7) is an example of a proposition that takes as its arguments the propositional representations of the individual clauses. Also, the reader may construct macropropositions on the basis of the explicit (micro) propositions in the text (Kintsch & Van Dijk, 1978; Van Dijk & Kintsch, 1983; Singer, 1990). If these macropropositions are interrelated, the connections between them can also connect several clusters of (micro)propositions.

Fletcher’s (1994) second argument is that the perception of causal relations is not restricted to the interpretation of text, but also occurs in the physical and social surroundings. This is, in fact, compatible with a view on coherence relations presented by Garnham (1983, 1991; Garnham & Oakhill, 1996). According to this view, coherence in text does not arise from the text itself, but from the fact that the events described in the text are causally related in the
In sum, theoretical observations seem compatible with the view that coherence relations are represented at both levels of representation. If the text explicitly indicates the relation is causal, it seems straightforward to assume that causal coherence relations are represented at the level of the textbase and the situation model and that its linguistic indicator is presented at the level of the surface code. If the relation is implicit, it is still possible that the relation is represented at the textbase level, but it may be more likely in this case that implicit relations are represented at the situational level only. Note that we have not considered another possibility that causal relations are only represented at the level of the textbase. We do not believe this is plausible, because it is highly unlikely that readers will only construct a textbase representation (Graesser et al., 1995). However, in principle it is not impossible and in some scenarios it may even be likely. For instance, when the text contains overt signals of the underlying coherence relations and the reader does not have enough background knowledge to create a situation model of the text. This may be the case when readers read an expository text describing a topic unfamiliar to them (Britton, 1994).

As stated above, empirical evidence that distinguishes between the different possibilities of the representational level of causal relations is not available. However, studies that investigate the impact of marking of text structure on the representation of text (e.g. McNamara & Kintsch, 1996), as well as from studies that investigate the effects of marking on on-line processing of causal relations (e.g. Cozijn, 2000; Millis & Just, 1994) do indicate that there is some indirect evidence suggesting that causal relations are represented at both levels.

Let us first review some of the off-line evidence (the on-line processing results are presented in section 1.3.3). McNamara and Kintsch (1996) obtained results that seem to indicate that the presence of explicit markers of text structure has differential effects on the textbase and the situation model, depending on the background knowledge of the reader. In their experiment, they manipulated the presence of cohesive devices to obtain an incohesive and a cohesive version of a text. In the cohesive version, several linguistic indicators of text structure were present in the text, whereas in the incohesive version these indicators were not present. These experimental text versions were present to readers, who had high or low knowledge of the text topic.

The effect of the presence of the cohesive devices was measured by using tasks that tap either the textbase or the situational level of representation. For the textbase measures all of the information that is needed to perform the tasks is directly provided by the text. By contrast, for the situational measures the information is not directly provided but requires the textual information to be integrated with background knowledge. For example, this happens when the task requires the reader to access relations that are not explicitly indicated in the text (in the low cohesive text). One of the situation model tasks that was used was a key word sorting task. This task measures how concepts in the reader’s knowledge base are related to each other. The sorting task can be used as a situational measure by comparing the sorting pattern before reading to the pattern after reading. If the scoring pattern has (substantially) changed, this indicates that the relations between concepts in the reader’s knowledge have changed on
The Processing and Representation of Causal Coherence Relations

The basis of the information in the text. Such a change can only arise if the textual information is integrated in the reader’s background knowledge. Therefore, a change in knowledge only arises when a well developed situation model is constructed by the reader. Thus, the amount of change between pre- and post-reading sorting offers an indirect assessment of the quality of the situation model.

McNamara and Kintsch (1996) hypothesize that the effect of cohesive devices is different for textbase and situation model representations, depending on the knowledge of the reader. The more cohesive texts are supposed to increase the quality of the textbase, and the quality of the situation model, but only for low knowledge readers. For high knowledge readers, it is expected that the more cohesive text improves the textbase, but not the situation model. On the contrary, for high knowledge readers the presence of cohesive devices in the text should have a negative impact on the situation model.

The theoretical background for these hypotheses is provided by the construction-integration model (Kintsch, 1988; 1998). In this model, comprehension of a text involves a construction and an integration phase. During the construction phase, concepts in the text and knowledge of the reader are activated. During the integration phase, the concepts that are relevant for the comprehension of the text mutually increase their activation, whereas irrelevant concepts decrease in activation. The result of the integration phase is a stable network of activated concepts, both derived from the explicit text and background knowledge.

The hypothesis on the interaction of cohesion and background knowledge – the high knowledge readers who read a cohesive text develop a poor situation model – follows from the idea that text comprehension involves the integration of text and knowledge. A cohesive text will restrain the reader from actively using background knowledge. If a text is not cohesive, the reader will have to use background knowledge much more intensively in order to construct a coherent representation. This will lead to stronger connections between the explicit text and background knowledge. As the situation model reflects how well the text and knowledge are integrated, it is expected that less cohesive texts will lead to a better developed situation model.

McNamara and Kintsch’s (1996) results confirmed their hypotheses. For textbase measures the more cohesive text leads to increased performance for both low and high knowledge readers. For the situation model, inference questions and problem solving questions show how the cohesive text leads to better performances for low knowledge readers, but for high knowledge readers, the less cohesive text lead to a better performance. The results on the key word sorting task show that for low knowledge readers the amount of change between the pre- and post-sorting task was larger in the cohesive condition, whereas for high knowledge readers the amount of change was larger for the incohesive text. These results replicate the findings reported by McNamara, Kintsch, Songer, and Kintsch (1996).

What do these findings tell us about the level of representation of coherence relations? First, the cohesive devices that were used in the experiment are comparable to the linguistic markers of coherence relations that were discussed before. It was proposed that these are represented at the level of the surface code, but that they may also influence the processes involved in the interpretation of causal coherence relations. McNamara and Kintsch’s (1996) results (indirectly) indicate that the presence of linguistic indicators indeed influence the
processes involved in the construction of the textbase and the situation model. Second, the results are consistent with the idea that coherence relations are represented at both the textbase and the situational level, because cohesive devices influence performance on textbase and situation model. Indeed, if readers had not constructed these levels of representation, the differential influence of cohesion on textbase and situation model would not have been detected.

However, there are several problems with the methodology that was used in the experiments. The first problem is that several cohesive devices were manipulated simultaneously (Kamalski, 2007; Sanders & Noordman, 2000). This means that although the results are in line with the idea that marking of coherence relations has effects on textbase and situation model, or even different effects for textbase and situation model (for high knowledge readers), it is unclear what kind of influence a particular cohesive device has on these levels of representation. The second problem with the McNamara and Kintsch (1996) study is that they manipulated only one experimental text. As a result, both the internal and external validity of the experiment are under dispute (Meuffels & Van den Bergh, 2006). With respect to the internal validity, the problem is that idiosyncratic properties of the text are confounded with the experimental manipulation. Any effect that was obtained may therefore be due to how this particular individual text behaves under different conditions, instead of an effect of the manipulation that is not restricted to the actual text that was used. This also leads to problems with the external validity of the study. There is no way of knowing that the effects found for this particular text generalize to other texts. As Tukey (1969) puts it:

“No internal evidence can tell us about how much other repetitions – of those things we actually repeated only once – might differ. External evidence there may be, but rarely will it have been assembled for careful scrutiny by others.” (p. 84)

Consider the following analogy. Someone claims that people (in general) perform better in stressful situations. When asked for an argument to support this claim, the person says: “Well, I always do more work when I am under a lot of stress”. It is unlikely that anyone would seriously accept the general claim on the basis of this particular argument. Now consider this conclusion drawn by McNamara and Kintsch (1996):

“We conclude, with Britton and Gulgoz and other experimenters, that revising instructional texts to be more coherent and explicit can indeed foster better text memory and learning. But, we confirm McNamara et al.’s claim that such revisions are beneficial only for low knowledge readers and may be counterproductive for students how have the necessary knowledge background to understand low-coherence texts on their own.” (p. 282)

We can hardly maintain that the results on the one text that was used in McNamara and Kintsch’s experiment license drawing such general conclusions. In other words, the generalizability of the results is questionable. Still, the conclusion that cohesive devices might have differential effects for textbase and situation model measures is highly interesting.
The methodological problems identified here were partly solved in other more recent studies. For instance, only one particular kind of cohesive device was used, for instance connectives (e.g. Cozijn, 2000; Degand, Lefèvre, & Bestgen, 1999; Degand & Sanders, 2002; Millis, Graesser, & Haberlandt, 1993; Millis, & Just, 1994) or other linguistic indicators for coherence relations (Sanders & Noordman, 2000). With few exceptions (e.g. Kamalski, 2007) these studies did not make a distinction between textbase and situation model, so there is no way of telling whether the effects obtained (or the absence of effects, for that matter) can be contributed to the textbase or the situation model.

1.3 Cognitive processes of discourse interpretation

1.3.1 Discourse inference processes

When a text is fully comprehended by a reader, the text representation will be a coherent structure in which explicit textual information is interconnected and also integrated with background knowledge (Kintsch, 1988; 1998). The mutual connections between text elements and the connections between the text and background knowledge are the result of knowledge based inference processes (see among many others: Graesser, Singer, & Trabasso, 1994; McKoon & Ratcliff, 1992; Singer, 1990, 1994; Van den Broek, 1994). The focus of this section is on the knowledge based inferences involved in the interpretation of causal coherence relations.

Knowledge based inferences are pieces of information that are derived from background knowledge and that are encoded in the cognitive representation of the text (McKoon & Ratcliff, 1992; Graesser et al., 1994; Singer, 1990; 1994; Van den Broek, 1994). In general, researchers agree that information that is not explicitly given in the text but is actually encoded in the representation is considered to be an inference, although some researchers also consider information that is activated but not necessarily encoded as an inference (Van den Broek, 1994). Note that the term inference is also used for the process of making the inference (Singer, 1990). Thus, ‘knowledge based inference’ refers to the process of making the inference as well as the product of the inference: the actual information derived from background knowledge.

Some inferences contribute to the coherence of the representation, whereas others add to the completeness of the representation (Cozijn, 2000; Noordman & Vonk, 1992; Singer, 1988), which means that these latter inferences elaborate or enrich the representation that is based solely on the explicit information in the text. In general, inferences that are necessary for the coherence of the text representation are assumed to be made consistently during reading, whereas elaborations of the explicit text are not made consistently (Singer, 1994; Van den Broek, 1994).

For the present purposes, an important distinction is one between two types of knowledge based inferences: text connecting inferences and extra textual inferences (Graesser et al., 1995). Text connecting inferences result in a link between the current clause and a previous explicit statement in the text. Inferring the type of connection between explicit statements in
the text is an example of a text connecting inference. A text connecting inference may result in establishing a causal link between explicit statements in the text. Thus, establishing the (causal) coherence relation between sentences may involve making a text connecting inference. In the case of extra textual inferences, information is encoded in the meaning representation that is not explicitly mentioned in the text (Graesser et al., 1995). This information is derived from the reader’s background knowledge. This distinction between text connecting inferences and extra textual inferences is somewhat vague, as both types of inference result in information derived from background knowledge being added to the representation of the literal text. The following example may serve to clarify the difference between text connecting inferences and extra textual inferences.

(9) The child pricked the balloon with the pin. The balloon burst. (Example adapted from Singer, 1990).

When reading these sentences, a possible interpretation is that the balloon burst because the child pricked it with the pin. In this case, a text connecting inference results in the link between the sentences. A text connecting inference amounts to recognizing that the events described in the text are causally related, and adding the causal link between the sentences to the meaning representation. Thus, the causal coherence relation between the sentences is established by means of a text connecting inference.

In the following example, an extra textual inference is needed for a causal connection between the events.

(10) A burning cigarette was carelessly discarded. The fire destroyed many acres of virgin forest. (Example from Keenan & Kintsch, 1974).

A tentative interpretation of this sequence of events is that the events are causally related. Note, however, that there is no direct causal link between the events described sentences. For this reason, the causal connection between the events cannot be established by a text connecting inference, but the causal interpretation of the events requires the reader to infer the additional event that the burning cigarette caused the fire. This additional information is not explicitly mentioned in the text and is constructed by the reader on the basis of his background knowledge. The combination of this additional information and the explicit text, explains why the event in the second sentence takes place. In other words, both the inferred information and the explicit text are antecedents of the consequence described in the second sentence. The inferred extra textual information can therefore be regarded as a causal antecedent inference (Graesser et al., 1995; Singer, 1994; Van den Broek, 1994). The difference between text connecting inferences and extra textual inferences can also be illustrated by means of making the relations explicit. The relation underlying (9) can be made explicit with lexical signaling:

(9') The child pricked the balloon with the pin. *As a result*, the balloon burst.
An explicit version of (10), on the other hand, requires more information than a “simple” link
between the sentences: a whole sentence or clause, at least containing a full proposition,
which constitutes an event or state-of-affairs on its own.

(10’) A burning cigarette was carelessly discarded. The burning cigarette caused a fire. The
fire destroyed many acres of virgin forest.

The italicized parts of (9’) and (10’) are the linguistically realized counterparts of respectively
the text connecting inference and the extra textual inference. The TCI can be paraphrased by a
“relational proposition” that connects two explicitly realized text segments and which can be
paraphrased by one lexical chunk or connective expressing the causal coherence relation
between the segments. The ETI can be any part of information not explicitly mentioned in the
text, which, in combination with the explicit text explains the type of connection between the
text segments.

Both inferences that may be made during reading of (9) and (10) are examples of inferences
that establish coherence. It is by virtue of these inferences that the explicit statements in the
text are connected in the text representation. Whereas text connecting inferences are always
involved in establishing coherence, extra textual inference may or may not be necessary for
coherence (Graesser et al., 1995). For example, in (9) a possible extra textual inference is that
after the balloon burst the child got scared from the bang that accompanied the bursting of the
balloon. This inference is also causally related to the events described in the text, since it
consists of a consequence of the second event, but the inference is not needed for establishing
coherence between the two sentences. Thus, there is a difference between inferences that
establish coherence (text connecting or extra textual) and extra textual inferences that do not.

This brings us to another important distinction frequently made in the literature: the difference
between bridging inferences and elaborative inferences (Singer, 1990; 1994; Van den Broek,
1994). Bridging inferences bridge the conceptual gap between explicit text statements. As an
example, consider the following classic example:

(11) We checked the picnic supplies. The beer was warm. (Example from Haviland & Clark
(1974)).

Coherence between the sentences arises only after the reader infers that the beer must have
been part of the picnic supplies. Stated differently, if the reader fails to infer that the beer was
part of the picnic supplies a conceptual gap (incoherence) exists between the sentences. This
implies that the possible causal inferences underlying (9) and (10) can be considered to be
bridging inferences. Without these inferences the statements are not conceptually connected
by means of coherence relations. Thus, both text connecting inferences and extra textual
inference that establish causal coherence between explicit statements in the text can be
considered as bridging inferences. This leads to the conclusion that coherence relations are the
result of bridging inferences.

Other inferences are not necessary for coherence. These inferences are generally called
elaborative inferences (Singer, 1990; 1994; Van den Broek, 1994), which are elaborations of
the explicit text content. For instance, the causal consequence inference that the child got scared does not result in coherence between the sentences describing the child pricking the balloon with the pin and the balloon bursting. Although elaborative inferences are not involved in establishing coherence, they may be important for improving the completeness of the representation (Cozijn, 2000; Noordman & Vonk, 1990; Singer, 1988).

In sum, bridging inferences are involved in the construction of coherent representations. The bridging inference may consist of inferring a causal link by means of a text connecting inference or by means of an extra textual inference that involves deriving additional information from background knowledge. Extra textual inferences may also be elaborative inferences that are not involved in establishing coherence, but that contribute to the completeness of the representation.

Although many theorists in the field of discourse processing agree that causal bridging inferences are routinely made during reading because of coherence requirements, there is also disagreement between theories as to whether this includes both local and global coherence requirements. In the next subsection, two opposing leading theories are compared with respect to the causal inferences they predict to be routinely made during reading.

1.3.2 The role of causal relations in discourse inference theories

Theories of inferential processing show remarkable differences with respect to the types of causal inferences predicted to be routinely made during reading. This difference is especially apparent in the discussion between researchers inspired by a constructionist theory as proposed by Graesser, Singer, and Trabasso (1994) and researchers inspired by the Minimalist Hypothesis (McKoon & Ratcliff, 1992). Each theory has a different domain of interest, and as a result their perspective on the reading process differs (Van den Broek & Lorch, 1993).

The constructionist theory proposed by Graesser et al. (1994) is presented as a theory that offers precise predictions about the kinds of inferences readers will make during text comprehension. Like in other constructionist approaches, it is assumed that comprehenders attempt to construct meaning out of text, social interactions, and perceptual input. With respect to the reading process, this effort-after-meaning principle is elaborated by three specific assumptions.

First, the reader goal assumption asserts that the meaning representation that is constructed by the reader satisfies the goals of the reader. This means that the representation that is constructed when the goal is to learn from a text differs from when the goal of the reader is to proof-read a text for spelling errors. Second, the coherence assumption states that readers attempt to construct a meaning representation that is both locally and globally coherent. Local coherence is coherence that is established between adjacent clauses or short sequences of clauses that co-occur in the reader’s Working Memory (WM). Establishing local coherence

---

1 Although the theory of Graesser, Singer, and Trabasso (1994) is ‘only’ one instance of a theory that embraces constructionist principles, below the theory will sometimes be referred to as The Constructionist Theory. Several other theories, for instance those inspired by Mental Models Theory, can also be seen as constructionist theories.
The Processing and Representation of Causal Coherence Relations includes linking the information in WM by means of detecting referential connections and coherence relations. Global coherence, on the other hand, refers to coherence that is established by the organization of local information in to higher order information or to coherence that is established between the current clause and an explicit statement earlier in the text, which is no longer active in WM.

The first kind of global coherence can be demonstrated as follows: the reader infers “John likes sports” when the text contains the local propositions “John likes tennis”, “John likes basketball”, “John likes football”, and “John likes volleyball”. The second kind of global coherence occurs when an event is mentioned early in the text, whereas its consequence is described much later. Connecting the consequence to this antecedent requires that the latter statement be reinstated in WM.

Third, the explanation assumption asserts that readers attempt to explain why actions, events and states are being described in the text. This explanation driven processing is also a fundamental aspect of the causal theories of narrative comprehension that will be discussed later in this section. Thus, the reading process from the perspective of Graesser et al.’s (1994) Constructionist Theory can be seen as an essentially reading-goal driven process in which the reader attempts to construct a coherent representation by actively trying to explain the occurrence of events.

A completely different stance is taken in the Minimalist Hypothesis (McKoon & Ratcliff, 1992). The minimalist hypothesis focuses on (minimal) automatic inference processes that take place during reading. These processes occur independent of strategic processing or specific reading goals the reader might have. Minimal inferences provide the input for strategic inferences that are constructed during reading, and they result in a minimalist representation of the text in memory upon which more elaborate representations can be constructed. Automatic inferences only occur if the information upon which they are based is quickly and easily available. This information may come from the explicit text and the reader’s general knowledge. An automatic inference that is derived from general knowledge is only made if the information in the reader’s knowledge base is well-known. If this is the case, it is assumed that the information is quickly and easily available from background knowledge.

Automatic inferences that are based on the explicit text are used to establish local coherence. It is hypothesized that inferences that establish global coherence are only made when local information does not provide coherence.

The major difference between the Constructionist Theory and the Minimalist Hypothesis is that from the perspective of the Constructionist Theory, reading is considered as fundamentally goal directed and explanation based, whereas the Minimalist Hypothesis refers to cases that are independent of the specific reading goals and strategies of the reader. A similar way of looking at this difference is in terms of depth of understanding: the Constructionist Theory is a theory of inferential processes that are involved in deeper understanding of text, whereas the Minimalist Hypothesis pertains to low level text

---

2 This is an example of an application of the generalization rule in the formation of macropropositions (Van Dijk, 1980), which states that a more general proposition entailed by a sequence of propositions will be constructed by the reader. The macrostructure of the text is the representation of the global structure of the text, in which the macropropositions are interconnected.
understanding. In such a view, the two theories define different domains of interest and might even be compatible (Van den Broek & Lorch, 1993).

Let us now turn to the question which role causal coherence relations play in these theories of discourse inference processes. In order to make the discussion less abstract, some examples of causal inference are presented that might occur during reading. Both the following text and the example inferences are taken from Graesser et al. (1996, p. 13).

(12) The Czar and His Daughters
Once there was a Czar who had three lovely daughters. One day the three daughters went walking in the woods. They were enjoying themselves so much that they forgot the time and stayed too long. A dragon kidnapped the three daughters. As they were being dragged off they cried for help. Three heroes heard their cries and set off to rescue the daughters. The heroes came and fought the dragon and rescued the maidens. Then the heroes returned the daughters to their palace. When the Czar heard of the rescue, he rewarded the heroes.

Possible causal inferences that may be made during reading of this text are the following.

Superordinate Goal: A goal that motivates an agent’s intentional action. E.g. “The dragon wanted to eat the daughters.”
Subordinate Goal/Action: A goal plan or action that specifies how an agent’s intentional action is achieved. E.g. “The dragon grabbed the daughters.”
Causal Antecedent: An inference on a causal chain, or bridge, between the explicit action, event or state being comprehended and the prior passage. E.g. “The dragon saw the daughters.”
Causal Consequence: An inference on a forecasted causal chain emanating from the explicit clause being processed. These include physical events and new plans of agents, but not emotions. E.g. “Someone rescued the daughters.” (In the text this inference would have been made prior to the text describing that heroes came and rescued the daughters; the reader could have made this inference upon reading that the daughters were kidnapped)
Character Emotion: An emotion experienced by a character in response to an action, event, or state. E.g. “The daughters were frightened.”

Note that all of these inferences are extra textual inferences (Graesser et al., 1995). The inferences refer to information that is not explicitly mentioned in the text and is derived from the reader’s background knowledge. The question is, then, what the role of text connecting inferences is in these theories. Neither Graesser et al. (1994) nor McKoon and Ratcliff (1992) give an explicit answer to this question. However, the answer seems to be straightforward: since text connecting inferences establish local coherence, it is plausible to assume that both theories predict that these inferences will normally be generated during reading.

According to the Minimalist Hypothesis causal inferences that are encoded automatically during reading should be those that are quickly available from the reader’s knowledge or those that are required for local coherence (McKoon & Ratcliff, 1992). As the causal antecedent inference is the only inference necessary for establishing local coherence, this
The Processing and Representation of Causal Coherence Relations

According to the Constructionist Theory, inferences that establish local and global coherence should be generated on-line. Superordinate goal inferences and character emotion inferences are two types of inference that are important for establishing global coherence (Graesser, Bertus et al., 1995; Graesser et al., 1996). Therefore, these inferences should be generated on-line. Furthermore, causal antecedent inferences should be generated on-line, because they are important for establishing local coherence. Also, from the explanation assumption it follows that causal antecedent inferences and superordinate goal inferences should be generated on-line, since these inferences explain why events occur or why an agent performs certain actions. The Constructionist Theory predicts that subordinate goals and causal consequence inferences are not consistently made during reading. The reason for this is that these inferences are elaborative inferences that are neither necessary for coherence, nor provide explanation for why events occur (Graesser et al., 1995; Graesser et al., 1996). Another reason why causal consequence inferences are normally not generated on-line is that there are too many hypothetical plots that could be forecasted and most of them will not be substantiated later in the text (Graesser, 1981). Because there are so many hypothetical consequences, the amount of cognitive resources necessary for constructing causal consequence inferences is likely to be very large. Therefore, given the restricted amount of processing resources available during reading, it is implausible to assume that readers construct causal consequence theories. However, there are circumstances under which causal consequence inference may in fact be produced. For instance, if the consequence is highly constrained by the context and few alternative consequences are likely to unfold (Graesser et al., 1995), or when the explicit text provides sufficient explanation for the consequence to occur (Van den Broek, 1990; 1994; Van den Broek, Risden, Fletcher, & Thurlow, 1996).

Although in general circumstances causal consequence will not be created on-line, one model proposed by Graesser et al. (1994) does predict that causal consequence inferences are generated consistently during reading. This model is called the prediction-substantiation model. In this model, reading is not only considered to be explanation driven, but also expectation driven. During reading, readers construct expectations about future events in the plot of the narrative, and these expectations guide the interpretation of the incoming text in a top down way. For instance, a text might trigger the activation of a schema or script underlying the narrative (Bower, 1976; Mandler & Johnson, 1977; Shank & Abelson, 1977; Thorndyke, 1977). A schema is a knowledge structure that represents knowledge of the form or structure of a typical story. Also, the existence of schemas for cause, plans and goals have been proposed (Shank, & Abelson, 1977) and demonstrated (for an overview see Singer, 1990). If a text triggers a restaurant script and subsequently describes two people entering the restaurant, the reader creates the expectations that the characters will eat, talk and be served food (Graesser et al. 1994). Later mention of these events, or, in other words, substantiation of the expectations in the text should facilitate processing of these events. The expectations created by the reader are more likely to be abstract than fully detailed. Thus, the reader is likely to expect that the protagonists will order something in the restaurant, but will probably not infer they will order a specific meal. The possibility that top-down expectations are
generated by the reader on the basis of schematic knowledge will become important later in this chapter (see section 1.3.4 and also chapter 5).

To recapitulate, the two opposite theories of inference processing differ in their perspective on the reading process. The Constructionist theory assumes that readers normally aim at constructing representations both locally and globally coherent and that readers explain the occurrence of events. The minimalist hypothesis focuses on lower level meaning representations. These minimal representations are furnished by automatic inferences that establish local coherence between text segments and that are based on knowledge that is easily available. In terms of causal inferences, the theories differ with respect to two types of inference: global inferences and character emotions. Only the Constructionist Theory predicts that these inferences will be made during reading. According to the minimalist hypothesis, character emotion inferences will not be made automatically, and global inferences will only occur, when the reader is unable to establish local coherence.

Although the Minimalist Hypothesis has received much criticism, not surprisingly especially from research adopting constructionist views on text comprehension (e.g. Garnham, 1992, 1993; Glenberg, & Sashi, 1992; Singer, 1993; Singer, Graesser, & Trabasso, 1994), the important observation for the present purposes is that both the minimalist hypothesis and the constructionist theories support the idea that causal coherence relations are being inferred during reading. Indeed, in both theories, coherence requirements are one of the determining factors for the generation of knowledge based inferences. For a cognitive account of coherence relations, however, the constructionist theory is the most attractive theory. Especially the minimalist assumption that inferences that establish global coherence are normally not created is unattractive for coherence relation theories. For example, suppose a text consists of two paragraphs of which the first paragraph describes a problem and the second paragraph describes a solution to this problem. In a coherence account of text structure these paragraphs would obviously be connected by means of a Problem-Solution relation. This analysis of the text structure is a description of the global structure of the text. According to the Minimalist Hypothesis this global coherence structure will normally not be inferred by the reader. Still, it is counterintuitive to assume that readers will have understood the text without noticing how the different paragraphs of the text are related to each other.

The picture that emerges from the discussion so far can be described as follows. Coherence relations are the product of knowledge based bridging inferences. Coherence may arise from two kinds of knowledge based inference: text connecting inferences and extra textual inferences. Text connecting inferences establish local coherence between adjacent sentences in the text by linking the sentences by means of a causal coherence relation. Extra textual inferences may result in both local and global coherence. For example, causal antecedent inferences result in local coherence, and superordinate goal inferences result in global coherence. Elaborative extra textual inferences, such as causal consequence inferences, may also be generated by the reader on the basis of knowledge about the typical structure of text.
1.3.3 Empirical evidence for causal inferences

As was explained in the introduction, a cognitive account of coherence relations predicts that differences between types of coherence relations are reflected in processing and memory measures. Also, such an account predicts that making coherence relations explicit should influence processing. In the previous sections it was proposed that coherence relations are the product of text connecting and/or extra textual inference. The main goal of this section is to review the empirical evidence that is available with respect to causal inferences. This review will lead to the formulation of hypotheses regarding the inferential processes involved in the interpretation of (local) causal coherence relations.

Several studies indicate that causal inferences are indeed generated during reading (see for overviews: Graesser et al., 1994; Graesser et al., 1995; Graesser et al., 1996; Singer, 1990; 1994; Van den Broek, 1994). Processing results obtained by Keenan, Baillet, and Brown (1984) and Myers, Shinjo and Duffy (1987) are generally interpreted as showing that causal inferences are made during reading. In these studies, the strength of the causal relation is manipulated. Consider for instance, the following experimental task used by Keenan et al..

(13) Joey’s brother punched him again and again. The next day his body was covered with bruises.
(14) Racing down the hill, Joey fell of his bike. The next day his body was covered with bruises.
(15) Joey’s crazy mother became furiously angry with him. The next day his body was covered with bruises.
(16) Joey went to a neighbor’s house to play. The next day his body was covered with bruises.

The difference between the relations underlying the sentences (13) through (16) is the amount of causal strength of the relation. In (13) the relation is very strong; in (16) the relation is very weak. In fact, the relation in (16) is so weak it can even be argued that the relation in (16) is not causal at all (see also Chapter 5). The reading times of the second clauses, which are identical in all the examples, has been found to vary as a function of causal strength. In (13) the second clause is read fastest, whereas in (16) the sentence times are highest.

This pattern in reading times can and has been interpreted as showing that readers were inferring causal connections between the sentences, but the weaker the relation gets, the more cognitive resources are needed to establish a causal connection. For a strong causal relation, inferring the causal link is easiest. In terms of the previous sections, the causal link can be established by means of a text connecting inference in (13), and maybe even (13), whereas the reader will have to make an extra textual inference in (15) and (16). In the case of (15) and (16) the reader will have to infer an additional causal antecedent (Van den Broek, 1994). For instance, in (15) the reader might infer the intervening event that Joey’s mother punched him. Thus, the pattern in reading times suggests that readers try to connect sentences by means of causal relations and this will happen faster as it gets easier to do this.
At this point, it is worthwhile to notice the necessity of an explanation assumption besides a coherence assumption for the prediction of causal bridging inferences. According to the coherence assumption, readers will perform a bridging inference because of coherence requirements (e.g., Graesser et al., 1994; McKoon & Ratcliff, 1992). However, such an assumption alone cannot explain why readers would infer a causal connection when a coherence relation of temporal succession would suffice to connect the sentences. In contrast, an additional explanation assumption does predict that readers will try to establish a causal connection even though they could have settled for a temporal relation. Thus, the results of Keenan et al. (1984) indicate that readers do not only try to establish a connection between the events described in the text, but that they will also try to connect statements by means of a causal relation.

The explanation assumption accounts for the fact that reading times increase as the strength of the causal relation decreases. In a very strong causal relation, the explanation of the event is described in the text itself. In (13), for example, the fact that Joey was covered with bruises is explained by the explicit statement that he got punched repeatedly. In (15), on the other hand, additional information is needed to sufficiently explain the fact that Joey was covered with bruises. Thus, one of the text characteristics predicting whether a text connecting or an extra textual bridging inference will be made by the reader is the amount of explanation provided by the explicit text (Van den Broek, 1990; 1994).

If the text itself provides sufficient explanation, the relation can be established by means of a text connecting inference. This must not be taken to imply, however, that in these circumstances the reader does not have to use his background knowledge. Indeed, a text connecting inference is a knowledge based inference too. But the point is that readers must somehow know that a text connecting inference is appropriate when the relation is sufficiently explained in the text. This implies that readers check whether or not the explanation provided in the text is actually sufficient for the consequence to occur. For example, if readers make a text connecting inference in order to connect the sentence in (12), it is likely to be the case that their knowledge about the relation between punching and bruising has enabled them to verify that being repeatedly punched is a sufficient explanation for getting bruised. A similar point is made by Singer et al. (1992; see also: Halldorson & Singer, 2002; Singer, 1994; Singer & Halldorson, 1996; Singer, Revlin, & Halldorson, 1990), who present a validation model of causal bridging inferences. According to this model, a tentative causal bridging inference is only accepted by the reader after he has validated this inference on the basis of his world knowledge. Evidence that readers indeed activate additional background knowledge has been obtained by comparing sentence pairs like (17) and (18).

(17) Dorothy poured the bucket of water on the fire. The fire went out.
(18) Dorothy placed the bucket of water by the fire. The fire went out.

A text linguistic way to test whether the text provides sufficient explanation is to check the paraphrase of the relation by making it explicit with a connective or a lexical signaling phrase. In (13) and (14), the lexical signalling phrase ‘as a result’ can be used, whereas in (15) and (16) this is highly unnatural. Compare “Joey’s brother punched him again and again. As a result, the next day his body was covered with bruises.” and “Joey’s mother was furiously angry with him. As a result, the next day his body was covered with bruises.”
The difference between (17) and (18) is that a causal interpretation is compatible with the events described in (17) whereas this is unlikely for those in (18), for which a temporal succession of events is a more likely interpretation. According to the validation model, the causal bridging inference that connects the sentence pairs in (17) is accepted by the reader after the causal connection is validated against background knowledge. The required knowledge is hypothesized to be the missing major premise of the syllogistic reasoning pattern underlying the causal relation. To be more precise, the causal relation between the events can be analyzed in terms of a syllogism in which the major premise is missing. The syllogistic pattern is called an enthymeme. For sentence pairs (17) the reasoning pattern can be (informally) analyzed as follows:

Major premise: Water extinguishes fire.
Minor premise: Dorothy poured water on the fire.
Conclusion: (therefore) The fire went out.

The explicit text describes the minor premise and the conclusion. The major premise is left implicit in the text, and in order to ‘close’ the syllogism, this major premise will have to be derived from the readers’ knowledge base. The validation model predicts that if the tentative bridging inference is causal, the reader will activate this validating knowledge (the missing major premise); if the bridging inference is not causal, this knowledge will not be activated. In their studies, Singer and colleagues used a question answering task in which the participants had to answer a yes/no question. This question addressed the validating knowledge. For instance, subjects had to answer the question “Does water extinguish fire?” Now, the prediction is that when the validating knowledge is indeed activated during reading, the activation of this knowledge after reading should be higher than when the knowledge is not activated during reading. Because of this, participants are expected to need less time to answer the question if the validating knowledge is activated during reading. This is exactly what the results showed.

However, these results should not be taken to imply that knowledge about the relation between water and fire was not activated in the temporal condition at all. It is highly likely that this knowledge actually is activated during processing of this temporal sequence of events. For instance, this would be predicted by Kintsch’s (1988, 1998) construction-integration model, in which it is assumed that, during the construction phase, concepts in the knowledge of the reader are activated regardless of whether these concepts are relevant for the text currently processed. During the integration phase, the concepts that are not pertinent to the processing of the text will lose activation. Accordingly, the knowledge that water extinguishes fire is likely to be activated in the temporal condition too. However, according to the validation model, this knowledge is only needed to validate the tentative causal link in the causal condition, since no such link is established in the temporal condition.

The important observation is that inferring the causal connection requires the reader to access knowledge about the relation between the events that are described in the text. Whether this knowledge arises from syllogistic reasoning or whether the knowledge is actually used for validation of the relation is a less relevant issue. In any case, the knowledge on the basis of
which the relation is constructed is not explicitly expressed in the text. A question one may now ask is whether this knowledge only serves as a catalyst in the interpretation of the causal connection between the events, or whether the activated knowledge is also encoded in the representation of the text. In the latter case, the process of activation of the underlying knowledge can be seen as an inferential process that leads to the encoding of extra textual information. According to the definition of inferences described above, a knowledge based inference is information derived from background knowledge that is encoded in the representation. As was explained above, knowledge that is activated during reading, but not actually encoded in the representation, is generally not considered as an inference (Graesser et al., 1994; Van den Broek, 1994).

The question whether the knowledge underlying the causal relation is actually encoded during reading, or whether the knowledge ‘only’ serves as a catalyst for inferring the causal link was investigated by Halldorson and Singer (2002). In their experiment, they used a delayed priming technique. The reasoning behind this technique was as follows. If the knowledge underlying the relation becomes part of the text representation, explicit statements from the text and this underlying knowledge should mutually prime one another. That is, if connections between the explicit text and the underlying knowledge exist in the representation, presenting the explicit text should activate this underlying knowledge and presenting the underlying knowledge should activate the representation of the explicit statements. Consider the following example of the experimental materials that were used.

(19) The hiker examined the injured deer. The deer died.
(20) The hiker shot the injured deer. The deer died.

In (19) there is no apparent causal relation between the sentences, whereas (20) is compatible with a causal interpretation of the events. According to the validation model, the underlying knowledge about the relation between shooting and killing should only be activated and represented when (20) is being read. In order to test this hypothesis, Halldorson and Singer (2002) presented target questions preceded by a prime question. If the target question assessed the underlying knowledge (e.g. ‘Do bullets kill animals?’) the explicit antecedent was assessed in the preceding prime question. For example, in the causal condition the question “Did the hiker shoot the deer?” was presented, whereas in the non-causal condition the question “Did the hiker examine the deer?” was presented. If the target question contained the explicit antecedent question, the underlying knowledge was assessed in the prime question. Thus, both underlying knowledge and the explicit antecedent sentence were presented in the target questions and the prime questions.

The experiment tested the prediction that the validating knowledge is incorporated in the representation, and hence the information derived from background knowledge and the explicit causal antecedent information in the text should mutually prime each other. As the validation model predicts that the validating knowledge is only activated and encoded in the causal condition, the mutual priming effect should only be observed in the causal condition, but not in the temporal condition. If the hypothesis is correct that the underlying knowledge is incorporated in the representation, the explicit antecedent target sentence should be primed by
the underlying knowledge question and a difference in answer times between the causal and non-causal condition can be expected. However, if the underlying knowledge only serves as a catalyst for comprehension, the delayed priming task should not show this answering time difference on the explicit antecedent questions between causal and non-causal relations. The results show that for both prime types (explicit antecedent and underlying knowledge) the responses on the target questions were faster in the causal condition than in the non-causal condition. This indicates that the underlying knowledge is indeed incorporated in the representation. The most important conclusion for the present purposes is that constructing the causal link between the sentences does not only involve a text connecting inference that creates the causal link between the sentences, but also involves an extra textual inference. This follows from the fact that the knowledge that is needed to construct the causal link, is also incorporated in the cognitive representation of the text.

Let us now turn to explicit causal relations. If a causal connection is explicitly indicated with a connective like because, the reader will not have to make a text connecting inference. It might therefore be hypothesized that the extra textual inference that was just described in the context of the validation model is not made either. There seems to be no reason to assume that additional background knowledge is activated and encoded when the causal relation is explicitly indicated in the text. Surprisingly, processing evidence indicates that this is nevertheless the case. For instance, Simons (1993) used an expert-novice paradigm to investigate whether the inference underlying the causal relation indicated by because is made during reading. Simons presented his participants with texts containing sentences like the following.

(21) American exports have been suffering a decline, because rising inflation has produced a harmful effect on the competitive position of the USA.

The knowledge underlying the causal relation is that a decline in the competitive position of a country leads to a decrease in exports. Prior to the experiment it was established that the experts had this knowledge, whereas the novices did not. This inferential information was either explicitly presented in the text or was left out. If the extra textual inference is indeed made during reading, it can be expected that providing this information explicitly, decreases processing time of the sentences; after all: readers do not have to make the inference on their own as the text already gives the information. The data show that the explicit information leads to a processing advantage for the experts only. This suggests that the inference was made by the experts, but not by the novices. Thus, if the relation is signaled with the connective because, the extra textual inference is made during reading. However, this is only true for readers that have the knowledge that enables them to do so. For readers that do not have the required knowledge, making the inference amounts to the derivation of new knowledge on the basis of the text. Apparently, the derivation of new knowledge does not occur spontaneously during reading (Cozijn, 2000; Simons, 1993; Noordman, Vonk, & Kempff, 1992).
Millis and Just (1994) and also Cozijn (2000) investigated the comparison between implicit causal relations without a connective and explicit relations marked with *because*. The review of the results in the previous paragraphs suggests that the difference between the implicit and explicit conditions in terms of inferential processes is that in the implicit condition a text connecting inference (i.e. establishing the causal link) and an extra textual inference are made (i.e. activating and encoding the knowledge underlying the causal relation), whereas in the explicit condition only the extra textual inference is made. This leads to the prediction that the processing times in the implicit condition are longer than in the implicit condition. The results obtained by Cozijn (2000) and Millis and Just (1994), indicate, however, that this is only partly the case. Consider an example from the materials used by Cozijn.

(22) On his way to work that morning he experienced a big delay. There was a traffic jam on the highway.

(23) On his way to work that morning he experienced a big delay, because there was a traffic jam on the highway.

These sentence pairs were embedded in short narrative texts. Cozijn (2000) used both eye-tracking and self-paced-reading methods to investigate the influence of the connective on the processing of the second clause. The results indicate that the connective speeds up processing of the first and middle part of the sentence (i.e. there was a traffic jam) but actually slows down the processing of the last part of the sentence (i.e. on the highway). These speeding up and slowing down effects were also observed by Millis and Just (1994).

Cozijn (2000; see also Noordman & Vonk, 1997) explains these results as follows. The speeding up effect is due to the faster integration of the sentences in the explicit condition. This is the *integration effect of because*. In terms of the terminology used in this section so far, integration of the sentences involves establishing the causal link between the sentences. In the terms of this chapter, then, the integration of the sentences in the implicit condition is the result of a text connecting inference. This connecting inference does not have to be made in the explicit condition, because the connective already provides information on the type of coherence relation that holds between the sentences. Therefore, the type of connection does not have to be inferred. The slowing down effect can be attributed to the making of the extra textual inference underlying the causal relation. This is the *inference effect of because*. The inference is the extra textual inference that was discussed in the paragraphs above. Since making this inference takes time, the reading times are longer in the explicit condition.

The conclusion that the extra textual inference is indeed made when the connective is present, was corroborated by a verification task that Cozijn (2000) used in his experiments. In this task, the content of the extra textual inference was presented to the participants (i.e. Traffic jams may cause delays). In the explicit condition the participants reacted faster to these statements than in the implicit condition. This result, combined with the reading times leads Cozijn (2000) and Noordman and Vonk (1997) to conclude that the extra textual inference is

---

4 Note that here the terms implicit and explicit are used for the presence of connectives. In the Simons (1993) study that was described in the previous paragraph implicit and explicit refer to the explicit presence of the information that is needed to make the extra textual inference underlying the causal relation.
only made when *because* is present. However, we have seen how Singer and colleagues have shown that this extra textual inference is also made in the case of implicit causal relations (e.g. Dorothy poured water on the fire. The fire went out).

How can this apparent contradiction in results be resolved? For an answer to this question we have to realize that both Cozijn (2000) and Singer and colleagues have used off-line measures after reading to corroborate the idea that extra textual inference was made during reading. That is, the reading time patterns were taken to indicate that an inference was made during reading, but off-line tasks were needed to validate this idea. The logic underlying both the tasks used by Singer et al. and Cozijn is that if the extra inferential information is presented to participants, reactions will be faster in conditions where this information is already activated during reading, in comparison to a condition where this information is not activated. However, since these tasks are off-line measures, performance speed also reflects retrieval phenomena. This means that the faster responses on the off-line task may just as well reflect differences in the quality of the text representation in stead of a difference in the activation of the information during reading. Indeed, representational advantages for causal relations as opposed to non-causal relations have been found by many studies (e.g. Black & Bern, 1981; Bradshaw & Anderson, 1982; Myers, O’Brien, Balota, & Toyofuku, 1984; Sanders & Noordman, 2000; Trabasso & Sperry, 1985; Trabasso & Van den Brock, 1985). Similarly, many studies have found memory advantages for explicit causal relations as opposed to implicit causal relations (e.g. Degand, Lefèvre, & Bestgen, 1999; Degand & Sanders, 2002; Millis, Graesser, & Haberlandt, 1993; Millis, & Just, 1994). Thus, the fact that in the Cozijn (2000) study the participants reacted quicker to the inferential information in the explicit condition than in the implicit condition may be due to the representational effect of the connective, in stead of to the inference only being made in the explicit condition. If this is the case, the results obtained by Cozijn no longer contradict the results obtained in the studies inspired by Singer’s validation model.

As a matter of fact, the reading time patterns found by Cozijn (2000) and Millis and Just (1994) can also be explained by the hypothesis that in the implicit condition the text connecting inference and the extra textual inference are made, whereas in the explicit condition only the extra textual inference is made. The explanation can be formulated as follows. In the explicit condition, the sentences are first connected by means of the causal link signaled by the connective, and then the extra textual inference is made (see Noordman and Vonk, 1997). Making the extra textual inference means that the knowledge underlying the causal relation is activated and encoded at that particular moment in time. Suppose that the same time course applies to the implicit condition: first the causal connection is established, and then the extra textual inference is made. Now, the encoding of the extra textual inference may proceed faster in the implicit condition than in the explicit condition, because the required knowledge was already activated during the construction of the text connecting inference which occurs earlier. Therefore, a slowing-down effect of the connective is observed at the end of the sentence (or better: a speeding-up effect of the absence of the connective was observed).

With respect to the speeding-up effect of the connective, the fact that the text connecting inference only has to be made in the implicit condition explains why the processing times are
longer in that condition at the time the causal link between the sentences is established. As we will see later, there is a qualitative difference in the integration process between implicit and explicit conditions: establishing the causal connection in the implicit condition requires a text connecting inference, whereas in the explicit condition the causal connection consists of encoding the relation signaled by the connective. Below, the term text connecting operation (TCO) will be used to denote the process of establishing the causal connection in explicit conditions.

This alternative interpretation of the processing results assumes that in the implicit condition both a text connecting and an extra textual inference are constructed and in the explicit condition the extra textual inference is constructed in addition to the TCO. The pattern of reading times that is predicted by this interpretation is exactly the pattern obtained by Cozijn (2000) and Millis and Just (1994). Note, however, that this also requires the additional assumption about the time course of the generation of these inferences: the text connecting inference is made before the extra textual inference is made. This is exactly the time course of the causal processes proposed by Noordman and Vonk (1997) and Cozijn (2000).

1.3.4 Top-down versus bottom-up processing

So far, the review of the empirical data has focused on the inferential processes involved in the processing of explicit and implicit causal relations. The results indicate that causal inferences are indeed made during reading. Implicitly, a bottom-up perspective on text processing was taken: All of the inferences described originate from the explicit text. For instance, because the current clause has to be connected to the previous one, the reader activates background knowledge pertinent to the contents of these clauses. For example, reading the clauses “Dorothy poured water on the fire. The fire went out.” triggers the activation of knowledge about the relation between water and fire going out. However, as was described in section 1.3, knowledge structures such as scripts and schemas might also guide the interpretation of text in a top-down manner. In that case, knowledge pertinent to the text structure is activated prior to the actual processing of the clause in question, and the clause is interpreted (both in terms of its content and in terms of the relation with the preceding text) in congruence with the activated knowledge structure (see also Noordman & Vonk, 1998).

To make things more concrete, consider an experiment reported by Sanders and Noordman (2000). In their experiment, short newspaper articles were presented to the participants. These newspaper articles each contained a target sentence that was either embedded in a (causal) Problem-Solution context or in an (additive) List context. In the Problem-Solution contexts, the target information provided the reader with the solution to the problem described in the previous context. In the List context, exactly the same information was part of an enumeration of events. In addition to this difference between causally and additively related sentences, a second experimental factor distinguished between implicit and explicit relations: The coherence relations were either left implicit in the text or marked with a signaling phrase. These phrases are in parentheses in the examples below.
(24) Another inhabitant of Veendam was killed in a traffic accident yesterday. The man crossed the street and was hit by a truck. For years the people of Veendam have now been campaigning to reduce the annoyance caused by the traffic in their town. This annoyance is caused especially by freight traffic that passes through the heart of the town. Because of the heavy traffic, crossing the street has become very dangerous. (A solution is in sight now). The construction of a subway in the center of Veendam will begin next year. When the subway is finished, pedestrians and cyclists will be able to cross the road safely.

(25) The traffic in the region of East Groningen will be having trouble the next year because of road construction. Especially in Veendam and its surroundings road construction will be going on. A new local road will be constructed between the towns of Stadskanaal and Veendam. The exit of the highway between Groningen and the German border, on the east of Veendam, will be re-asphalted in spring. (A third project is situated nearby). The construction of a subway in the center of Veendam will begin next year. This was decided at a meeting of the city council. When the subway is finished, pedestrians and cyclists will be able to cross the road safely.

The target sentence of interest has been italicized in the examples. In (24) this target sentence is related to the context by means of a Problem-Solution relation, whereas in (25) exactly the same information is presented as part of a List structure. The reading time results obtained by Sanders and Noordman (2000) show that the explicit versions are read faster than the implicit version, and that the Problem-Solution relation is read faster than the List relation. This latter effect of the nature of the coherence relation corresponds to other research that shows a processing advantage of causal versus non-causally related information (Black & Bern, 1981; Hallidorsen, & Singer, 2002; Haberlandt, & Bingham, 1978; Keenan et al., 1982; Myers et al., 1992; Singer et al., 1992).

The important issue here is that the results of both the explicit indicators and the relation type can be explained from a bottom-up and a top-down perspective. A bottom-up explanation of the signaling effect is as follows. Because the signaling phrase specifies the type of relation between the sentences, the reader does not have to infer this relation when reading the target sentence. (Note that this was also the explanation of the processing advantage of causal connectives as described in the previous section). A top-down explanation of the signaling effect may be that the signaling phrase triggers reader’s schematic knowledge about the structure of text. As a result, the incoming sentence is interpreted with reference to this knowledge structure and is readily integrated in the developing mental representation. This integration process is easier in the explicit than in the implicit condition, because the signaling phrase triggers the schematic expectations.

The effect of relation type can also be explained from a bottom-up and a top-down perspective (Sanders & Noordman, 2000). From a bottom-up perspective readers are expected to try to connect sentences by means of causal relations, as was explained in the previous sections. Processing times indicate the amount of cognitive resources required to establish the causal relation. In the Problem-Solution case, the relation is causal, so trying to connect the sentences by means of a causal relation will be relatively easy. This is not true for the List relation. If readers will try to connect the incoming sentence to the preceding context by
means of a causal relation, they will fail to do so, but they will still have used processing time in order to try to establish the causal connection. From a top-down perspective the reasoning is as follows. Knowledge of text structure may trigger predictions about the upcoming text (Sanders, & Noordman, 2000). For instance, in the case of the Problem-Solution text, identifying the Problem part of the relation trigger the reader to expect a Solution to be described in the text. Because of this prediction, the sentence that substantiates the prediction (i.e. the sentence describing the Solution part) will be processed faster in comparison to a situation in which the text segment does not meet this expectation, or in a situation in which the expectations are absent (Graesser et al., 1994). This latter state of affairs is likely to apply to the List condition, where no or weaker expectations about the upcoming text are generated (Meyer, 1985; Meyer & Freedle, 1984; Sanders & Noordman, 2000; see also chapter 5). In short, the processing advantages of explicit indicators of text structure as well as differences between relation types can be explained in a bottom-up or a in a top-down way. In a bottom-up perspective readers try to connect sentences by means of causal inferential processes (text connecting inferences and extra textual inferences). From a top-down perspective, readers use schematic knowledge about the structure of text.

1.4 The present study

The major aim of this study is to provide further insight into the processes and representation involved in the interpretation of causal coherence relations. Now that we have seen an overview of the basic literature we can further specify the general research questions formulated in section 1.1.2. The first set of research questions addresses the representational status of coherence relations and their linguistic markers. These questions are as follows.

RQ1: What is the representational status of causal coherence relations?
RQ2: What is the representational status of linguistic indicators of causal coherence relations?

Section 1.2 explained that the causal coherence relations are considered a property of the meaning representation readers construct on the basis of the text. On the other hand, the linguistic marker of the relation is represented at the level of the surface code representation. This assumption reflects the widely shared belief in cognitive accounts of coherence relations that the surface realization of the coherence relation should be distinguished from the relation itself. In other words, this crucial assumption can also be given a psychological interpretation: the distinction between surface code and meaning representation coincides with the distinction between linguistic marking and coherence relations. The meaning representation consists of the textbase and the situation model. The textbase representation contains the propositions that are explicitly expressed in the text, but may also contain inferred connections between these propositions if these links establish local coherence. The situation model representation relates concepts from text and knowledge. On
the basis of this distinction between textbase and situation model the following ideas were put forward. When the relation is explicit, the relation is likely to be presented both at the level of the textbase and the level of the situation model. For implicit relations, in which no linguistic marker is used to indicate the type of connection, there are two possibilities. One is that a separate textbase and situation model representation are constructed. The other is that implicit coherence relations are represented only at the level of the situation model. Thus, a preliminary answer to RQ1 is that coherence relations are either represented at both the textbase and the situation model or only at the level of the situation model.

Section 1.2 also explained that the ideas presented above, are based on theoretical observations and indirect evidence. This thesis aims at obtaining direct empirical observations regarding the representational status of causal coherence relations. The experiment which aims to achieve this goal is presented in chapter 4.

The second set of research questions address the psychological/psycholinguistic process involved in the interpretation of coherence relations. The questions were as follows.

RQ3: How can the process of causal coherence relation interpretation be characterized in psycholinguistic terms?
RQ4: How can the process of the interpretation of the linguistic markers of causal coherence relations be characterized in psycholinguistic terms?

The review of the literature showed that the process of coherence relations interpretation can be characterized as a knowledge based inference process. Coherence relations can be seen as the product of bridging inferences that establish conceptual links between the concepts derived from the text. With respect to the processes involved in the interpretation of local causal relations, it was explained that two processes can be distinguished. The first process is the integration process during which the causal connection is established. The second process is an extra textual inference process in which the knowledge underlying the causal relation is activated and encoded in the cognitive representation of the text.

The presence of a connective like the causal connective because influences both the integration and the extra textual inference process. The first effect of the connective is that it speeds up the integration process. The reason for this is that the connective specifies the type of relation that holds between the clauses. By contrast, when the connective is not present the integration of the sentences requires a text connecting inference. Therefore, since a text connecting inference is not needed when the relation is explicit, processing times are expected to be shorter when the connective is present. In order to emphasize that the causal connection is established both when the connective is present and when the connective is absent (i.e. the integration process takes place regardless of the presence of the connective), the integration process will be characterized as a text connecting operation (TCO) in the explicit case and as a text connecting inference (TCI) in the implicit case. The text connecting operation consists of the encoding of the coherence relation specified by its linguistic indicator, which acts as a processing instructor: it instructs the reader to encode the relation according to the information provided by the connective.
The presence of the connective also influences the extra textual inference process. In the implicit case, integration of the sentences requires the reader to access the knowledge underlying the causal relation. During the extra textual inference process this underlying knowledge is encoded in the meaning representation of the text. This extra textual inference process also takes place when the relation is explicit. However, it is expected to proceed faster in the implicit case, because the extra textual information was already activated during the integration phase.

This cognitive processing account (CPA) of the interpretation of local causal coherence, is based on hypotheses put forward by Noordman and Vonk (1997, 1998), the ideas and empirical data central to the validation model of Singer et al. (1992) and a reinterpretation of empirical evidence of the effects of connectives obtained by Cozijn (2000) and Millis and Just (1994). The CPA is summarized in the following table.

**Table 1.1. A cognitive processing approach to the interpretation of causal relations**

<table>
<thead>
<tr>
<th>Integration</th>
<th>Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implicit (without connective)</td>
<td>Text connecting inference</td>
</tr>
<tr>
<td>Explicit (with connective)</td>
<td>Text connecting operation</td>
</tr>
</tbody>
</table>

The CPA does not only concern the types of processes that take place during the interpretation of causal coherence relations, but it also explains how the linguistic marker functions as a processing instructor: it specifies to readers how to integrate the sentences, but also that an extra textual inference should be made.

With respect to the time course of the integration and extra textual inference processes, it is assumed that the integration process takes place before the extra textual inference process (Cozijn, 2000; Noordman & Vonk, 1997). Accordingly, the speeding-up effect that pertains to the integration process is expected to occur earlier in the sentence than the slowing-down effect that is associated with the extra textual inference process. This crucial prediction will be tested in a reading time experiment presented in chapter 3.

In section 1.2, Noordman and Vonk’s (1997) approach to the processing functions of the connective *because* was presented. An interesting aspect of their hypothesis is that the integration function and inference function of *because* are related to the construction of the textbase and the situation model respectively. This hypothesis allows us to further specify the CPA in terms of the representational status of causal coherence relations. This leads to the tentative hypothesis presented in Table 1.2.

**Table 1.2. A CPA of the processing and representation of causal coherence relations**

<table>
<thead>
<tr>
<th>Process</th>
<th>Integration</th>
<th>Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implicit</td>
<td>Text connecting inference</td>
<td>Extra textual inference</td>
</tr>
<tr>
<td>Explicit</td>
<td>Text connecting operation</td>
<td>Extra textual inference</td>
</tr>
<tr>
<td>Level of representation</td>
<td>Textbase</td>
<td>Situation model</td>
</tr>
</tbody>
</table>

The CPA presented in Table 1.2 is one of the central hypotheses in this dissertation and motivates the research presented in chapters 3 and 4. In chapter 3, the integration and
The Processing and Representation of Causal Coherence Relations

Inference effects are investigated (in line with RQ3 and RQ4) and in chapter 4 the representational status of coherence relations and their linguistic markers is investigated (RQ1 and RQ2).

As was explained in section 1.6, the CPA primarily focuses on a bottom-up perspective of text processing. For instance, knowledge-based inference are triggered from the bottom-up, because of coherence requirements, or because of the presence of a connective. It was argued that some of the processing evidence can also be explained in terms of top-down influences. For example, the processing advantage found for causal relations in comparison to additive relations, which was observed by Sanders and Noordman (2000) can be explained with reference to text structural expectations readers create on the basis of the text. Therefore, another aspect of RQ3 and RQ4 is the question to what extent processing results can be exclusively explained from a bottom-up perspective. In chapter 5, an experiment is presented in which bottom-up and top-down explanations are empirically disentangled.

As secondary aim of this dissertation is to investigate RQ1 through RQ4 in experiments in which non-narrative text is used. The reason for focusing on non-narrative text is that many processing results reviewed above, have been obtained with short texts and sentence pairs describing very simple stories. As a consequence, theories and hypotheses based upon these results, may, in fact, not reflect processes involved in the comprehension of other types of text. The stories that are used in the experiments, for instance, often have a very stereotypical structure, which means that readers can be expected to have very stereotypical expectations about the events described in the text (Sanders & Noordman, 2000). Expository texts, on the other hand, often inform readers about topics unfamiliar to the reader. Hypotheses based on results obtained with simple stories may therefore overestimate the amount of inferential processing and the depth of comprehension that would normally arise in the processing of expository text (Britton, 1994; Graesser et al., 1994; Graesser et al., 1995). Therefore, it is important to move away from the bias of using simple narrative texts, and focus more on expository text processing (Van Dijk, & Kintsch, 1983). In this study, short newspaper-like articles are being used as the materials. These materials are based on naturally occurring news articles describing regional, national and international events. All of the materials were constructed in teams of advanced students and researchers in the field of text linguistics and text processing. The materials used were short articles of at most 400 words long.

Finally, the fifth research question central to this dissertation focuses on the statistical analysis of experimental data. This research question is:

RQ5: What is the preferred statistical technique for the analysis of experimental data from common designs used in discourse processing studies?

As was explained in section 1.2, researchers should preferably use more than one text in their experiments. However, using multiple items in an experiment creates problems for the statistical analysis of the experimental data. The core of these problems is the fact that using multiple items in experiments introduces a second random factor in the design, besides the
random participant factor. For this reason, all of the results of the experiments reported in chapters 3, 4, and 5 were analyzed with an alternative statistical technique: multilevel regression analyses (Goldstein, 1995) also known as hierarchical linear modeling. Because this technique is relatively unknown in the field of discourse processing, the reason why it is preferred to other, commonly used analyses, is explained in a separate chapter: chapter 2. The fact that the item factor should preferably be analyzed as a random factor, has been appreciated in the field of discourse processing ever since Clark’s (1973) article on the language-as-a-fixed-effect fallacy. However, the currently widespread practice of using separate F1- and F2-analyses, of which many researchers assume that they solve the fixed-effect fallacy, is actually not a solution to this problem. The main reason is twofold. First, the F1- and F2-analyses are irrelevant for the main purpose of the statistical analysis. This main purpose is to show that the variance created by the experimental manipulations outweighs the random variance of persons and items simultaneously. Most researchers incorrectly assume that this is what the F1- and F2-analyses do, because in F1 persons are treated as random effects and in F2 items are treated as random effects. However, as will be explained in chapter 2, doing separate analyses for the random effects of persons and items, is not the same as treating persons and items simultaneously as random effects in a single analyses. Only in the latter case, the researcher can show that the effects induced by the manipulation are not restricted to the specific samples of persons and items that were used in the experiment. The second reason is that the F1- and F2-analyses may be severely biased. As a consequence of this bias, researchers will too often conclude there is an effect while there is in fact no statistical ground for concluding this. That is, the probability of making a type I error (incorrectly rejecting the null-hypothesis) is much larger than the 5% the researcher would normally assume. In other words, the problem with F1- and F2-analyses is that they give an incorrect answer to a question that was not actually asked by the researcher. In chapter 2, it will become clear that this latter statement is less enigmatic than it at first may seem.
2.1 Introduction

From a methodological perspective, the main objective of this dissertation is to provide generalized statements about the effects of relation type and connectives on text processing and representation. This means that the aim is to draw conclusions that are not limited to the specific respondents and items that were used in the experiments. In order to accomplish this, a series of methodological and statistical choices were made. This chapter focuses primarily on the statistical choices. In particular, it discusses the use of multilevel modeling (MLM) as a tool for analyzing experimental results.

In the experiments reported in this dissertation, experimental designs were used that incorporate multiple participants. It is plausible to assume that differences between conditions are not equal across participants, because differences between participants (such as prior knowledge or reading ability) cause interaction with the experimental factor under investigation. If differences between participants are not systematically varied, participants constitute a random factor: they vary because of random (chance) factors. This is the usual interpretation of the participant factor in experimental research. Establishing the statistical reliability of the results found in an experiment amounts to determining the extent to which the systematic variability caused by the experimental manipulation outweighs the random variability associated with the participants. If the results are indeed statistically reliable (or significant) they are not merely characteristics of the participants in the sample, but also of the population from which the participants were sampled. In other words, the results can be generalized from sample to population.

Besides multiple participants, multiple items are used in the experiments reported in the subsequent chapters. As it likely that the effects found will vary from text to text, text variability will also need to be accounted for in the statistical analyses. If researchers use multiple items in the experiment, a choice will have to be made between two options for the statistical analyses. Either they will consider the effect of text to be fixed, or they will consider text to constitute a random factor. As we shall demonstrate below, this choice will have consequences for both the statistical analysis and the generalizability of the conclusions that can be drawn on the basis of the analysis.

Ever since Clark’s (1973) famous article on the language-as-fixed-effect-fallacy, researchers in the experimental tradition have recognized that items have to be treated as a random variable too (although there is some discussion, see for instance Rietveld & Van Hout, 2005). If items are treated as fixed instead of random, the statistical test is not only biased, but the conclusions drawn on the basis of the experiment do not extend beyond the specific sample of items that was used in the experiment.

Accounting for the random variability between items and participants introduces a problem for the statistical analysis of experimental results. The reason for this is that two random
factors will have to be dealt with simultaneously: participant and text. Traditional techniques such as Analysis of Variance (ANOVA) are not fully equipped for simultaneously handling participants and texts as random factors (Richter, 2005). Solutions have been offered in the literature (e.g. Clark, 1973; Raaijmakers, Schrijnemakers & Gremmen, 1999) but in most cases, the application of ANOVA remains cumbersome and for some commonly used designs even impossible.

We propose the use of multilevel modeling (MLM) as an alternative. MLM solves many of the problems associated with ANOVA and related procedures, such as Multiple Regression (MR). In the remainder of this chapter, it will be shown how designs with multiple items and participants – the type of experimental design widely used in studies of discourse processing – can be analyzed. First, ANOVA is discussed. Statistical solutions and problems will be illustrated by analyzing a simple repeated measures design. Second, the use of MLM is demonstrated. It will be shown how using this technique overcomes many of the problems associated with ANOVA.

2.2 Analysis of Variance

Let us first demonstrate the logic of the ANOVA procedure for a simple repeated measures design. In this design each participant was observed once under each condition. The design is summarized in Table 2.1.

<table>
<thead>
<tr>
<th>Respondent</th>
<th>Condition A</th>
<th>Condition B</th>
<th>Condition C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( y_{11} )</td>
<td>( y_{12} )</td>
<td>( y_{13} )</td>
</tr>
<tr>
<td>2</td>
<td>( y_{21} )</td>
<td>( y_{22} )</td>
<td>( y_{23} )</td>
</tr>
<tr>
<td>[…]</td>
<td>[…]</td>
<td>[…]</td>
<td>[…]</td>
</tr>
<tr>
<td>I (i)</td>
<td>( y_{i1} )</td>
<td>( y_{i2} )</td>
<td>( y_{i3} )</td>
</tr>
</tbody>
</table>

Note: \( y \) denotes the score on the dependent variable, the subscript \( i \) refers to the \( i \)-th person, where \( i \) ranges from 1 to \( N \) (the total number of respondents).

The ANOVA starts with specifying a linear model that describes the relations between the factors. In this case the following model applies:

\[
Y_{ij} = \mu + \alpha_i + \pi_j + \alpha\pi_{ij} + e_{ij} \\
(i = 1, 2, ..., N; j = 1, 2, ..., J)
\]

(1)

\( Y_{ij} \) is the score of the \( i \)-th participant under condition \( j \), \( \alpha_i \) is the effect of condition \( j \), \( \pi_i \) the effect of participant \( i \), \( \alpha\pi_{ij} \) the interaction between condition and subject, and \( e_{ij} \) is a residual or error.
The significance of the effect of the Treatment is tested with an F-test: an F-ratio is calculated by dividing two Mean Squares (variances) that only differ with respect to the component associated with the effect of the Treatment under consideration. The resulting F-ratio can be tested for significance by comparing its value with the F-distribution. To illustrate the logic of the test consider the expected mean squares in Table 2.2.

Table 2.2 Expected Mean Squares for design 1

<table>
<thead>
<tr>
<th>Factor</th>
<th>Df</th>
<th>Expected mean squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>t-1</td>
<td>$n\sigma^2_{\theta} + \sigma^2_{\text{tp}} + \sigma^2_{\epsilon}$</td>
</tr>
<tr>
<td>Participant</td>
<td>n-1</td>
<td>$\sigma^2_{\theta} + \sigma^2_{\text{tp}} + \sigma^2_{\epsilon}$</td>
</tr>
<tr>
<td>Treatment x Participant</td>
<td>(t-1)(n-1)</td>
<td>$\sigma^2_{\text{tp}} + \sigma^2_{\epsilon}$</td>
</tr>
</tbody>
</table>

Note: t = number of treatment levels, n = number of participants

As the table shows, the difference between the expected mean square for the treatment (MST) and for the treatment by subject interaction (MSTxP) is the presence of $n\sigma^2_{\theta}$. This is the component associated with the effect of the treatment. The larger the difference between the means is, the higher the value of the component will be, and the higher the ratio of MST and MSTxP will be. Under the null hypothesis the resulting F–ratio follows an F-distribution, therefore the difference between the means can be tested for significance. The analysis of this design is straightforward and can be executed with standard statistical software (for instance with the Repeated Measures ANOVA procedure in SPSS).

Now let us consider a design in which multiple items are used. An example of such a design is presented in Table 2.3. In this design, a participant is presented with items 1 to 4 under condition A, items 5 to 8 under condition B and 9 to 12 under condition C.

Table 2.3 Repeated measures with item nested under condition

<table>
<thead>
<tr>
<th>Respondent</th>
<th>Condition A</th>
<th>Condition B</th>
<th>Condition C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Y111 Y121 Y131 Y141</td>
<td>Y152 Y162 Y172 Y182</td>
<td>Y193 Y210 Y211 Y212</td>
</tr>
<tr>
<td>2</td>
<td>Y211 Y221 Y231 Y241</td>
<td>Y252 Y262 Y272 Y282</td>
<td>Y293 Y210 Y211 Y213</td>
</tr>
<tr>
<td>[...]</td>
<td>[...]</td>
<td>[...]</td>
<td>[...]</td>
</tr>
<tr>
<td>1</td>
<td>Y111 Y121 Y131 Y141</td>
<td>Y252 Y262 Y272 Y282</td>
<td>Y303 Y310 Y311 Y312</td>
</tr>
</tbody>
</table>

For this design the following linear model applies:

$$Y_{ik} = \mu + \alpha_i + \beta_{j(k)} + \pi_{ij} + \omega_{ik} + \beta\pi_{ij(k)} + \epsilon_{ik}$$

$$i = 1, 2, ..., N; j = 1, 2, ..., F; k = 1, 2, ..., K$$

5 As the variance associated with the Treatment by participant interaction cannot be separated from the error variance, error variance is not incorporated as a separate factor.

6 To indicate random variance $\sigma^2$ is used, to indicate variance associated with a fixed factor $0^2$ is used.
The additional parameters in model (2), as opposed to model (1), are $\beta_j$, the effect of item (within treatment), and $\alpha_{ij}$, the interaction of item (within treatment) and subject.

The test of the treatment effect would, like before, involve constructing the appropriate F-ratio. The first step is to specify the expected mean squares. In Table 2.4, the expected mean squares are given for each factor in the analysis. Here, the treatment factor is considered as fixed and the participant and item factors as random.

**Table 2.4 Expected Mean Squares for design 2.**

<table>
<thead>
<tr>
<th>Factor</th>
<th>df</th>
<th>Expected Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>t-1</td>
<td>$\theta T_{iT}^2 + n \sigma_{T}^2 + i \sigma_{T}^2 + \sigma_{1}^2 + \sigma_{2}^2$</td>
</tr>
<tr>
<td>Item within Treatment</td>
<td>t(i-1)</td>
<td>$\theta T_{i}^2 + \sigma_{iT}^2 + \sigma_{j}^2$</td>
</tr>
<tr>
<td>Participant</td>
<td>n-1</td>
<td>$t \sigma_{i}^2 + \sigma_{iT}^2 + \sigma_{j}^2$</td>
</tr>
<tr>
<td>Treatment x Participant</td>
<td>(t-1)(n-1)</td>
<td>$i \sigma_{T}^2 + \sigma_{iT}^2 + \sigma_{j}^2$</td>
</tr>
<tr>
<td>Item x Participant</td>
<td>(t-1)(n-1)</td>
<td>$\sigma_{T}^2 + \sigma_{iT}^2 + \sigma_{j}^2$</td>
</tr>
</tbody>
</table>

In constructing the appropriate F-ratio we now face a problem: there is no MS available for testing the effect of treatment, because no MS differs from the Treatment MS only with respect to the component $\theta T_{iT}$, which is the component associated with the effect of Treatment. In a regular F-test, two MS’s are selected that differ only with respect to the presence of the treatment component. Because these are not available in the current analysis, the regular F-test cannot be performed. In stead, researchers will have to resort to an approximation of the F-test by calculating a quasi-F. Unfortunately, this cannot be done fully automatically with standard statistical software like SPSS or SAS. In the next section it will be demonstrated how a quasi-F can be calculated for the current design.

### 2.2.1 Calculating quasi F-ratios

The procedure for testing the Treatment effect in a design with the two random variables participant and item the procedure is as follows. In the first step, all the relevant factors in the design are identified. In the second step, the expected mean squares of all factors are determined. In the third step, the quasi-F ration is constructed. This is done in such a way that the numerator only differs from the denominator with respect to the component that is associated with the treatment effect. In the fourth step, the quasi-F ratio is calculated. In order to do so, the observed mean squares have to be obtained. In the fifth step, an approximation of the degrees of freedom is calculated. In the sixth and final step, the tail probability of the quasi-F is determined, by referring to the F distribution.

A major drawback of this procedure is that there is no unified solution for every design. The expected mean squares vary from design to design, and therefore also the construction and calculation of the quasi-F also differs from design to design (see Meuffels & Van den Bergh (2006) for an overview of the construction of the quasi-F for some common research designs).

A second disadvantage of the procedure is that standard statistical software can only be used
in obtaining the observed mean squares. The quasi-F-ratio and its associated degrees of freedom have to be calculated manually. In all, calculating the quasi-F-ratio requires a firm understanding of the relationship between experimental design and expected mean squares and how the relevant observed mean squares can be obtained.

Let us review the procedure for a repeated measures design like the one in Table 2.3. The relevant factors and expected mean squares (obtained in steps 1 and 2) are presented in Table 2.4 above. Rules for obtaining these expected mean squares can be found in the statistical literature (e.g. Jackson & Brashers, 1994; Winer, 1971).

The test of the treatment effect can be accomplished by constructing the following quasi-F (Clark, 1973).

\[
F' = \frac{MS_T + MS_{Txp}}{MS_{T} + MS_{iT}}.
\]

If we express this in terms of the expected mean squares, we can see that the component associated with the treatment effect, is only present in the numerator, whereas all the other components are present both in the numerator and denominator.

\[
E(MS_T) + E(MS_{Txp}) = \frac{(n\sigma_T^2 + n\sigma_{T}^2 + n\sigma_{T}^2 + \sigma_{T}^2 + \sigma_{T}^2 + \sigma_{T}^2 + \sigma_{T}^2 + \sigma_{T}^2 + \sigma_{T}^2)}{(n\sigma_{T}^2 + n\sigma_{T}^2 + n\sigma_{T}^2 + \sigma_{T}^2 + \sigma_{T}^2 + \sigma_{T}^2 + \sigma_{T}^2 + \sigma_{T}^2 + \sigma_{T}^2)}
\]

The required mean squares can be obtained by performing two MANOVA procedures in SPSS (Meuffels & Van den Bergh, 2006). For the purpose of illustration, we have analyzed simulated data for 15 participants who were presented with 5 items per condition (see appendix for the data and the SPSS syntax). In the first MANOVA, SPSS tests the treatment effect against the interaction of respondent and treatment. From this first run, the mean squares of the treatment (MST) and the interaction of respondents and treatments (MSTxp) are selected. In the second run, we used MANOVA to test the factor item nested under treatment. SPSS tests the effect of item (within treatment) against the interaction between respondents and items (within treatment). From the second analysis we selected the Mean Square associated with item (MSi(T)) and for the interaction of item and participant (MSi(T)xp).

The two analyses have provided us with the ingredients for calculating the value of F' (of Equation (3)):

<table>
<thead>
<tr>
<th>Table 2.5 Ingredients for calculating F'.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
</tr>
<tr>
<td>Treatment (MS_T)</td>
</tr>
<tr>
<td>Treatment x Participant (MS_{T}xp)</td>
</tr>
<tr>
<td>Item x Participant (MS_{i(T)}xp)</td>
</tr>
<tr>
<td>Item (MS_{i(T)})</td>
</tr>
</tbody>
</table>
The obtained approximation of the degrees of freedom is rounded to the nearest integer.

In our case, the following simplified formula applies (Jackson & Brashers, 1994):

\[
df' = \frac{(MS_I + MS_T)^2}{df_i + df_T}
\]

(6)

For the denominator of the F'-value in the present example, applying this formula leads to:

\[
df'_{den} = \frac{(MS_T + MS_{(I)})^2}{df_{(I)}^* + df_{(T)}^*} = 14.98
\]

(7)

Applying the same formula for the calculation of the degrees of freedom of the numerator we get a value of 2.02. Thus, the F' for the treatment effect is 24.18 with 2 degrees of freedom for the numerator and 15 for the denominator. This value is significant at the 0.001 level (F'(2,15) = 24.18, p < .001).

The example analysis shows how complicated the analysis of even a fairly simple design can be. However, there is also a reward: since we have incorporated both items and participants as random factors, the obtained result can be generalized to both a population of items and a population of subjects simultaneously. In the end, this is the main reason for performing this analysis.

Given the practical complexities involved in the calculation of F' and the degrees of freedom, it comes as no surprise that alternative computational procedures are being used. One way to deal with the problem is to make use of the minimal value of F', in short minF'. If this minimal value of F' is significant, F' itself is, of course, also significant. The minimum value of the quasi F can be quite easily calculated.

According to Clark (1973), the value for minF' can be calculated by making use of F1 and F2. As we explained above, F1 corresponds to the F that is obtained in an analysis of variance after averaging the scores per condition for each respondent by collapsing over items. The F2 value on the other hand is obtained by performing an analysis of variance on average scores per condition per item by collapsing over respondents. On the basis of F1 and F2, minF' can be calculated by using (8):
\[ \text{minF'}(i, j) = \frac{F_1 F_2}{F_1 + F_2} \tag{8} \]

The degrees of freedom of the numerator (i) are the same as the degrees of freedom in the numerator of both the F1 and the F2. The degrees of freedom for the denominator (j) is the nearest integer of the result of the following formula, where \( n_1 \) equals the degrees of freedom for the numerator of F1 and \( n_2 \) equals the degrees of freedom for the numerator of F2 (see Clark, 1973, p. 347).

\[ j = \frac{(F_1^2 + F_2^2) n_1 n_2}{n_1 n_2} \tag{9} \]

Let us demonstrate the use of F1 and F2 for calculating minF’ by making use of the dataset from the previous section. The results of the analyses are presented in Table 2.6.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>F</th>
<th>df(_{\text{num}})</th>
<th>df(_{\text{den}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject (F1)</td>
<td>265.747</td>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>Item (F2)</td>
<td>26.945</td>
<td>2</td>
<td>14</td>
</tr>
</tbody>
</table>

The value for minF’ is obtained by applying formula (8) to the values obtained in the F1 and F2 analyses.

\[ \text{minF'} = \frac{F_1 F_2}{F_1 + F_2} = \frac{265.747 \times 26.945}{265.747 + 26.945} = 24.46 \tag{10} \]

The degrees of freedom for the numerator equal 2, and the degrees of freedom for the denominator equal the nearest integer that is the result of the following formula:

\[ df_{\text{den}} = \frac{(F_1^2 + F_2^2) n_1 n_2}{n_1 n_2} \tag{11} \]

In the case of the present example, the result of this formula is:

\[ df_{\text{den}} = \frac{(265.747^2 + 26.945^2) 15.009}{14 + 28} = 15.009 \tag{12} \]
In this case, a minF' value of 24.46 with 2 degrees of freedom for the numerator and 15 for the denominator is highly significant (p < .001). Thus, in the present example, using minF' as a replacement of F' will also lead to rejection of the null hypothesis. Of course, since minF' is the lower bound of F', it may be possible that F' is significant but minF' is not.

The example analysis shows that for the design in Table 2.2 there are ANOVA solutions available for dealing with the problem of the two crossed random factors. In general, however, F' and minF' are very rarely present in published work. One of the reasons for this is probably that calculating F' can be very complicated, and is sometimes even impossible. For instance, in designs with structural empty cells F' cannot be calculated (cf. Richter, 2005).

Another reason is that F' is known to be conservative (Meuffels & Van den Bergh, 2006; Raaijmakers, Schrijnemakers & Gremmen, 1999; Richter, 2005).

The conservative behavior of F' may have lead to the current statistical practice of only reporting F1 and F2 (Richter, 2005). Researchers who use this so called F1 x F2 criterion (Raaijmakers, Schrijnemakers & Gremmen, 1999) will conclude that there is a Treatment effect when both the F1 and F2 tests prove to be significant. In doing so, they incorrectly assume that using this criterion equals the minF' procedure in which item and participant variability is accounted for simultaneously (Raaijmakers, Schrijnemakers, & Gremmen, 1999). The example analysis confirms this: F1 and F2 are both larger than minF' although for F1 the difference is the most dramatic (see Table 2.6). Due to this bias, the use of the F1 x F2 criterion might lead to the rejection of the null hypothesis while in fact there was no effect of the Treatment.

The bias in the F1- and F2-analyses is a result of the fact that in both analyses a random factor is assumed to be fixed. In the F1-analysis, item is considered to be a fixed factor. This was pointed out by Clark (1973), and because item is treated as a fixed factor in stead of a random factor Clark named the F1-analysis an instance of performing the language-as-fixed-effect fallacy. In the F2-analysis participant is treated as a fixed factor. Since in both analyses a random factor is considered fixed the generalization of the effect that is found is restricted to either the sample of items (F1) or the sample of participants (F2) but not, and this is important, to two populations simultaneously. In the F1-analyses the effect is only generalizable to a population of participants to whom the exact same items are being presented, and in the F2-analyses only to a population of items that are presented to the exact same participants. Thus, the F1 x F2 criterion is entirely inadequate, not only from a statistical point of view (i.e. the results are biased), but also from a methodological point of view.

2.2 ANOVA: preliminary conclusion

In the previous sections we have shown how complicated the analysis of even simple designs can be if we use Analysis of Variance. In order to get an F' or minF' value, several analyses have to be performed and the degrees of freedom have to be calculated for each specific analysis.
Calculating \( \text{minF}' \) by making use of \( F_1 \) and \( F_2 \) is easier to accomplish than calculating \( F' \). This might lead one to expect that \( \text{minF}' \) can be used in every possible design. This is not true, however. Indeed, this use of \( F_1 \) and \( F_2 \), (that is averaging over items or respondents) is suitable only for the analysis of designs in which items are nested under treatments and if there are no other factors that need to be considered. In any case, reporting and interpreting \( F_1 \) and \( F_2 \) separately is never a statistically sound procedure, as we explained above. In the nested designs in the previous sections, one aspect of the problem is that differences between the treatment levels are partly due to random variability between items. The statistical analyses presented above, are statistical solutions for controlling for this random variability. If the random effect of item (and participant) is accounted for, a significant treatment effect shows the true differences between the treatments. The researcher could also have used a different experimental design to control for the random effect of items. A frequently used design is the Latin Square Confounded Factorial design (see: Kirk, 1982). This design is illustrated in Table 2.7.

<table>
<thead>
<tr>
<th>Respondents</th>
<th>Treatment A</th>
<th>Treatment B</th>
<th>Treatment C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>items 1 2 3 4</td>
<td>items 5 6 7 8</td>
<td>items 9 10 11 12</td>
</tr>
<tr>
<td>Group 2</td>
<td>items 9 10 11 12</td>
<td>items 1 2 3 4</td>
<td>items 5 6 7 8</td>
</tr>
<tr>
<td>Group 3</td>
<td>items 5 6 7 8</td>
<td>items 9 10 11 12</td>
<td>items 1 2 3 4</td>
</tr>
</tbody>
</table>

In the design portrayed in Table 2.7, each respondent is randomly assigned to one of the three participant groups. We can see that for group 1, items 1 through 4 are presented under treatment level A, items 5 through 8 under condition B, and items 9 through 12 under treatment C. Group 2 and 3 are presented with the same items but under different treatment conditions than in group 1. Thus all groups differ with respect to the items the groups are presented with under each level of the treatment factor.

The major advantage of using this design is that each item is presented under each level of the treatment factor. This means that the main effect of item does no longer contribute to the difference between the treatments, when compared to a nested design. However, applying ANOVA to data from Latin Square designs is very complex. The reason for this is that it is virtually impossible to construct the relevant (quasi) F-ratio. Solutions have been offered in literature, for example, Winer (1971, p. 371-375) gives rules for expected mean squares for all possible designs. However, these rules provide “little insight into the mathematical rationale underlying the end product” (p. 371). Furthermore, the application of these rules is not always simple. Besides, this requires that all the relevant Mean Squares have to be estimated as well, before the correct F-ratio can be derived. Not surprisingly, these statistical complexities have led to (incorrect) simplifications.

The computational complexities involved in performing ANOVA, are a direct result of taking account of item as a random effect. When item is considered as a fixed effect, the analysis is far less cumbersome. The question is whether the lack of generalizability to a population of items is the only reason for considering items as a random effect, or if there are more reasons for not treating items as a fixed factor? After all, a researcher might be interested in providing
(merely) existential evidence by showing that it is possible in principle to obtain a specific result. One of the possible drawbacks is that when item is incorporated in the analysis as a fixed factor, it becomes impossible to enter item-level explanatory variables (like difficulty or text length) in the analysis as covariates. This is a result of the fact that when a fixed item factor is present any systematic differences between the items are linearly related to the item-factor. Because of this multi-collinearity the ANOVA procedure cannot uniquely identify variance associated with items and variance associated with the text level explanatory variables (Winer, 1971). As a result, the effects cannot be adjusted for systematic differences between the items (besides the treatment effect), which may have a direct bearing on the reliability of the analysis. Traditionally, this problem has led to averaging over items, such that systematic differences attributable to items, no longer cause variance in the dependent variable (Richter, 2005). However, this also results in the unnecessary loss of useful information. (Quené & Van den Bergh, 2004; Richter, 2005) and to biased estimates of the Treatment effect (as in the F₁-analysis). If item is treated as a random effect, more detailed analyses of item-level explanatory variables are possible.

In sum, the analysis of repeated measures data in which two random variables are crossed (in this case respondents and items) can be done with analysis of variance, but this requires several analyses; it can not be done with a standard procedure. For each design the relevant mean squares have to be specified and obtained, before a correct F-ratio can be calculated. This makes the use of ANOVA cumbersome, since there is no unified solution for every possible design.

Laying the practical complexities aside, ANOVA can lead to unreliable results because it requires underlying statistical assumptions to be met, of which some are hardly ever met (Quené & Van den Bergh 2004). Among these assumptions are homoscedasticity (i.e. equal variance in the treatment groups) and sphericity (i.e. equal variance of the difference scores between conditions). When these assumptions are not met, the ANOVA model is no longer valid, and estimates based upon the model become unreliable.

A third drawback of ANOVA is that the information the data provide is not fully used. The procedure will give an F-ratio or quasi-F ratio to indicate whether the effect of interest is significant, but it is not possible, especially in the case of quasi-F, to further investigate the residual variances.

Multilevel modeling (MLM) promises to solve each of the problems described above. First, the basic model applies independent of the design that is chosen. That is, the procedure does not depend on the specific design in the way that the calculation of F does, and most of the analyses involved can be done fully automatically. Therefore, from a practical perspective, MLM is much more appealing than ANOVA. Second, MLM enables modeling and testing underlying statistical assumptions directly. Third, because MLM can separate various variance components simultaneously, inspection of the residuals is possible. As will be shown below, inspection of the residual variance associated with items can lead to the formulation of new hypotheses which will ultimately lead to further theoretical progress. In the next section the MLM procedure will be illustrated by analyzing the sample data from design 2.
2.3 Multilevel modeling

Multilevel analysis or multilevel modeling (Goldstein, 1995) is a technique for (multiple) regression analysis. Unlike standard Ordinary Least Squares (OLS) regression, the technique takes into account that data may be clustered (or nested) within larger units. In repeated measures designs this clustering is found at the level of the subject: the measurements of a person are nested within this person. Another way of saying this is that the measurements of a person are correlated or dependent.

It is important that we do not ignore the fact that the data are dependent. The reason is that due to the dependency between the measurements of the same person, the effective sample size is smaller than the nominal sample size. Overestimating the effective sample sizes leads to underestimation of standard errors associated with parameters, since the standard error is a function of the sample size. As many statistical tests use (derivations of) the standard error, underestimation of the standard error will lead to incorrect conclusions about the significance of parameters. Traditional Ordinary Least Squares (OLS) regression does not take into account the dependency of the data from the same person and therefore cannot be used for designs with repeated measures. Of course, this has been recognized and General Linear Modeling (GLM) procedures have been developed to account for dependency of the data. An example is the Repeated Measures ANOVA.

In MLM, dependency is taken into account by simultaneously modeling residual terms at different levels of the sampling. In a basic two-level model, the dependency is found between the lowest level units (level 1) which are nested in the highest level units (level 2). For example, in the simple repeated measures design 1 the data within a person are dependent. In this case, then, participants are situated at level 2 and the individual measurements at level 1. Or, stated differently, the measurements of a person are nested within this person.

The basic two level model can be described as follows.

\[
Y_{ij} = \beta_0 + (u_{0j} + e_{0ij}) \\
(i = 1, 2, \ldots J; j = 1, 2, \ldots J)
\]  

In this model, \(Y_{ij}\) is the score of item \(i\) nested under person \(j\). \(\beta_0\) indicates the intercept term, which, in this model, represents the population mean. It is part of the fixed part of the model. \(u_{0j}\) is a person level residual (or level 2 residual), it indicates the difference between the general mean \(\beta_0\) and the mean of person \(j\). The term \(e_{0ij}\) is the within person residual (or level 1 residual), which indicates the difference between the person mean and the score on item \(i\) for that person. The presence of the two residual terms distinguishes the MLM model from standard OLS regression models. In the latter models, only one error term is present. In MLM it is assumed that both the person level residual \(u_{0j}\) and the within person residual \(e_{0ij}\) follow a normal distribution with a mean of zero and variance \(\sigma^2_{0j}\) for the person level residual and \(\sigma^2_{e}\) for the within person level residual.

In MLM, the parameters are estimated by making use of iterative procedures. First the coefficients of the fixed part of the model are estimated, and then values of the random
parameters. In the next step the parameters in the fixed part of the model are estimated again, by making use of the previously estimated values for the random part of the model. The estimation procedure ends, when the change in the estimates is below a certain threshold (e.g. less than 1% change in the estimates between successive iterations). The parameters can be estimated with several estimation procedures. Examples of such procedures are Iterated Generalized Least Squares (IGLS) (Goldstein, 1995) and Expectation-Maximization (EM) (Bryk & Raudenbush, 1992; Raudenbush & Bryk, 2002). These procedures produce the same estimates as long as the residuals follow a normal distribution and have a zero mean.

In MLM testing the effect of the treatment may proceed as follows. First a model is fitted that contains no explanatory variables. Such a model is therefore called an empty model. Model (13) above is an example of such an empty model. The empty model consists of the general population mean (the intercept term $\beta_0$) and residual terms associated with each of the levels of the model. Special purpose software like MLwiN (Rasbash, Browne, Goldstein, Yang, Plewis, & Healy, 2000), estimates the value of the intercept term and the distribution of the residuals or, in other words, it estimates the residual variance at all levels in the model. In the case of model (13), residual variance between participants (level 2) and within participants (level 1) will be estimated. Also, a goodness-of-fit index will be provided. This index plays a crucial role in subsequent steps of the modeling procedure, as will be demonstrated below.

Second, additional parameters are added to the model. For instance, parameters associated with the effect of the treatment may be estimated. This has been done in the following model.

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (u_{ij} + e_{ij})$$

The effect of treatment is incorporated in the model by making use of two dummy variables $x_1$ and $x_2$. If indicator coding is used, these explanatory variables have the value 1 if the measurement was obtained in the respective conditions and the value 0 if the measurement was obtained in another condition. The regression weight $\beta_1$ represents the mean of the reference condition. If the reference condition is condition 1, $\beta_1$ is the difference between the conditions 2 and 1, and $\beta_2$ represents the difference between condition 3 and the condition 1.

The new model will yield a different goodness-of-fit index than the empty model. If the model provides a better goodness-of-fit than the empty model, the new model is selected for further analysis. One aim in the analysis procedure is to find the model that gives the best goodness-of-fit with the lowest number of parameters. Thus, part of the analysis of data with MLM involves choosing the most economical model that captures all relevant aspects of the data.

As an illustration of how the analysis works, the example dataset of design 2 was analyzed with the first two models. These first two models were chosen because Quené & Van den Bergh (2004) and Richter (2005) have used similar models in their articles on the use of MLM in language research. The program MLwiN was used to obtain the estimates, which are presented in the following table.
Table 2.8 Parameter estimates of models 13 and 14.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model 13</th>
<th>Model 14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>2.261 (0.148)</td>
<td>0.814 (0.170)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>1.980 (0.145)</td>
<td></td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>2.362 (0.145)</td>
<td></td>
</tr>
<tr>
<td>Random</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{\epsilon_0}$</td>
<td>0.199 (0.120)</td>
<td>0.275 (0.120)</td>
</tr>
<tr>
<td>$\sigma^2_{\epsilon_1}$</td>
<td>1.939 (0.189)</td>
<td>0.791 (0.077)</td>
</tr>
<tr>
<td>-2*loglikelihood</td>
<td>801.53</td>
<td>613.30</td>
</tr>
</tbody>
</table>

The difference between models (13) and (14) in terms of parameters is located in the fixed part of the model. In the fixed part of model (13) only the general population mean ($\beta_0$) is estimated. The population mean equals 2.261 and its standard error is 0.148. In the fixed part of model (14), the effect of condition is incorporated. Here, $\beta_0$ represents the mean in condition 1, which equals 0.814 with a standard error of 0.170. $\beta_1$ represents the difference between conditions 1 and 2. The difference between the means is 1.980, and the standard error of this difference is 0.145. $\beta_2$ is the difference between the means of conditions 3 and 1. This difference is 2.362 and its standard error is 0.145.

The significance of the parameters in the fixed part of the model can be tested by dividing the estimate by its standard error. This results in a large sample z-value. If the absolute value is 1.96 or larger, the parameter is significantly different from 0 (2-sided; $\alpha = .05$). Here, both the difference between conditions 1 and 2 ($z = 13.66$) and the difference between conditions 1 and 3 ($z = 16.29$) are highly significant.

The -2*loglikelihood (-2*ll) provided by MLwiN indicates the goodness-of-fit of the model. For model (13) the value is 801.53 for model (14) it is 613.3. The smaller the value of -2*ll is, the better the goodness-of-fit of the model is. In model (14), -2*ll is smaller than that of model (13). This means that incorporating the difference between the conditions in the model increases its goodness-of-fit. However, in model (14) two additional parameters were estimated. As the fit of a model increases when additional parameters are estimated, it is necessary to check whether model (14) provides a significant better fit than model (13).

As was said above, the goodness-of-fit index is used to assess whether a subsequent model provides a better fit to the data. If models are related hierarchically, that is, if one model is nested in another model, the difference between the -2*ll follows a chi-squared distribution with the difference in the number of parameters as its associated degrees of freedom. This means that it is possible to test whether adding additional parameters provides a significant better fit than the previous model. In this case, the models are hierarchically related, since restricting the two additional parameters in model (14) to zero yields model (13). The difference in -2*ll of the models is 188.23, and the difference in the number of parameters is 2; model (14) provides a significant better fit to the data than model (13) ($\chi^2(2) = 188.23, p < .001$). This means that incorporating separate estimates for the condition means has improved the goodness-of-fit of the model significantly.
The fact that the second model provides a better fit of the data than the first model means that the effect of condition is significant. Since random variability between and within persons (between items) is accounted for, the effect that is found can be generalized to the two populations. If we compare this to the amount of work that was needed to concluded this with the use of $F'$, it becomes immediately apparent that MLM provides a much easier way of accounting for the random effect of participant and item than ANOVA.

The estimates of the random part of the model consist of the level 2 variance ($\sigma^2_{ju}$) and the level 1 variance ($\sigma^2_{ije}$). Both Quené and Van den Bergh (2004) and Richter (2005) interpret the level 2 variance as between participant variance and at level 1 the variance represents between item variance (within participants). However, in design 2 items are not nested within participants but crossed with participants: every participant is presented the same items, and not different items. This means that the models described by the authors do not fully capture the design. In this dissertation only designs are used in which participants and items are crossed, therefore, in the following models will be used in which this crossing is accounted for.

A second difference between the models presented above and the models that will be used throughout this book, is that in this book cell means models will be used in stead of effect models like model (14). In a cell means model, all means in the cells of the experimental design are estimated, in stead of the mean in a reference condition and parameters that represent the differences between conditions, as is done in effect models. The cell means model and the effect model are equivalent, but in MLwiN, there are circumstances when the cell means model provides a more convenient way of data interpretation.

As an illustration, the cell means variant of model 2 reads as follows.

$$y_{ij} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + (u_{ij} + e_{ij})$$

(i = 1, 2, ..., I; j = 1, 2, ..., J) \hspace{1cm} (15)

The explanatory variables $x_1$ through $x_3$ are dummy variables that have the value 1 if the measurement was obtained in the respective conditions, and 0 if the measurement was obtained under a different condition.

In Table 2.9 the estimates of this model are presented next to the estimates of model 2.
Table 2.9 Parameter estimates of effect and cell means model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model 14</th>
<th>Parameter</th>
<th>Model 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed (\beta_0)</td>
<td>0.814 (0.170)</td>
<td>Fixed (\beta_1)</td>
<td>0.814 (0.170)</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>1.980 (0.145)</td>
<td>(\beta_2)</td>
<td>2.794 (0.170)</td>
</tr>
<tr>
<td>(\beta_2)</td>
<td>2.362 (0.145)</td>
<td>(\beta_3)</td>
<td>3.176 (0.170)</td>
</tr>
<tr>
<td>Random (\sigma^2_{\nu})</td>
<td>0.275 (0.120)</td>
<td>Random (\sigma^2_{\nu})</td>
<td>0.275 (0.120)</td>
</tr>
<tr>
<td>(\sigma^2_{\nu})</td>
<td>0.791 (0.077)</td>
<td>(\sigma^2_{\nu})</td>
<td>0.791 (0.077)</td>
</tr>
<tr>
<td>(-2)\text{loglikelihood}</td>
<td>613.30</td>
<td>(-2)\text{loglikelihood}</td>
<td>613.30</td>
</tr>
</tbody>
</table>

The values of the \(-2\)\text{ll} indicate that the models are equivalent. In the cell means model the difference between the condition means cannot be obtained directly as is the case in the effect model. In the cell means model, differences between conditions are tested by specifying contrasts for the fixed parameters. Each of the contrasts corresponds to an amount of variance that is accounted for by this contrast. The amount of variance accounted for by a contrast follows a chi-squared distribution (Winer, 1971, p. 849). This means that the significance of a contrast can be tested against a chi-squared distribution with 1 degree of freedom (see: Goldstein, 1995). The difference between conditions 1 and 2 and the difference between condition 1 and 3 is significant (condition 1 versus 2: \(\chi^2(1) = 185.67, p < .001\); condition 1 versus 3: \(\chi^2(1) = 264.34, p < .001\)).

2.4 Multilevel modeling of random cross classifications

As was argued above, Quené and Van den Bergh (2004) and Richter (2005) model the between item variance as variance fully nested within participants. In a design in which participants and item are crossed, there is not only variance between items within participants but also variance between items across participants. This variance between items across participants can be obtained by using a model with cross classification.

In the case of repeated measurements, like the design presented in Table 2.1 above, three sources of random variation need to be considered: subjects, items and the interaction between subjects and items (the latter source cannot be distinguished from random noise). From a conceptual point of view the data are hierarchically structured; observations are nested within individuals. The same applies to items, since for each item there are multiple observations. Therefore, subjects and items are considered at a higher level than the actual observations.

Let \(Y_{i(jk)}\) be an observation nested within the combination of subject \(j\) and item \(k\). As subject and item are experimentally independent, the correlations between both will be zero. Now, the empty model can be written as:

\[
Y_{i(jk)} = \beta_k + (\mu_{i(m)} + \nu_{(m)} + \epsilon_{i,k})
\]

\((i = 1, 2, \ldots I_{(j)}; j = 1, 2, \ldots J; k = 1, 2, \ldots K)\)

(16)
The additional term $v_{0(0k)}$ now represents the difference between the population mean and the item mean. The term $e_{i(jk)}$ represents the residual associated with the interaction of participant and item and random error. The residuals are assumed to follow a normal distribution with an expected mean of zero and a variance of respectively $\sigma^2_u$, $\sigma^2_v$, and $\sigma^2_e$.

For modeling the data according to Table 2.1, this model does not satisfy, as the three conditions are not represented in this model. If we define three dummy variables, we can estimate the condition means. The model now can be written as:

$$y_{i(jk)} = \beta_0 + \beta_1 x_i + \beta_2 x_j + \beta_3 x_k + (u_{i(jk)} + v_{i(jk)} + e_{i(jk)})$$

where $i=1,2,...,I$, $j=1,2,...,J$, $k=1,2,...,K$.

As before, $\beta_1$, $\beta_2$, and $\beta_3$ are estimates of the population treatment means for conditions 1, 2, and 3, respectively.

The various estimates of the parameters in models (16) and (17) are presented in Table 2.10.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model 16</th>
<th>Model 17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>2.261 (0.330)</td>
<td>0.841 (0.263)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td></td>
<td>2.794 (0.263)</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td></td>
<td>3.176 (0.263)</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_u$</td>
<td>0.308 (0.131)</td>
<td>0.298 (0.125)</td>
</tr>
<tr>
<td>$\sigma^2_v$</td>
<td>1.980 (0.315)</td>
<td>0.209 (0.093)</td>
</tr>
<tr>
<td>$\sigma^2_e$</td>
<td>0.579 (0.058)</td>
<td>0.574 (0.058)</td>
</tr>
<tr>
<td>$-2\log$likelihood</td>
<td>597.89</td>
<td>572.95</td>
</tr>
</tbody>
</table>

The table shows, that, as before, adding the difference between the conditions in the model, decreases the $-2\log$. In this case, the difference between the models is significant ($\chi^2(2) = 24.49, p < .001$). This means that there is a main effect of the treatment factor. For the more detailed analyses of the differences between the condition means, contrasts will be formulated for the fixed part of the model. If the all the pairwise differences between the means are investigated, it may be prudent to adjust the significance level for each test, because multiple pairwise comparisons will lead to inflation of the experiment wide alpha level. For instance, a Bonferroni correction can be applied, by dividing the significance level by the number of pairwise comparisons. In this case, this will amount to dividing the alpha level by 3. For model (17) the pairwise comparisons show that there is a significant difference between the means of condition 1 and 2 ($\chi^2(1)=39.57, p <.001$) and between the means of conditions 1 and...
Multilevel Modeling of Experimental Data

The difference between the conditions 2 and 3 is not significant ($\chi^2(1) = 1.48, \ p = .22$). As random variability between items and between participants is accounted for, the effects that are obtained can be generalized to two populations simultaneously. This demonstrates once again that the analysis of a design with two crossed random factors is very straightforward in MLM. Two models are formulated, one with and one without the treatment factor. The difference in goodness-of-fit provides an immediate test of the treatment effect. If the second model provides the best fitting model, this model is selected for more detailed comparisons between the means and for additional modeling of relevant factors. Of course, adding additional parameters to the model, may lead to the formulation of a model in which the treatment effect is no longer significant.

2.5 Additional differences between MLM and ANOVA

Besides the practical advantages for dealing with designs with two random factors, there are several other characteristics of MLM that make it a valuable research tool. This was demonstrated by Quené and Van den Bergh (2004) and Richter (2005). First of all, the same basic model can be applied to both simple and more complex designs. For instance, the basic model for the cross-classified designs can be applied to more elaborate cross-classified designs without much trouble. The basic model (16), for example, can readily be applied to Latin Square Designs that are very hard to analyze with ANOVA.

Second, the researcher does not have to make the same crucial assumptions about the data as he has to make when using ANOVA. To be more precise, for valid repeated measures ANOVA the data have to show compound symmetry (or at least sphericity) and the variance has to be homoscedastic. The data show compound symmetry if all covariances between the conditions are equal, the data are homoscedastic if the variances in all conditions are equal. In practice, these requirements are hardly ever met (Quené & Van den Bergh, 2004). In MLM, these properties do not need to be assumed. Rather, whether or not the data meet these restrictions can be modeled and tested, by modifying the model.

For instance, in models (16) and (17) it is assumed that the data show compound symmetry and homoscedasticity. These assumptions can be tested by allowing the residuals to vary across conditions. This can be achieved by indicating that the explanatory variables should be incorporated in the random part of the model. In the next model, the assumption of homoscedasticity is tested for the level 1 residuals. Compound symmetry does not apply for this level, since each specific measurement is only performed under one of the treatment conditions, thus there is no covariance between the conditions at level 1. To test the assumption of homoscedasticity, we will allow the residuals at level 1 to vary across conditions, as in model (18).

\[
\begin{align*}
\gamma_{i,j,k} & = \beta_0 x_{i,j,k} + \beta_1 x_{i,j,k} + \beta_2 x_{i,j,k} + u_{i,j,k} + v_{i,j,k} + e_{i,j,k} + e_{i,j,k} + e_{i,j,k} \\
& (l=1,2,\ldots, I; j=1,2,\ldots, J; k=1,2,\ldots, K)
\end{align*}
\]
As we can see, model (18) differs from model (17) in the presence of condition specific residuals \(e_{1ij(k)}, e_{2ij(k)}\), and \(e_{3ij(k)}\). The parameter estimates are presented in Table 2.11.

<table>
<thead>
<tr>
<th>Parameter estimates of model (18)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed</strong></td>
</tr>
<tr>
<td>(\beta_1)</td>
</tr>
<tr>
<td>(\beta_2)</td>
</tr>
<tr>
<td>(\beta_3)</td>
</tr>
<tr>
<td><strong>Random</strong></td>
</tr>
<tr>
<td>Level 2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Level 1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>-2*loglikelihood</td>
</tr>
</tbody>
</table>

If the assumption of homoscedasticity is not valid, model (18) will have a better goodness-of-fit than model (17). The difference in -2*ll equals 1.05, while the difference in number of parameters is 2. The difference between the models is not significant \(\chi^2(2) = 1.05, p = .59\). It may therefore be concluded that the assumption of homoscedasticity is valid for the level 1 variance, although the level 1 variance in condition 3 is somewhat larger than in the other two conditions (0.658 as opposed to 0.551 and 0.515). But what if the assumption proved to be invalid? In that case, the new model would have provided a better fit to the data and would have been selected for further analysis. Any observed treatment effect would then have been corrected for heteroscedasticity.

A third advantage of MLM is that residuals at the various levels can be investigated. Especially inspection of the residuals at the text level can be a very powerful tool in linguistic research. In the next figure the standardized residuals for the 15 texts in model (17) are presented.
Figure 2.1 Text level residuals of the simulated dataset

Note that the first 5 texts were presented in condition 1, texts 6 through 10 in condition 2, and texts 11 through 15 in condition 3. Since the effect of condition was part of the fixed part of the model, the treatment factor does not explain differences between these residuals. This means that text 8, a text in condition 2, scores higher than all the other texts, for reasons other than the treatment condition. Similarly, texts 2 and 12 score much lower than the other texts. Apparently, there are unexplained factors in the data that cause this behavior. The search for an explanation for this state of affairs might lead to the formulation of hypotheses regarding text characteristics that further contribute to differences in cognitive processing. Testing these hypotheses in subsequent research will ultimately lead to a deeper understanding of the relationship between text characteristics and cognitive processing.

2.6 Conclusion

The main objective of this chapter was to show how MLM can be used in the analysis of designs with the random factors participant and item. Accounting for the random variability between participants and items in the statistical analysis of experimental data means that the obtained estimates of the treatment effect can be generalized to both a population of possible participants and a population of items.

The largest part of the chapter was devoted to illustrating that using ANOVA to analyze designs with multiple participants and items gives rise to practical complexities. Each design requires a unique ANOVA solution. All of the relevant Mean Squares in the design will have to be identified, several ANOVA runs will have to be executed to obtain these Mean Squares,
and the observed Mean Squares will have to be combined for the calculation of a quasi F and the accompanying degrees of freedom. Furthermore, RM ANOVA requires a full design. Designs with empty cells (like the Latin Square design), cannot be analyzed with ANOVA, unless auxiliary factors as Lists or Groups are used in the analysis. These factors require averaging over items, however, and this will lead to the loss of useful information.

With MLM, the analysis is much more user-friendly. The same basic model applies to more elaborate designs and empty cells in the design (i.e. missing values, either random or by design) pose no problem for the analysis (Goldstein & McDonald, 1988). The random effect of participant and item can be modeled in one single analysis, there is no need to compute anything by hand, and the analysis is rather straightforward: successive models are compared, and the most economical model is selected for significance testing. Analyzing experimental data with MLM is a matter of finding the best fitting model and then testing the difference between the conditions means.

Besides the practical complexities associated with ANOVA the researcher will have to make statistical assumptions. Notably, the assumption that variances and covariances are equal across conditions. These assumptions are often violated, and violating the assumptions may result in biased F-ratios. In MLM, these assumptions can be tested and if the assumptions are violated a model can be used that accommodates these violations.

A final advantage of MLM is that residual variance and covariances can be further investigated. The residuals at all levels in the model can be inspected and this may lead to the formulation of new hypothesis regarding the effect. This increased possibility for interpretation makes MLM a very powerful research tool.

In conclusion, MLM promises to be an adequate alternative to ANOVA, since it solves many of the problems associated with ANOVA. With the use of MLM, the language-as-fixed-effect fallacy will no longer be present in experimental research. This means that theories can be developed in which statements about the relation between text characteristics and cognitive processing are grounded on statistical evidence. The conclusions based upon an experiment are no longer restricted to the specific sample of items that was used, but may be extended to the population of possible items. It is for this reason that the experimental data presented in the subsequent chapters are analyzed with MLM.
Chapter 3: The Role of Text Connecting and Extra-textual Inferences in the Processing of Causal Coherence Relations

In chapter 1, the interpretation of both implicit (without connective or signaling phrase) and explicit causal coherence relations (with connective or signaling phrase) were discussed. In this chapter, a reading time experiment is described that was conducted to further explore the influence of causal connectives on the processing of causal relations. Before the results of this experiment are presented, a short summary is given of the issues that inspired this research. This background is described more extensively in chapter 1.

3.1 The processing of implicit and explicit causal relations

In chapter 1, a cognitive processing account (CPA) was proposed in which it is assumed that the interpretation of an implicit causal relation involves the reader making two knowledge based inferences. The first is a text connecting inference (TCI) on the basis of which the reader establishes the causal connection (Graesser et al., 1995). The second inference is an extra textual inference (ETI), which involves the encoding of knowledge that is needed for establishing the causal connection between the sentences. This knowledge underlies the causal connection between the sentences. With respect to explicit causal relations, marked with a connective, it was argued that readers only make the ETI. There is no need to make the TCI, because the connective specifies the type of connection between the sentences. It was proposed that in the explicit case, the connective functions as a processing instructor: it signals the reader to encode the type of relation specified by the connective by means of a text connecting operation (TCO). Thus, there is a qualitative difference in the process involved in the interpretation of explicit and implicit coherence relations. More specifically, establishing a causal link (integration) involves a TCI in the implicit case, and a TCO when the relation is made explicit with a connective. The ETI is expected to be made irrespective of the presence of the connective. It is important to note that the CPA only applies when the causal relation between the sentences is strong. If the relation is weak, the reader needs additional elaborative processing to establish causal coherence (Van den Broek, 1994).

The processing evidence that supports the CPA comes from two sources. The first source is the observation that the connective because results in a speeding-up and a slowing-down effect of the clause it introduces (Cozijn, 2000; Millis & Just, 1994). The speeding-up effect occurs in the first and middle parts of the sentence, and the slowing down effect in the last part of the sentence. Noordman and Vonk (1997, 1998; see also Cozijn, 2000) have argued that these effects should be contributed to two different processes underlying the interpretation of causal relations.

The first effect of the connective is a reflection of an underlying integration process. Integration means that the two clauses are integrated in a common representation. This common representation is the causal coherence relation that connects the clauses. The
integration effect of the connective shows that integrating the sentences is easier to accomplish with a connective than without a connective. This integration effect can also be expressed in the terminology used in chapter 1: in the presence of a connective, integration of the sentences does not require a TCI. This connecting inference is only needed when the relation remains implicit; in the explicit case a TCO is performed. As inferences take time, the reading times are shorter when a connective is used to express the kind of connection between the sentences.

The second effect of the connective reflects an inference process that occurs at the end of the sentence. During this inference process the reader activates and encodes the knowledge underlying the causal relation (Cozijn, 2000; Noordman, & Vonk, 1997, 1998). As was explained in chapter 1, this means the reader constructs an extra textual inference. The slowing down effect of because may therefore be characterized as an inference effect (Noordman & Vonk, 1997, 1998). In short, the speeding up and slowing down effects of the connective because reveal two underlying processes: an integration process in which the causal connection is established and an inference process in which extra textual information is activated and encoded in the meaning representation of the text.

It was argued in chapter 1 that the slowing down effect of because should not be interpreted as an indication that the extra textual inference is only made when because is present. In essence, the reasoning was as follows: when the connective is not present, readers rely on the extra textual information to be able to construct the causal link. Together with the causal link, the extra textual information is also encoded in the representation of the text (Halldorson & Singer, 2002). Under the assumption that integration precedes the encoding of the inference (Cozijn, 2000; Noordman, & Vonk, 1997, 1998), the slowing down effect of because can be related to the fact that making the extra textual inference takes more time, since readers of explicit causal relations do not benefit from the earlier activation of the extra textual information that occurs in the implicit condition. The following example should serve to illustrate this point.

(1) On his way to work that morning he experienced a big delay. There was a traffic jam on the highway. (example from Cozijn (2000)).

During the interpretation of (1) the integration by means of a text connecting inference should occur at was a traffic jam (cf. Cozijn, 2000). This text connecting inference results in the encoding of the causal connection between the sentences. Establishing the causal link, however, requires that the reader knows that traffic jams may cause delays. The combination of the explicit text and this extra textual information leads to the construction of the causal link. In the last part of the sentence, on the highway, the extra textual information is encoded in the representation. In other words, during the integration phase the underlying knowledge is activated and during the inference phase the information is encoded.

Let us compare example (1) to its explicit counterpart (2).
On his way to work that morning he experienced a big delay, because there was a traffic jam on the highway.

In examples (1) and (2), the causal link is probably established during the processing of *was a traffic jam*. However, in example (2), the integration of the sentences does not require an inference, because the connective specifies the kind of connection between the sentences. Readers make the extra textual inference during reading of *on the highway*. Thus, readers activate *traffic jams may cause delays* and encode this information in the representation. Now note the difference between (1) and (2). Integration is faster in (2) than in (1), because a text connecting inference is needed in (1) but not in (2). In addition, the extra textual information *traffic jams may cause delays* is activated during the integration phase in (1), but not in (2). After the integration phase the extra textual inference is made. In both (1) and (2) this requires encoding the extra textual information, but, contrary to (2), in (1) this does not require the activation of this information at this point because it was activated already. Consequently, the extra textual inference takes more time to complete when the connective is present. In short, the speeding-up effect of *because* can be explained by faster integration of the sentences, and the slowing-down effect of *because* by slower extra textual inferencing.

The proposal outlined in chapter 1 was not restricted to the causal processes involved in the interpretation of causal relations. Clear hypotheses were formulated with respect to the levels of representation each of these processes result in. In line with Noordman and Vonk (1997, 1998) and Cozijn (2000), it is assumed that the integration process is involved in the construction of the text base representation and that the extra textual information is encoded as an inference at the level of the situation model (Halldorson & Singer, 2002).

This concludes a short review of the hypotheses put forward in Chapter 1. Let us now focus on the current chapter. In this chapter, the results of an experiment are described that served the following two purposes. First, the experiment was designed to replicate the processing results obtained by Cozijn (2000) and Millis and Just (1994). There are several reasons why replication of the results is interesting. First of all, the processing and representation hypotheses put forward in Chapter 1 rely heavily on the speeding up and slowing down effects of *because*. Second, the experimental texts used in previous studies were either narrative sentence pairs (Millis & Just, 1994) or short narrative texts (Cozijn, 2000). In the current study the effect of connectives is investigated by using naturalistic non-narrative text (news paper articles). In sum, the experiment was designed to replicate the speeding-up and slowing-down effects of *because* and to investigate whether these results also hold for non-narrative text.

Before the method and results of the experiments are described, some of the methodological issues of previous studies will be discussed. The focus will be primarily on the studies conducted by Cozijn (2000) and Millis and Just (1994). Thus, the main purpose of the next section is a critical review of earlier investigations showing the speeding-up and slowing-down effects of *because*. 
3.2 Processing evidence for the integration and inference effect of connectives

The processing evidence supporting the integration and inference functions of because is rather sparse. In only two studies where the presence of connectives was manipulated the speeding up and slowing down effects have actually been observed (Cozijn, 2000; Millis & Just, 1994). More recent studies by Maury and Teisserenc (2005) and Kamalski (2007) were not able to reproduce both effects.

Let us first consider the experimental method used by Millis and Just (1994). In their study, the participants read sentence pairs like the following:

(3) The elderly parents toasted their only daughter at the party. Jill had finally passed the exams at the prestigious university.
(4) The elderly parents toasted their only daughter at the party, because Jill had finally passed the exams at the prestigious university.

In a self-paced reading experiment the authors collected reading times word-by-word. The word reading times of the second sentence showed an interaction effect of the presence of the connective and location of the word in the sentence. The words following the connective were read faster, but the last word of the sentence was read slower in the presence of the connective.

It is important to note that the speeding-up effect of the connective on the first word after the connective sentence may be due to the fact that in the implicit version this word is the first word of a new sentence. First words in new sentence are known to be processed longer than first words in subordinate clauses (Gernsbacher, 1990). The effect of the connective on the reading times on the first word is therefore likely to be confounded with sentence initial processing (Cozijn, 2000). The interaction pattern, however, was also found by Millis and Just (1994) when the reading times on the first word were excluded from the analysis, which indicates that the speeding up and slowing down pattern was not exclusively the result of the faster reading times on the first word after the connective.

Interestingly, Millis and Just (1994) interpreted the slowing-down effect they observed as an integration effect but not the speeding-up effect they found. They explain the results in terms of their delayed integration hypothesis. According to this hypothesis, when because is present the integration of the sentences takes place at the end of the sentence. This involves the reactivation of the information in the first clause. Presumably, this information has decreased in activation in the presence of because, since the information only needs to remain activated when the reader has to infer the causal connection, as is the case in the implicit version. Thus, when the connective is not present, the information in the first clause has to remain active in working memory, as readers are using this information to infer the link between the sentences. The slower reading times at the end of the sentence in the because condition are therefore compatible with the idea that readers integrate the sentences at that point and will have to reactivated the information in the first clause. The speeding up effect of the connective is
consistent with the idea that readers in the implicit condition are constantly trying to establish the type of connection between the sentences. If the delayed integration hypothesis is true, we may expect that differences between explicit relations in terms of the nature of the relation will show up at the end of the sentence and not earlier. This prediction was tested by Traxler, Bybee, and Pickering (1997). In their experiment subjects read sentences like the following:

(5) Heidi felt very proud and happy, because she won first prize at the art show.
(6) Heidi could create and imagine things, because she won first prize at the art show.

The relation between the two clauses in (5) can be described as a causal relation of the type consequence-cause and in (6) as a relation of the type claim-argument, where the second clause serves as an argument for the claim made in the first clause (Sanders, Spooren, & Noordman, 1992). Differences between the sentences on eye tracking measures appeared in the middle part of the second clause (won first prize). Similar results were obtained in a study by Traxler, Sanford, Aked, and Moxey (1997), who used identical materials. The results clearly contradict Millis and Just’s (1994) delayed integration hypothesis. Readers integrate the sentences in the middle part of the sentence, or in other words, the type of connection between the sentences is established in the middle part of the sentence, and not at the end of the sentence. For this reason, and contrary to Millis and Just’s interpretation, their results are in line with the idea that the speeding-up effect of the connective can be interpreted as an integration effect.

Even though these results seem to corroborate the CPA, a potential problem with studies such as Millis and Just (1994) and Traxler et al. (1997) is that they have used sentence pairs in their experiments. These sentence pairs were presented without context. The question is whether processing results obtained with isolated sentence-pairs have much to say about actual text processing. It can, for instance, be argued that full comprehension of the events in these sentence pairs, requires more information than can be provided by these sentences alone. When a full text is presented, the context will normally allow the reader to construct a coherent situation model of the events. With reference to (5), for example, the context can provide information about who Heidi actually is and about the general circumstances in which the events took place. For this reason, processing effects that are obtained with sentence pairs may not generalize to the processing of entire texts. It seems important for text comprehension research to focus on the processing of complete texts, which allow the reader to fully comprehend the events portrayed in the causally related sentences.

In contrast to Millis and Just (1994), Cozijn (2000) did indeed embed the sentences in short narrative texts. In his experiments participants read short texts each containing one target sentence that was connected to the previous sentence by means of a causal relation. The second clause of this relation was either introduced by the connective because as in (7a) or the relation was not marked as in (7b).

(7a) He experienced a big delay, because there was a traffic jam on the highway.
(7b) He experienced a big delay. There was a traffic jam on the highway.
Cozijn (2000) used both a self paced reading paradigm in which the texts were presented in sections (as in (7c)) and eye tracking, after which the reading times of the same sections as in the self paced reading experiment were analyzed. Cozijn found a facilitative effect of the presence of the connective in the beginning of the second clause (regions 3 and 4), but found longer reading times at the end of the second clause (region 5). The interaction between presence of the connective and region (4 and 5) reached significance.

As was explained in chapter 1, these results can be explained in terms of hypotheses first put forward by Noordman and Vonk (1997;1998), who argue that the connective because has three functions in the reading process. The first function is the segmentation function: because signals to the reader how the linguistic material is structured. For instance, because signals that a new clause begins and that it is a subordinate clause. The speeding up effect on region 3 can be interpreted as a segmentation effect. The second function is the integration function: the connective signals to the reader that the incoming clause is causally related to the first clause. In other words, it signals that the incoming clause should be connected to the first clause by means of a causal relation. The speeding up effect on region 4 reflects the integration function of because. A third function is referred to as the inference function: because is an instruction to the reader to infer the extra textual information that underlies the causal connection signaled by because. The slowing down effect on region 5 reflects the inference function of because. As was argued above, these interpretations of the results also apply to the processing pattern obtained by Millis and Just (1994).

Given these results and the fact that they fit in nicely with the CPA presented in chapter 1, the question rises why we replicated Cozijn’s (2000) study. Basically, for two reasons: the experimental materials that were used and other possibly confounding factors which co-varied with the experimental manipulation. Both issues are discussed in some detail below.

The first reason for replication has to do with the materials used by Cozijn (2000). Although the target sentences were embedded in a text, the texts are still fabricated narratives that would not be likely to occur in the real word, outside the lab. The risk of using this kind of experimental texts is that results obtained with them do not generalize to ‘real texts’ (see for a similar point of view: Graesser et al. 1997). It is important to use naturalistic texts, if we want to be able to draw ecologically valid conclusions about processing mechanisms (cf. Sanders, Land, & Mulder, 2007).

A second issue concerns the experimental manipulation: Cozijn (2000) compared identical sentences that either did (explicit) or did not (implicit) contain the causal connective because. The experiments were in Dutch and the explicit version contained the subordinate connective omdat. As a result, the two experimental conditions did not only differ in the presence of the connective, but also in the syntactic status: in the implicit version there was simple coordination of two clauses, in the explicit version the second clause was subordinate to the main clause. This also implies a different word order for the coordinated clause in the coordinated versus the subordinated clauses in the explicit version, see (8a) and (8b) below (adapted from Cozijn, 2000):
(8a) Hij ondervond een flinke vertraging. Er was een file ontstaan op de snelweg.
(8b) Hij ondervond een flinke vertraging, omdat er een file was ontstaan op de snelweg.

The italicized sections of (8a) and (8b) illustrate the difference in word order. The italicized part of the sentence corresponds to region 4 (see example 7 above), where the speeding up effect of *because* was detected. Thus, the experimental manipulation (implicit versus explicit) is confounded with a word-order difference, and therefore, the speeding-up effect may reflect an effect of the difference in word-order. In the current study, we set out to control for this possible confounding by using the coordinating causal connective *want*, in addition to *omdat*.

One of the main objectives of the present study is to investigate the processing of causal relations in non-narrative text. Two recent studies by Maury and Teisserenc (2005) and Kamalski (2007) have investigated the role of connectives and comparable signaling devices in respectively expository texts and persuasive texts. Unfortunately, both studies suffer from problems, as a result of which we can draw no valid conclusion on the possibility that the speeding-up and slowing down effects can be generalized to non-narrative text. These problems have to do with the quality of the materials and the statistical analyses of the results. Maury and Teisserenc (2005) presented short expository texts. All of the texts consisted of 4 sentences. Consider the following text:

(9) Plastic is made of artificial resins. The plastic is supplied in tablets of different diameters. The plastic spreads over the mould, because the chemist heats the plastic to a high temperature.

The question is whether results obtained with experimental materials like these have much to say about discourse processing. It may be hard for readers to construct a situation model on the basis of these four sentences without a preceding context. For instance, a context in which it is made clear that the production process of plastic will be described. Without this additional information it is very hard to figure out what the coherence relations are between the sentences in the text. For example, what is the coherence relation between the second and the third sentences in (9)? If readers are not able to determine the coherence relations between the sentences, the only source of coherence in (9) is referential coherence which results from repeated reference to plastic. For this reason, it can be argued that the text lacks a fundamental property of texts: relational coherence. As a result, the results obtained with texts like these are not only very difficult to interpret, but also hardly informative as far as text comprehension is concerned. In contrast, the materials in the Kamalski (2007) study do not suffer from this lack of quality. Therefore, the results obtained in her experiment are potentially interesting for the present purposes. In her experiment, participants read naturalistic short persuasive texts such as the following:

7 For the interested reader, the uninterpretable results showed a speeding up effect of *because* on the first and middle part of the sentence but not a slowing down effect at the end of the sentence. Thus, the results shows a speeding up effect in line with the segmentation and integration function of *because*, but not a slowing down effect as predicted by the inference function of the connective.
Chapter 3

The state Board of Higher Education is considering a standardized test for college students. Such a test will assess students’ writing ability, reasoning and computer skills. Consequently, the Board can ensure that students don’t graduate without possessing basic skills. Although some state officials worried that such a test would cause professors to gear their classes toward the test, this will ultimately have a positive effect on the quality of the state college system. On that account, an exit exam for college students needs to be implemented as soon as possible. All students will benefit from such a system.

Kamalski manipulated the presence of connectives and lexical signaling phrases for both objective and subjective causal relations (Pit, 2003, 2005; Stukker, 2005; compare Pander Maat & Sanders, 2001). Examples (5) and (6) above are examples of objective and subjective relations respectively. An example of an objective relation manipulated by Kamalski is the following:

Such a test will assess students’ writing ability, reasoning and computer skills. Consequently, the Board can ensure that students don’t graduate without possessing basic skills.

This sentence pair was presented either with or without marking. In (11) the underlying objective causal relation is marked with the connective Consequently. Kamalski (2007) analyzed the reading times of the two segments of the second sentence and the reading times of the whole sentence. The first segment corresponds to The Board can ensure and the second segment with possessing basic skills.

For relation types such as in (11), only the total sentence reading times are presented, but the reading times for the different segments are not presented. The obtained results indicate that the presence of the connective did not have a significant effect on the processing of the sentence in its entirety. However, Kamalski (2007) claims she in fact did find an effect, but her conclusion is based on a misinterpretation of the statistical test. For example, Kamalski incorrectly applies the F1 x F2-criterion (see chapter 2). For instance, when one of the two F-ratios fails to reach significance, she nevertheless concludes the effect is significant.

Another more serious problem has to do with the textual materials used by Kamalski (2007). It is questionable whether the materials are suitable for detecting faster integration and slower extra textual inferencing. Compare, for instance, (11), repeated below in (12), with its implicit version (13).

Such a test will assess students’ writing ability, reasoning and computer skills. Consequently, the Board can ensure that students don’t graduate without possessing basic skills.

Faster integration means that the detection of the causal relation proceeds quicker in the explicit than in the implicit condition. With respect to the integration effect, we should not
focus on the first segment of the sentence. As was explained above, the reason for this is that the effect of marking on the first segment of the new clause is confounded with the effect of sentence initial processing (Cozijn, 2000; Gernsbacher, 1990). This is exactly the reason why Cozijn, for instance, did not interpret the effect found in this segment as an integration effect. On the contrary, the shorter reading times on this first segment are interpreted as a segmentation effect (Noordman & Vonk, 1997, 1998) of the connective. The crucial region where an integration effect can be detected is that point in the sentence where readers of the implicit relations have enough information to construct the causal link. In the case of (13), the readers of the implicit relation have enough information to establish the causal link at the end of the sentence, thus on Kamalski’s (2007) second segment (possessing basic skills). Note that in (13), the first segment (The board can ensure) does not provide enough information to establish the causal connection, which is an additional argument why we should not focus on the first segment to detect an integration effect. It can be concluded that the reading times Kamalski collected at the end of the sentence actually reflect a mixture of integration and extra textual inference processes. Therefore, it is not clear what kind of effect the longer reading times in the explicit condition represent. As a matter of fact, it can be maintained that the absence of effects in the objective condition can be attributed to this mixture of processes too.

On the whole, the review of the processing evidence in this section leads us to conclude that the empirical basis for CPA of the interpretation of causal relations proposed in chapter 1 is not firmly established. The available evidence is either based on the processing of sentence pairs (Millis & Just, 1994) or on unnatural narrative text, where the speeding up effect of the connective might be the result of a difference in word-order (Cozijn, 2000). With respect to the other studies that provide more detailed views on the processing of clauses and sentences, the conclusion is that they either used ill-formed text (Maury & Teisserenc, 2005) or target sentences that do not permit the separation of integration and inference effects of linguistic marking (Kamalski, 2007).

This state of affairs clearly asks for an experimental study in which the strong aspects of previous studies are incorporated, and therefore an attempt is made to eliminate the problems associated with each of these individual studies. The experiment reported below aims at doing just that. The extent to which this was indeed successful will be evaluated in 3.4. In the experiment, naturalistic texts are used (as in Kamalski, 2007) with target sentences that allow for the separation of integration and extra textual inference effects (as in Cozijn (2000) and Millis and Just (1994)).

### 3.3 The experiment

In order to test the influence of the presence of connectives on reading times a self-paced reading time experiment was conducted. For practical reasons, the experiment had to be run in two parts. The two parts were identical except for the filler texts that were used.
Chapter 3

The subjects read one of three possible versions of a causal relation. In the explicit versions, the relation was introduced by either the causal connectives want or omdat. In the implicit version, the relation was not marked with a connective. Besides the effect of the presence of the connective on reading times, the effect on the reaction times on verification statements was investigated.

There are three major differences between this study and the experiments by Cozijn (2000). The first difference is that expository texts instead of narrative texts were used. The second difference is that these texts were based on real news articles that were found in national and local news papers. As was explained in previous chapters, one of the aims of the present study is to obtain evidence that the results found for narrative text can be generalized to expository text. The third difference is that besides omdat the connective want was used. The reason for this is that want, unlike omdat, which was used in the Cozijn study, does not change the syntactic structure of the sentence it introduces. In line with Cozijn (2000) and Millis and Just (1994), and unlike the study reported by Kamalski (2007), the causal relations allowed the separation of integration and extra textual inference effects.

In line with the CPA outlined in chapter 1, it is hypothesized that the presence of connectives leads to a processing advantage during the integration phase and a processing disadvantage during the inference phase. The reason for this is that during the integration phase, the participants will have to make a text connecting inference in the implicit versions in order to integrate the sentences. This will lead to an increase in reading times. At the same time, readers can benefit from the activation of the extra textual information that is needed for the integration of the sentences, as a result of which processing times during the inference phase will be shorter in the implicit condition than in the explicit condition. The integration effect of the connectives is hypothesized to occur in the middle part of the second sentence; and the extra textual inference effect will occur at the end of the second sentence.

With respect to the verification times, no specific hypotheses are formulated. As was explained in chapter 1, the statement verification task is an off-line measure that does not only reflect the activation of the extra textual information during, but also retrieval from the text representation. Since it is hypothesized that the extra textual inference is made regardless of the presence of the connective, no more than a trivial reaction time difference is expected. However, because the presence of connectives may have an effect on the representation, a reaction time advantage of the connective conditions may still be observed.

3.3.1 Method

Participants

112 subjects from the Faculty of Arts of Utrecht University took part in the experiment and were paid for their participation. The average age of the participants was 21 years. 65 percent of the participants was female and 35 percent male.
The Role of Text Connecting and Extra-textual Inferences

Materials

The participants read 12 short expository texts that contained 1 causal relation, of which the second part served as the target sentence. Besides the 12 experimental texts, the participants read 12 filler texts. Both the experimental texts and the filler texts were selected from a wide variety of sources. The texts were based on actual texts found in local and nation-wide newspapers, the Internet, and the Dutch Teletext service.

The texts were presented in sections which correspond to major constituents, as shown in example (15). This example was taken from the text presented below (14). The text in (14a) is the Dutch original. Its English translation is presented in (14b).

(14a) Een Nederlandse tanker geladen met salpeterzuur is nabij het Duitse Krefeld lekgeslagen. Bij een aanvaring met een Duits vrachtschip ontstond een scheur in de romp. Vrijwel direct erna ontstond een felle brand. Omwonenden moesten ramen en deuren sluiten, omdat er giftig gas vrijkwam bij het ongeval. Het vuur was na enkele uren onder controle. Er zijn geen slachtoffers gevallen.

(14b) A Dutch tanker loaded with nitric acid has struck a leak near German Krefeld. During a collision with a German freighter the hull got damaged. Almost immediately a fierce fire broke out. Local residents had to close windows and doors, because poisonous gas escaped during the accident. The fire was under control after a few hours. There were no casualties.

(15) / Omwonenden / moesten / ramen en deuren sluiten/, omdat / er / giftig gas vrijkwam / bij het ongeval. / Het vuur /

Local residents / had to / close windows and doors/, because / (there) / poisonous gas escaped / during the accident. / The fire /

The target sentences were divided into 4 target sections. The sections of interest are the first section following the connective (target 1), the middle part of the sentence (target 2), and the last part of the sentence (target 3) and the first part of the sentence following the causal relation.

Local residents / had to / close windows and doors, / because / (there) (1) / poisonous gas was released (2) / at the accident. (3) / The fire (4) /

The statements used in the statement verification task contained the information that was hypothesized to be activated during the inference-stage. In the case of example (15) the statement was:
“If poisonous gas escapes in the vicinity, windows and doors have to be closed.”

Design and Procedure

Three experimental lists that were constructed using a Latin Square Confounded Factorial design (Kirk, 1982). The experimental procedure was as follows. The texts were presented at a computer screen using a self-paced moving-window paradigm. Each time the participant pressed the space bar the next constituent/segment appeared. The reading time was defined as the time between the text appearing on the screen and the space bar being pressed. The presentation of the text and the measurement procedures were controlled by the program E-prime.

After reading a text, the participant had to verify three statements. For each statement the participant had to decide whether the statement was true or not according to the information provided in the text. The participant had to press the ‘f’ button on the keyboard when the statement was false, and the ‘j’ button if the statement was correct. The ‘f’ button was marked with a red color and the ‘j’ button was green. All the target statements were correct. The statements for the filler texts were incorrect. As a result, half of the statements were correct and half of them were incorrect.

Statistical Analysis

As was explained, the experiment took place in two sessions. The experimental texts were identical but the filler texts differed between the parts. Reading times of the two parts of the experiment were entered in the analysis simultaneously. This allowed us to check for the unlikely event that the effect of the connectives interacted with the part of the experiment. The reading times on the targets 1, 2, 3, and 4 were entered into a multilevel analysis (see chapter 2). Although target segment 4 is not of particular interest, it was included in the analysis to assess the possibility of spill-over effects that may have occurred as a consequence of the self-paced-moving window paradigm (Cozijn, 2000). For each participant, the reading times of a particular text were included if the corresponding target statement was responded to correctly. This restriction resulted in the deletion of 17% of the cases. A total number of 1148 cases were included in the analysis.

The reading times and the reaction times on the statement verification task were checked for outliers in a multivariate outlier analysis (see Bollen, 1989, p. 28 for definitions and computational details). The analysis resulted in the detection of one outlier for targets 3 and 4, two outliers for target 1, three outliers for target 2, and five outliers on the statement verification task.

The reading time data were analyzed in a multivariate multilevel model that accounted for variance of and covariance between the targets at level 2 (texts and participants) and level 1 (error). The effect of the connectives on the statement verification times was assessed in a univariate multilevel model.
3.3.2 Results

The reading time data were analyzed with a multivariate multilevel model (see chapter 2). Two models were compared: a model in which only the intercepts of the respective targets was estimated in the fixed part was compared to a model that accounts for the effects of the manipulation. Both models included level 2 variance and covariance estimate, both between texts and between participants and level 1 error variance and covariance. The -2LL-chisquare difference between the empty model and the full model was 80.0 with 8 degrees of freedom. This indicates that the presence of the connectives had a multivariate effect on the reading times ($\chi^2(8) = 80.0, p < .001$). In Table 3.1 the results of the fixed part of the multivariate model are presented.

| Table 3.1 Fixed parameter estimates of the multivariate regression analysis of the reading times (scores are milliseconds, s.e. in parentheses) |
|---|---|---|---|---|
| | Target 1 | Target 2 | Target 3 | Target 4 |
| Implicit | 549.4 (21.8) | 854.1 (49.7) | 749.5 (42.4) | 686.1 (49.4) |
| Implicit versus want | -73.8 (11.6) | -26.4 (24.4) | 45.9 (22.4) | 26.1 (18.0) |
| Implicit versus omdat | -90.2 (11.5) | -15.7 (24.3) | 21.3 (22.3) | 47.9 (17.9) |

The results suggest that several effects of the presence of the connectives have been detected. For instance, the reading times in target 1, which is the first part of the second sentence, indicate a processing advantage of the connectives as opposed to the implicit condition. In target 2, where the integration effect of the connectives was expected, on the other hand, the reading times do not seem to differ that much between the implicit and explicit conditions. The reading times in target 3, where the slowing down effect of the connective was expected, do seem to be longer in the connective conditions than in the implicit condition. In addition, in target 4, the first segment of the sentence following the causal relation, the reading times are longer in the explicit conditions than in the implicit condition.

With respect to the reading times in target 1, the estimated population difference between want and implicit was -73.8 milliseconds (s.e. = 11.6; $\chi^2(1) = 40.7, p < .001$), indicating that the reading times are indeed shorter in the want condition than in the implicit condition. The estimated difference between omdat and the implicit condition was -90.2 ms (s.e. = 11.5; $\chi^2(1) = 61.4, p < .001$), which indicates that the reading times were shorter in the omdat condition.

---

8 The reading times and the reaction times on the statement verification task were checked for influential cases and non-normality of level 1 residuals. Outliers were determined in a multivariate outlier analysis (see: Bollen, 1989, p. 28 for definitions and computational details). The analysis resulted in the detection of one outlier for targets 5 and 6, two outliers for target 3, three outliers for target 4, and five outliers on the statement verification task. The distribution of the level 1 residuals shows a large negatively skewed distribution. The data were therefore reanalyzed with normalized reading and reaction times and without the outliers. The reanalysis did not lead to a different pattern of results than the one described in this subsection.
than in the implicit condition. No effect was detected in the reading times of target 2, which is the segment were a speeding-up effect of the connective was expected. The estimated population difference between want and implicit is -26.4 ms (s.e. = 24.4, $\chi^2(1) = 1.73$, p = .19) and the estimated difference between omdat and implicit is -15.7 ms (s.e. = 24.3; $\chi^2(1) = .42$, p=.52). The reading time in target 3, where a slowing down effect was expected, is an estimated 45.9 ms (s.e = 22.4) longer in the want than in the implicit condition ($\chi^2(1) = 4.212$, p <.05). The estimated difference between omdat and implicit of 21.3 ms (s.e.=22.3) does not reach significance ($\chi^2(1) = .92$, p = .34). In target 4, the difference between want and implicit of 26.1 ms (s.e = 18.0) is not significant ($\chi^2(1) = 2.1$, p =.14), whereas the difference between omdat and implicit of 47.9 ms (s.e. = 17.9) is ($\chi^2(1) = 7.16$, p <.01). None of the targets show a significant difference between the want and omdat conditions.

Analysis of the verification data shows that the reaction times did not differ between conditions ($\chi^2(2) = 4.2$, p = .12). The mean reaction time in the implicit condition is 3259 ms (s.e. = 99.7), in the want condition the mean equals 3123 ms (s.e. = 100.1), and in the omdat condition the mean reaction times are 3112 ms (s.e. = 99.8).

Data exploration

The results show that the presence of the connectives did not significantly speed up the processing of target 2, while this was one of the crucial expectations. To further investigate these unexpected results the processing times on target 4 were analyzed further in a univariate analysis. The results of this analysis are shown in Table 3.2.

### Table 3.2 Regression parameters in the reading time analysis of target 4 (s.e of parameter estimates between brackets).

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Want</th>
<th>Omdat</th>
<th>Implicit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td>-</td>
<td>830.5 (38.9)</td>
<td>839.8 (42.3)</td>
<td>858.4 (40.4)</td>
</tr>
<tr>
<td>Random</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between items</td>
<td>11999.3</td>
<td>(5567.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Want</td>
<td>23210.4</td>
<td>(7887.0)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>omdat</td>
<td>43455.8</td>
<td>(7947.9)</td>
<td>45902.7</td>
<td>-</td>
</tr>
<tr>
<td>Implicit</td>
<td>32171.3</td>
<td>(6611.3)</td>
<td>42184.9</td>
<td>29685.7</td>
</tr>
<tr>
<td>Error</td>
<td>109560.4</td>
<td>(9471.3)</td>
<td>137330.0</td>
<td>132116.0</td>
</tr>
</tbody>
</table>


The estimated difference between the mean reading times in the want and implicit conditions is -27.9 ms. This difference does not reach significance ($\chi^2(1) = 1.462, p = .22$). The 95% confidence interval of this difference ranges from -73.0 ms to 17.3 ms. The difference between omdat and implicit is -18.6 ms. This difference is not significant ($\chi^2(1) = .554, p = .46$). The 95% confidence interval for the difference between omdat and implicit ranges from -67 ms through 30.3 ms. Thus, both confidence intervals are compatible with a wide range of possible population values, including the 0 difference between conditions. For this reason, the results of the present experiment are inconclusive with respect to the integration effect of causal connectives.

The variance estimates show an interesting pattern. At level 1, the variance in the implicit condition (132116.0) and the omdat condition (137330.0) are somewhat higher than in the want condition (109560.4). With respect to the between participant variance the estimates show that the variance in the omdat condition (45902.7) is much larger than the variance in the want condition (23210.4) or the variance in the implicit condition (29685.7). These results indicate that differences between persons influence the reading times in the omdat condition more than in the other conditions. This may indicate a large variation in the processes involved in the interpretation of the omdat sentences, an interpretation which will be further explored below.

### 3.3.3 Conclusion

It was hypothesized that the presence of the connectives leads to both a speeding-up and a slowing-down effect. In particular, it was expected that the segment reading times in target 2, the middle part of the sentence, is processed faster in the connective conditions than in the no-connective condition. The results indicate that the speeding-up effect of the connectives is not statistically reliable. A speeding-up effect was detected in target 1, the first segment of the second clause, but, for reasons explained above, this effect cannot be interpreted as an integration effect. This issue will be taken up in section 3.4.

The slowing-down effect of the connectives was hypothesized to occur at segment 3, the last segment of the second sentence. A slowing-down effect was indeed detected for the want condition, but not for the omdat condition. As a matter of fact, in the omdat condition, the reading times were longer in target 4, but it is difficult to interpret this effect. It may be a spill-over effect, due to the processing of target 3, but as it is the first segment of the sentence following the causal relation, it may just as well reflect other processes (Cozijn, 2000).

On the verification times no effect of the presence of connectives was detected. Although this is in line with the CPA, the results should only be reluctantly interpreted as such. The reason for this is that the research hypothesis that at most a trivial connective effect on the reaction times occurs is in fact a null-hypothesis. Drawing conclusions about the null-hypothesis based on non-significant results is always suspect, since failure to reject the null-hypothesis may be a lack of statistical power.

In conclusion, no evidence was obtained for the occurrence of an integration effect of the connective, and only in the want condition an extra textual inference effect was detected. In
other words, the results of the experiment do not seem to be compatible with the hypotheses. In the next section, explanations for these rather disappointing results will be discussed, as well as the consequences the results may have for the processing account I have presented.

3.4 General Discussion

According to the CPA presented in chapter 1, the interpretation of implicit causal coherence relations involves the construction of a text connecting inference and an extra textual inference. The text connecting inference is constructed during the integration phase, where the type of connection is established. The extra textual inference is made during the inference phase, where the extra textual information is encoded in the representation of the text. When the relation is explicitly marked with a connective, only the extra textual inference needs to be made. In this case, the integration phase involves a text connecting operation.

On the basis of these hypotheses, two effects of the presence of the connective were expected. The integration process was hypothesized to occur faster when the connective is present. This faster processing results from the fact that the type of connection does not have to be inferred. The inference process was hypothesized to proceed slower when the connective is present, because in contrast to readers of implicit relations, readers of explicit relations do not benefit from the activation of the underlying knowledge during the integration process. The results of the experiments only partly support these hypotheses, because no integration effect was detected, and only one of the connectives (want) shows the extra textual inference effect at the end of the sentence. The slowing-down effect of omdat may have spilled-over to the first section of the sentence following the causal connective, but this result must be interpreted with great care.

The hypotheses were based on earlier processing studies in which the speeding-up and slowing down effects were indeed detected (Cozijn, 2000; Millis & Just, 1994) and on the interpretation of these effects by Noordman and Vonk (1997, 1998). For this reason, then, we have to conclude that the results obtained in this experiment only partly agree with previous research. This state of affairs clearly asks for an explanation. In the discussion below, I will first focus on the integration effect, because the major difference between earlier studies and the present one are particularly evident with respect to the speeding-up effect of the connectives. The discussion will primarily focus on methodological flaws in the present experiment, but also theoretical considerations will be given.

It has been concluded that no integration effect of the connectives was found. However, the presence of want and omdat resulted in a clear speeding-up effect on the segment immediately following the connective. Why is this not an integration effect? It was explained in section 3.2 that the speeding-up effect on the segment immediately after the connective is most likely due to the fact that in the implicit condition this first segment is the first part of a new sentence. As sentence initial processing requires more processing than the processing of first words in subordinated or coordinated clauses (when explicitly marked with a connective), this speeding-up effect of the first segment is likely due to sentence initial processing (Cozijn, 2000; Gernsbacher, 1990). For this reason, this effect is interpreted as a segmentation effect.
The Role of Text Connecting and Extra-textual Inferences

of the connective, which refers to the segmentation function of connectives (Noordman & Vonk, 1997; see section 3.2 and chapter 1), and not as an integration effect. Thus, the results show a segmentation effect of the connective, but not an integration effect.

The first question is, then, why was an integration effect of the connectives not observed? First of all, the statistical test may have lacked statistical power to detect the effect. We should therefore be prudent in concluding that an integration effect does not occur, since we run the risk of making a type-II error (incorrectly not rejecting the null-hypothesis). However, referring to statistical power as to why the effect did not occur is rather unsatisfying, since this entails that the effect probably does exist but was simply not detected. There are methodological and theoretical reasons why the integration of the sentences in the reading of these experimental texts did not necessarily proceed faster when the relation was made explicit.

A methodological problem may have been the materials that were used. Although naturally occurring news articles were used in the experiment, the target causal relations and connectives may have been unsuitable for detecting an integration effect. The reason for this is that the target sentences had to be compatible with both the connectives want and omdat. Want is a typical marker for subjective causal relations, e.g. claim-argument relations, whereas omdat in itself is ambiguous, because it can be used for both subjective and objective causal relations, e.g. claim-argument and consequence-cause relations (Degand & Pander Maat, 2003; Pit, 2003). It is crucial for the detection of the integration effect that the readers in the connective conditions interpret the same relation as readers in the implicit condition, but because of the ambiguity of the connective (omdat), there is a chance that the interpretations differed among the participants. From a statistical point of view, the ambiguity of relation type signaled by omdat may have resulted in increased error variance on the integration segment. Note for instance, that Traxler, Bybee, and Pickering (1997) and Traxler et al. (1997) have shown that the interpretation of subjective causal relations takes more time than the interpretation of objective causal relations. This increased error variance on the integration segment may have resulted in the inability to detect significant differences between the conditions.

As a matter of fact, the ambiguity of the relation type signaled would also explain why an inference effect was detected for the want condition but not for the omdat conditions. The reasoning behind the inference effect is that readers of explicit conditions take more time to process the final segment of the second clause because they have to activate the extra textual information at that point, whereas readers in the implicit condition can take advantage of the fact that this knowledge was already activated during the integration phase. Now, because of the ambiguity of the connective, the extra textual information has to be accessed in both the omdat and the implicit conditions during the integration phase, but not in the want condition. The activation advantage for the implicit condition is therefore only expected to occur in comparison to want and not in comparison to omdat. This is exactly the pattern of results observed in the experiment.

If we take into account the ambiguity of the relation type signaled by omdat, it is possible to predict that the error variance on the integration segment will be larger for the omdat and implicit conditions than for the want condition. This prediction was tested in a multilevel
model of the processing times of the integration segment presented in the results section. Overall, these variance estimates indicate that reading times in the ondat condition varied much more than in the other conditions. Thus, these results are perfectly in line with the assumption that ambiguity of ondat with respect to the relation it signals influences the reading times. However, it should be noted that another difference between ondat and the other conditions is that the clause introduced by ondat is a subordinated clause whereas in the case of the want condition and the implicit the clause is coordinated. This syntactic difference is reflected in a difference in word order between the conditions (see section 3.2). Although it is not clear how a difference in word order may have resulted in an increase in between-participant variance in the ondat condition, this shows that the difference between the conditions may be exclusively attributable to the ambiguity of ondat.

Still, these considerations lead to the conclusion that the materials used in the experiment were probably not suitable for detecting an integration effect. It can thus be concluded that the integration effect has not been properly investigated, and therefore the absence of the detection of the effect cannot be interpreted as evidence that the hypothesis about the integration effect is incorrect. On the contrary, on the basis of this experiment no substantive conclusion can be drawn about the integration effect of the connective.

However, besides these methodological issues, there are also some theoretical reasons why an integration effect may not occur when the materials do in fact enable the testing this effect. Three of these reasons are given here. The first reason addresses text genre, and the other two explain why inferring a causal connection may not require more cognitive resources when the relation is implicit.

A first theoretical reason why an integration effect may not have been observed refers to a difference in the genre of the materials between the current study and previous studies. The studies by Cozijn (2000) and Millis and Just (1994) used short narrative text, and in the present study short news paper articles were used. Such an explanation, however, requires the assumption that fundamental differences exist between the processes involved in the interpretation of causal connections in news paper articles and short narrative texts. Although such a fundamental difference can be expected between simple stories and expository texts describing relatively new information about technical or scientific topics, e.g. the knowledge that is needed to infer causal connections may be absent for texts describing unfamiliar topics (Britton, 1994; Graesser et al., 1995), such a large difference is not expected for narratives and newspaper articles that describe events in the domain of general world knowledge. It is therefore unlikely that the absence of the effects can be attributed to a difference in the genre of the text. On the contrary, the detection of the inference effect may lead one to suspect that causal processing is largely independent of text genre.

A second reason why no difference in the time needed to make connecting inferences was detected is that these inferences may be made very quickly (McKoon & Ratcliff, 1992; Van den Broek, 1994). If the causal relation is strong and the underlying knowledge is easily available, it can be predicted that making the connecting inference will have only little impact on the cognitive resources needed to establish the causal relation. For this reason, it is unsure whether or not a speeding-up effect of the connective should occur during reading at all. This is in line with research reported by Golding, Millis, Hauselt, and Sego (1995), who have
shown that (total) sentence reading time differences between implicit and explicit conditions only arise for moderately related causal relations. The question is, then, whether the causal relations used in the present experiment were also of moderate strength. No independent information is available on the basis of which the causal strength of the relations can be determined, but inspection of the materials hints at the observation that the strength may be relatively low, and therefore no speeding up effect of the connectives was observed.

A third and final reason also makes reference to the assumption that inferring the causal connection in the implicit condition may not be very much more difficult than establishing the relation in the explicit version. It is conceivable that readers – within certain limits – assume that the relation between sentences is causal. This latter hypothesis is the CAUSALITY-BY-DEFAULT hypothesis (Sanders, 2005b; Sanders & Noordman, 2000). According to this hypothesis, readers will, by default, try to connect sentences by means of a causal relation. Presumably, a non-causal relation will only be established when a causal reading of the text is not possible. It may therefore be that establishing the causal link is only slightly slower when the relation is not marked with a connective. The validity of the CAUSALITY-BY-DEFAULT hypothesis will be investigated further in chapter 5.

In sum, there are both methodological and theoretical considerations that explain why an integration effect of the connectives was not detected. An inference effect was observed, however, for the want condition. As was explained above, the explanation of this effect is that readers in the implicit condition benefit from the fact that the extra textual information was already activated during the integration phase, and as result the encoding of the extra textual information is faster in the implicit condition than in the explicit condition. The fact that no such effect was found for the omdat condition (although there may have been a spill-over of the effect in this condition) was explained with reference to the ambiguity of the connective omdat. Because of this ambiguity, readers also had to access the extra textual information in the omdat condition, and therefore the encoding advantage of the implicit condition disappears. The inference effect in the want condition and the absence of the effect in the omdat condition leads to the tentative conclusion that the inference effect may only occur when the connective is unambiguous with respect to the relation it signals.

The fact that a clear slowing-down effect was not observed with omdat, whereas it was for want, means that previous findings by Cozijn (2000) and Millis and Just (1994) were partly replicated. It therefore seems to be plausible to conclude that the extra textual inference effect observed in these studies is similar to the findings in the present study, and that the effect is not a by-product of the unnaturalness of the materials. Moreover, the results indicate that the extra textual inference is made independent of the genre of the text. In this study, the effect occurs in short newspaper articles that describe events in the domain of general world knowledge. Previous studies have shown that the extra textual inference is made in short narrative sequences (Cozijn, 2000; Halldorson, & Singer, 2002; Millis & Just, 1994; Singer et al, 1992) and in expository text (Noordman, Vonk, & Kempff, 1992; Simons, 1993).

The results on the verification task are in line with the idea that the extra textual inference is made independent of the presence of the connective. However, as was argued above, the fact that no significant effect was found on the reaction times should not be taken as evidence for
the null-hypothesis. Note also that the verification task itself is not a valid measure to investigate whether the inference was made on-line because it is an off-line measure. This means that although there should be some correlation between the reaction times and the occurrence of the extra textual inference, the verification times also reflect retrieval phenomena. Therefore, the verification data obtained by Cozijn (2000), Singer, et al. (1992), and Halldorson, & Singer, 2002) should also be interpreted with great care. In the worst of possible circumstances none of the studies (including the present one) have convincingly demonstrated that the reaction times are a function of making the extra textual inference during reading.

In conclusion, the results of the present experiment show that the presence of connectives does not necessarily speed up the integration process. The observed extra textual inference effect was, however, observed for one of the connective conditions. The absence of the integration and inference effect in the omdat condition can possibly be explained by the ambiguity of the connective, as a result of which readers had to access extra textual information in order to establish the causal link. Most importantly, this explanation allows for the hypothesis that interpreting the causal relation requires a text connecting inference that establishes the causal link and the encoding of the extra textual information, although a clear speeding-up and slowing down effect would be more convincing in this respect. That is, the explanation of the results clearly hinges on the assumption that integration and inference processes occur during reading.

In chapter 1, it was argued that the causal link between the sentences is represented at the level of the text base and that the situational level of representation includes not only the causal connection between the events, but also the extra textual information underlying the causal connection. The interpretation of the current results is compatible with the hypothesis that an integration process and an extra textual inference process occurred during reading of the causal relations. Therefore, indirect evidence was obtained for processes involved in the construction of the textbase and the situation model representations of causal relations. We have seen how the connective, as part of the surface code of the text, provides processing instructions with respect to these processes. In the next chapter, it is investigated whether direct evidence can be obtained for a separate text base and situation model representation of explicit causal relations.
Chapter 4: The Representation of Causal Coherence Relations in Expository Text.

4.1 Introduction

In chapter 1 it was explained that a cognitive account of coherence relations claims that coherence relations are part of the mental representation readers construct of a text (Sanders, Spooren, & Noordman, 1992, 1993). However, it was also argued that speaking of “the mental representation” seems to be an oversimplification (Sanders & Noordman, 2000). Ever since Van Dijk and Kintsch (1983) and Schmalhofer and Glavanov (1986) there seems to be a general agreement in discourse psychology that the mental text representation should in fact be regarded as a mental structure that consists of different levels of representation (see among many others Fletcher, 1994; Fletcher & Chrysler, 1990; Graesser, Millis, & Zwaan, 1997; Kintsch, 1998). At least the levels surface code, textbase, and situation model can be distinguished. These levels differ both with respect to the cognitive processes involved in their construction, and to their memory characteristics. It was argued that this widely shared belief in three distinct levels of representation implies a challenge for a cognitive theory of coherence relations (Sanders et al., 1992, 1993; Sanders & Spooren, 2001), to the extent that such a theory should be precise about the level of representation at which coherence relations are represented. The statement: ‘Coherence relations are part of the text representation’ is simply too general.

Unfortunately, discourse processing studies do not provide much help in answering the question at what level coherence relations should be represented. Although the results from experiments show that causal coherence relations are part of the meaning representation rather than of the surface code representation, the question regarding the exact nature of this meaning representation remains. The general view seems to be that causal relations are part of the situation model (see for instance, Graesser, Millis & Zwaan, 1997; Zwaan & Radvansky, 1998), but the evidence on which this view is based, is both indirect and subject to alternative interpretations, as will be argued below. This situation clearly asks for a study in which direct evidence is obtained. Not only will such a study serve theories of discourse processing, but it will also provide an empirical foundation for further development of cognitive theories of coherence relations.

Hence, the study reported in this chapter focuses at the question at what level causal coherence relations are represented. First, the theoretical and empirical status of both the levels of representations and coherence relations is investigated. Then the results of an experiment are presented in which the research question was addressed directly: a sentence recognition paradigm was used to investigate the level of representation of local causal coherence relations in expository (informative) text. Finally, the implications of the results for cognitive theories of discourse representation will be discussed.
Chapter 4

4.2 Levels of representation of causal coherence relations

With the research reported in this chapter, we seek to corroborate the cognitive processing account (CPA) presented in chapter 1. The CPA advances the hypothesis that explicit causal relations are represented at both the textbase and the situation model levels of representation. Let us first reconstruct the argumentation in chapter 1 that leads to this hypothesis.

First of all, the general cognitive claim that coherence relations are part of the meaning representation entails that coherence relations are definitely not surface code phenomena. After all, the interpretation of coherence relations is independent of the presence of the surface indicator of the relations (Sanders & Pander Maat, 2006). For instance, both in sentence (1) and in sentence (2) a causal relation can be established between the clauses, even though this relation is made explicit with the connective because in (1), but remains implicit in (2).

(1) Bashir had to be hospitalized, because his heart problems had increased severely.
(2) Bashir had to be hospitalized. His heart problems had increased severely.

This implies that the interpretation of the relation does not depend on whether an explicit surface indicator of the relation is present (see also Graesser, McNamara & Louwerse, 2003). A related observation is that there is no one-to-one correspondence between the relation and the surface indicator. In (3), for instance, the same connective is used as in example (3), but the relation between the clauses is totally different: here we are dealing with a so-called epistemic, pragmatic or subjective causal relation (Sweetser, 1990; Sanders, 1997; Sanders, 2005b), which can be paraphrased as “It has to be the case that Bashir was hospitalized, because I saw an ambulance picking him up at home”.

(3) Bashir had to be hospitalized, because an ambulance picked him up at home.

Therefore, coherence relations should clearly be regarded as part of the meaning representation of the text and not as surface code phenomena. This means that they are part of the textbase and/or the situation model, whereas their linguistic surface indicators are represented at the level of the surface code representation.

It was explained in chapter 1 that the textbase is a structure in which the propositions that are derived from the explicit text are interconnected by means of referential links and, possibly, coherence relations. As such, the textbase describes the meaning of the text per se (Fletcher, 1994). The situation model, on the other hand, is a mixture of text derived and knowledge derived propositions (Kintsch, 1998). As was described, these knowledge derived propositions are the product of knowledge based inferences.

In the case of a causal relation marked with the connective because, the connective explicitly expresses the causal connection between the events. As the textbase represents the explicit information in the text, it may be assumed that the textbase representation of the relation contains the proposition that is expressed by because. To be more precise, a sentence such (4) may be represented as CAUSE((PUSH, BILL, JOHN), (FALL, JOHN)). This proposition
The Representation of Causal Coherence Relations in Expository Text

consists of the predicate CAUSE and its arguments are the propositional representations of the individual clauses.

(4) John fell, because Bill pushed him.

In the case of a completely explicit text, the situation model is fully described in the textbase (McNamara & Kintsch, 1996). Thus, the causal relation between the events in (4) is represented both at the level of the textbase and at the situational level.

This idea is highly compatible with a hypothesis put forward by Noordman & Vonk (1998) regarding the function of the causal connective *because* during reading. In fact, they hypothesize that *because* has three functions: it acts as marker of segmentation, integration and inference.

*Because* as a segmentation device indicates how the current sentence is linguistically structured with respect to the previous one (in the case of backward causal relations). For instance, it indicates that the current clause is a subordinate clause. The segmentation function may be related to the construction of the surface code representation as it is concerned with the way the information is linguistically structured. A second function of *because* is that of integration. As an integrative device, *because* signals how the information in the current sentence has to be integrated with the previous one: the current sentence provides the cause for the event described in the previous one. In other words, the connective signals how the two sentences can be integrated in a causal relation. In this sense a discourse internal relation can be established by the reader. The integration function of connectives is in line with the processing advantage of connectives that has been reported in several studies (Cozijn, 2000; Millis & Just, 1994; see also Britton, 1994; see chapter 3), and with the view on connectives as processing instructions that inform the reader on how to relate the previous information with the upcoming one (Britton, 1994; Sanders & Noordman, 2000; Sanders & Spooren, 2001, 2007).

The third function of the connective is the inference function. The connective is a signal to the reader to infer relevant background knowledge. This inference process may be related to the construction of the situation model: the information in the text is related to the background knowledge of the reader. The connective signals the activation of the missing major premise of the underlying reasoning pattern (Cozijn, 2000; Noordman & Vonk, 1998; Halldorson & Singer, 2002; Singer et al., 1992).

Interestingly, Noordman & Vonk (1997, 1998) hypothesize that the integration and inference function of the connective *because* may be related to the construction of the textbase and the situation model. During integration, discourse internal relations are established, and discourse internal relations are also part of the textbase. During the inference process, the reader activates relevant background knowledge and relates this information to the text. In other words, relationships are established between the readers’ knowledge and the explicit text. This is a key part of the construction of situation model representations (Kintsch, 1998). The hypothesis that integration and inference are processes in which the textbase and the situation model are constructed implies that causal coherence relations are represented at both of these levels.
However, in the discourse processing literature, the most prevalent view seems to be that causal relations should be considered as part of the situation model. The evidence that is presented as support for this view, however, is indirect and can be interpreted in alternative ways. Zwaan and Radvansky (1998) for instance, interpret the effects of connectives found by Millis and Just (1994) as adding support for the idea that causal relations are an important part of the situation model. By contrast, it was just explained that exactly the same results corroborate Noordman and Vonk’s (1997) hypothesis concerning the functions of the connective because, and hence can be taken to imply that causal relations are also represented at the textbase level. The question is therefore what exactly these findings contribute to insights at the situation model.

Although more recent research has shown additional evidence for the view that the presence of connectives leads to a more coherent mental representation (see, among others Degand, Lefèvre, & Bestgen, 1999; Degand & Sanders, 2002; Kamalski, Sanders, & Lentz, to appear; Millis & Just, 1994; Sanders, Land & Mulder, 2007), this does not necessarily mean that the relations expressed by the connectives are part of the situational representation. Indeed, the presence of the connective may also result in a more coherent textbase (see for instance McNamara & Kintsch, 1996, for their view on how cohesive devices like connectives, may improve the quality of the textbase).

Fletcher (1994) also concludes that causal relations play an important role in the comprehension of (narrative) text. He suggests that causal relations can be considered to be part of the situation model. One of his arguments is that causal relations can link several propositions. Fletcher (p. 600) considers the following example:

(5) The 750 pound orange and black cat walked into the bedroom and the burglar left very quickly.

Fletcher argues that a reader’s construction of a causal relation between the events involves mapping the six propositions on two events: a tiger entering the bedroom and the burglar running away. Because this process involves the linking of several propositions, it should be seen as process involved in a situation model type of representation. However, we could also analyze (5) as a causal relation between only two propositions. The fact that these propositions have embedded propositions does not necessarily mean that they should be considered part of the situation model; embedded propositions may just as well be considered to be part of the textbase (see Kintsch, 1998; Singer, 1990).

In sum, theoretical considerations lead to the hypothesis that explicit causal relations may be represented at both the textbase and the situation model level of representation, although only indirect evidence is available to support this hypothesis. For this reason, the experiment reported below, aims at obtaining direct evidence for the existence of the two levels of representation for explicit causal relations.
4.3 The experiment

4.3.1 Method

In the experiment we made use of a sentence recognition paradigm. In such a paradigm, participants are asked to indicate whether they remember having read a particular test sentence.

The sentence recognition task has proven to be a useful in showing the psychological validity of the distinction between surface code, textbase, and situation model (see Fletcher, 1994, for an overview). By making use of this task, researchers have repeatedly shown that recognition memory performance is influenced by the number of levels of representation that distractor sentences share with an original sentence (Fletcher & Chrysler, 1990; Kintsch, Welsh, Schmalhofer, & Zimny, 1990; Schmalhofer & Glavanov, 1986).

The logic of the sentence recognition paradigm is that the more levels of representation the test sentence shares with an original sentence, the more likely a subject is to say ‘yes, I recognize this sentence’. By manipulating the number of levels of representation the test sentence shares with the original sentence, the existence of each level can be determined. Although the specific materials and experimental procedures vary from study to study, the general experimental design is the same. In all studies there are distractor/test sentences, which are used in 4 conditions: A condition in which the test sentence shares all three levels of representation with the original sentence, a condition in which the test sentence differs at the level of the surface code, a condition that differs at the level of the surface code and the textbase, and a condition that differs at all three level of representation from the original sentence. The difference between the conditions is summarized in the following table:

<table>
<thead>
<tr>
<th>Condition</th>
<th>Surface code</th>
<th>Textbase</th>
<th>Situation Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: A ‘+’ indicates that the condition shares the specific level.

As an example, consider the experiment by Schmalhofer and Glavanov (1986). By making use of the sentence recognition task in combination with a speed accuracy trade off paradigm, Schmalhofer and Glavanov showed that retrieval from long term memory differs from level to level. In their study, one of the tasks of the participants was to determine whether or not a particular sentence was present in the text they had just read. In the example text, a manual for the programming language LISP, the following sentence appeared:

(6) PSY100 is a legal atom that concludes with a number. (Original)
There were four conditions in which this test sentence appeared during the recognition task. The first condition is the original condition, in this condition the test sentence appeared as it did in the text. The second condition is a paraphrased condition, as in sentence (7).

(7) PSY100 is a legal atom that ends what a numeral. (Paraphrase)

The difference between the two conditions situated at the level of the surface code. If it can be shown that subjects are less likely to say ‘yes’ to (7) than to (6) this can be taken as evidence for the existence of the surface code representation. In the third condition, not only the surface code was changed, but also the meaning of the sentence.

(8) PSY100 is a legal atom that begins with a letter. (Meaning changed).

Note that this sentence differs from the original one on two levels of representation, but that it differs from the paraphrased condition at only one level of representation. The meaning changed condition does not share the textbase representation of the original sentence, but the paraphrased condition does. Therefore, if it can be shown that subjects are less likely to say ‘yes’ to sentence (8) than to sentence (7), there is evidence for the existence of the textbase representation of the original sentence.

In the fourth and last condition, the correctness change condition, all three levels of representation that can be constructed on the basis of the test sentence differ from the original sentence.

(9) PSY.100 is a legal atom that contains a dot. (Correctness changed).

However, with respect to the previous sentence (8) the only difference is that the situation model in (8) corresponds to the model of the original sentence, whereas the model of sentence (9) does not. Again, if this difference results in a lower likelihood of a yes response, then we have evidence for the existence of a situation model representation of the original sentence.

Fletcher and Chrysler (1990) argue that conclusions with respect to the existence of the levels of representation can only be drawn if all the test sentences differ from the original sentence by the same degree of change, both at the level of the surface code and the textbase. In their experiment this was true for all materials:

(10) George says that his wife was angry when she found out that the necklace cost more than the carpet. (Original)
(11) George says that his wife was angry when she found out the necklace cost more than the rug. (Surface distractor)

---

9 The materials were not presented to the subjects in this form. Fletcher and Chrysler (1990) used a 2-alternative forced choice paradigm. An example for the surface test is ‘George says that his wife was angry when she found out the necklace cost more than the carpet/rug.’ Subjects had to indicate which of the two words carpet and rug appeared in the original text.
(12) George says that his wife was angry when she found out the necklace cost more than the painting. (textbase distractor).
(13) George says that his wife was angry when she found out the necklace cost more than the vase. (Model distractor)

Note that the original text describes the ordering of the objects in terms of what they are worth. The original ordering was CARPET $<$ PAINTING $<$ NECKLACE $<$ VASE.

As we can see, the test sentences all differ from the original sentence by only one word and/or proposition. Fletcher and Chrysler also present experimental evidence that supports the same amount of change was introduced from manipulation to manipulation.

In our study, we have adopted the sentence recognition paradigm in order to find out at what level of representation causal coherence relations are represented. Unlike previous studies, we are not interested in the representations of a particular event, but in the relationship between two events. Following Fletcher and Chrysler (1990) we have minimized changes between conditions by only changing one word at a time. We have established this by only changing the connective that was used to indicate the coherence relation between the events.

In the next section we will illustrate how this was done exactly, but the logic behind the task that we used was as follows. Consider the following example:

(14) John fell, because Bill pushed him.

This sentence expresses the following proposition CAUSE((PUSH, BILL, JOHN), (FALL, JOHN)). As we are only interested in the relation between the events the following short-hand version of this proposition can be used: CAUSE(P2, P1). As was explained above, this proposition is supposed to be represented at the level of the textbase. If we change because for a connective that, at the level of the explicit text, expresses a different proposition and the amount of incorrect yes increases in comparison to a paraphrased version of (14), we obtain evidence for the existence of a textbase level of representation of the relation between the clauses in (14). We have used the connective after for this purpose. For an example, see (15).

(15) John fell, after Bill pushed him.

Changing because in the original sentence for after, yields the proposition AFTER(P2, P1). Thus, on the level of the explicit text, a different proposition is expressed. However, the after-proposition is still compatible with the original causal model underlying (14). To obtain evidence for a situational representation of the causal connection we therefore need to include a condition which is not compatible with the original model. For this means we used the connective and. Changing because for and amounts to changing the causal proposition to a relations proposition of the type: CONJUNCTION(P1, P2). Note that logically speaking the and proposition is still compatible with the original model (as causality entails additivity), but a causal interpretation of the and sentence entails a different order of the events. Consider, for instance, (16) and (17) below:
(16) John fell and Bill pushed him.
(17) John fell, therefore Bill pushed him.

A causal interpretation of the sequence of events in (16) entails that Bill’s pushing was caused by John falling and is not compatible with (18), which is the original causal relation.

(18) John fell, because Bill pushed him.

If readers indeed construct a situation model representation of the original sentence, changing *because* for *and* should lead to a further decrease in incorrect *yes* responses in comparison to the meaning changed condition with *after*.

**Materials**

Participants read 20 newspaper articles (see appendix for examples of the original Dutch materials). In total, 32 causal relations were present in the texts. These relations were marked by the Dutch connectives *doordat* (“as a result of the fact that”) and *omdat* (“because”), which can both be translated by English connective *because*. The following sentence appeared in one of the texts:

(19) Bashir was in het ziekenhuis opgenomen, omdat zijn hartproblemen ernstig waren toegenomen. (Original)

‘Bashir had been hospitalized, because his heart problems had increased severely.’

During the recognition task, this target sentence appeared in one of four possible conditions: original, paraphrase, meaning changed, and model changed. The paraphrased version of the original sentence is as follows.

(20) Bashir was in het ziekenhuis opgenomen, doordat zijn hartproblemen ernstig waren toegenomen. (Paraphrase)

‘Bashir had been hospitalized, because his heart problems increased severely.’

In the paraphrase condition, we changed the Dutch connective *omdat* by *doordat*. In the English translation we used *because*, since *because* is the best translation of both connectives; *doordat* can only be translated literally in English as a cue phrase, for instance “as a result of (the fact that)” (Knott & Sanders, 1998). The difference between the sentences in (19) and (20) is located at the level of the surface code, because the connectives indicate the same relationship at the textbase and situation model level.

In the third condition we did not only change the exact wording but also the meaning of the relationship on the level of the textbase.
(21) Bashir was in het ziekenhuis opgenomen, nadat zijn hartproblemen ernstig waren toegenomen. (Meaning changed)

‘Bashir had been hospitalized, after his heart problems had increased severely’

In the final condition we replaced because by and, so that a maximal difference existed between the original sentence and this “model changed” condition:

(22) Bashir was in het ziekenhuis opgenomen en zijn hart problemen waren ernstig toegenomen. (Model changed).

‘Bashir had been hospitalized and his heart problems had increased severely.’

In this condition the sentence differs from the original at all three levels of representation.

Participants

33 participants took part in the experiment. They were all students at the Faculty of Arts at Utrecht University. They were paid for their participation.

Design and Procedure

The 20 texts were presented in a booklet. After each text two text comprehension questions were asked. The questions were used in order to keep the subjects from shallow processing of the texts. After the subjects read all the texts, a yes-no recognition task was administered. During this task the subjects had to judge whether they had read a particular test sentence. They had to circle ‘yes’ if they had and ‘no’ if they had not. In total 32 target relations were judged and 32 filler sentences.

A Latin Square Confounded Factorial design (Kirk, 1982) was used for the recognition task. This means that each target sentence appeared in each condition, and also that each subject judged sentences in each condition. In total the subjects judged 8 sentences per condition.

In order to prevent possible individual preferences for doordat or omdat (both connectives can be used to express the causal relations expressed in the experimental texts, but there may still be individual preferences for one over the other) from influencing the recognition performances we used both omdat and doordat in the target sentences. Half of the subjects read targets 1-16 in the omdat version and texts 17-32 in the doordat version, the other half read texts 1-16 in the doordat version and texts 17-32 in the omdat version. This design also had consequences for the paraphrase materials. Original omdat sentences were replaced by doordat sentences, and original doordat sentences by omdat paraphrases.
The analysis of recognition data

In the literature, the analysis of yes-no recognition data focuses on the computation of $d'$ values (Schmollhofer & Glavanov, 1986; Kintsch et al., 1990) or $A'$ values (Radvansky, Curiel, Zwaan, & Copeland, 2001), which is the non parametric counterpart of $d'$ (Neath, 1999). In recognition memory research, the values $d'$ and $A'$ are used to indicate how well subjects are able to discriminate between old and new items. The value of $d'$ is calculated by entering the proportion yes-responses to old items, the hit rate, and the proportion yes-responses to new items, the false alarm rate. If the hit rate and false alarm rate are equal, $d'$ gets a value of 0, which indicates that the subjects cannot distinguish between old and new items.

Schmollhofer and Glavanov (1986) applied the following procedure in calculating $d'$-values. With respect to the surface code, the yes-responses to original sentences were used to calculate the hit rate, and the yes-responses to paraphrased sentences were used to calculate the false alarm rate. The resulting $d'$-value indicates the strength of the surface code representation. In the case of the textbase, the yes-responses to the paraphrased sentences were taken as hits, and the yes-responses to meaning changed sentences were taken as false alarms. The strength of the situation model was determined by calculating a $d'$-value on the basis of yes-responses on meaning changed sentences (hits) and yes-responses on correctness changed items (false alarms).

Kintsch et al. (1990) used a somewhat different approach. In their analysis the yes-responses on original sentences were always treated as hits, and the yes-responses on the sentences in the other conditions as false alarms. For the strength of the surface code representation a $d'$-value was calculated on the basis of responses to original and paraphrases. In order to test for the textbase, first a $d'$-value was calculated by using the hit rate and a false alarm rate based on responses to easily inferable sentences (which they used in stead of meaning changed sentences). This particular value then, was compared to the $d'$-value for the surface test.

Although this use of $d'$ has proven to be very useful, a number of assumptions have to be made in order for this method to be reliable. Two of these assumptions are particularly important. In the first place, the use of $d'$ as a performance index is derived from signal detection theory (Green & Swets, 1966; Macmillan & Creelman, 1991; Neath, 1999; Smith and Duncan, 2004). In this theory it is assumed that the decision to answer yes to a particular test item depends on its familiarity value. If the familiarity of the item exceeds some familiarity value threshold, a yes-response will be given by the participant. This single threshold holds for both original items (old items) and new items, and therefore the model is a single threshold model. This means that if $d'$ is used as a measure we will also have to assume that recognition performance can be accounted for by this single threshold model. However, other models of recognition memory performance may be valid and these models may (or may not) require other measures than $d'$ (or $A'$) (see Slotnick & Dodson, 2005; Smith and Duncan, 2004, Yonelinas, 1994, 1997, 1999; Yonelinas, Dobbins, Szymanski, Dhaliwal, & King, 1996). In the second place, when using $d'$ or $A'$ the researcher assumes that there is no random variance between the specific items that were used. That is, the $d'$ or $A'$ value is computed on the basis of condition means/proportions per subject. This is an example of the
language-as-a-fixed-effect-fallacy (Clark, 1973). The problem associated with averaging over items, is that when this procedure is followed, one cannot generalize over the items that were used in the particular experiment (see chapter 2).

We prefer an analysis in which we do not have to make any assumptions with respect to the underlying recognition model, the distributions of the conditions, and the variance between items. For that reason we have chosen to model the probability of a yes-response by using a multilevel logit-model for binary responses. We modeled the probability of a yes-response in each condition, where we allowed the scores to vary both between items and participants and This enables us to provide evidence that the results can be generalized over both subjects and items.

4.3.2 Results

If the participants construct all three different levels of representation of the relation, we would expect that the probability of saying yes would show a decreasing pattern. The probability of a yes response should be highest in the original condition. If this probability is lower in the paraphrase condition, we take this as evidence for a surface code representation. If changing the meaning of the relation the propositional level decreases the probability even more, we will take this as evidence for the propositional textbase. In the model condition the probability should be the lowest, if a situation model is constructed.

In the following table the results are summarized.

**Table 4.2 Recognition performances in relation to condition**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Observed Proportion ‘yes’</th>
<th>Logit values of condition means ($\beta$)</th>
<th>Predicted Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>0.75</td>
<td>1.098 (0.150)</td>
<td>0.75</td>
</tr>
<tr>
<td>Paraphrase</td>
<td>0.76</td>
<td>1.149 (0.151)</td>
<td>0.76</td>
</tr>
<tr>
<td>Meaning Changed</td>
<td>0.72</td>
<td>0.954 (0.144)</td>
<td>0.72</td>
</tr>
<tr>
<td>Model Changed</td>
<td>0.36</td>
<td>-0.615 (0.135)</td>
<td>0.35</td>
</tr>
</tbody>
</table>

The predicted probabilities are presented in Figure 4.1.
Tests of significance show that all parameters are significantly different from 0, which corresponds to the logit value of a proportion of 0.5 (Original: $\chi^2(1) = 53.6, p < .001$; Paraphrase: $\chi^2(1) = 58.0, p < .001$; Meaning changed: $\chi^2(1) = 43.8, p < .001$; Model changed: $\chi^2(1) = 20.9, p < .001$).

With respect to the differences between the conditions, the Original, Paraphrase and Meaning Changed conditions do not differ from each other (All $\chi^2 < .87$). However, these three conditions do differ from the model changed condition in a simultaneous contrast ($\chi^2(1) = 111.1, p < .001$).

4.3.3 Conclusion

The aim of this experiment was to obtain evidence for the hypothesis that local causal relations are represented at the textbase and the situation model levels of representation. Clearly, the experiment failed to provide evidence for the local causal relations are represented at both levels. Indeed, changing the textbase of the text by changing the proposition that is expressed by the connective from causal (because) to temporal sequence (after), in the meaning change condition, did not result in a lower probability of a yes-response. In other words, readers were just as likely to incorrectly recognize the after sentence as correctly recognizing the original because sentence. The only significant decrease in the probability of a yes-response is found in the model changed condition. This leads to the conclusion that local causal relations can be best considered as part of the situation model readers construct of the text.
4.4 General Discussion

The results of this experiment are in agreement with the general view that causal relations are part of the situation model. Even if the causal relation is explicitly marked with a connective, the reader does not seem to construct a textbase representation of the relation between the events described in the text: readers cannot distinguish between a causal relation expressed by *after* and one expressed by *because*. This conclusion has implications for theories of coherence relations but also for theories of discourse processing.

Before we formulate these, we want to return to the question whether or not causal relations between sentences (or the events described in these sentences) are represented at the level of the textbase, too. The fact that we failed to find evidence that readers do not make a textbase representation of the relation, does not imply that no textbase representation is constructed at all. Individual propositions derived from the text and referential links (argument overlap) between them may still be part of the textbase. Coherence relations on the other hand are not part of the textbase representation.

The conclusion that local causal relations are not represented at the level of the textbase is based on the fact that changing the proposition that is expressed by the connective from causal (*because*) to temporal precedence (*after*) did not lead to a better rejection performance. This is a clear finding in favor of the situation model representation of causal coherence relations. However, from a methodological point of view it should be remarked that this is – strictly speaking – a null-result, which should be interpreted with some caution. Furthermore, arguments can be raised against our “causal relations are represented at the situation level model” conclusion.

One of the possible counter-explanations is that *after* does not only express a temporal proposition, but can also be used to express a causal proposition. That is, *after* can be interpreted as an underspecified indicator of the causal relation (cf. Spooren, 1997). The reasoning would be that *because* and *after* express the same proposition and that it is for this reason that we found However, the interpretation of such an underspecified connective is highly constrained by the context and this context enables the interpretation of the connective beyond its semantic domain (Fabricius-Hansen, 2005). In other words, *after* can be interpreted more specifically, that is causally, if the sequence as a whole leaves room for such a causal interpretation. For this reason, we do not accept the idea that *after* in itself expresses a causal proposition: *after* denotes a coherence relation of temporal sequence, *because* consequence-cause. And even if we would accept the stance that the two connectives do express the same coherence relation, it should be best considered as a similarity at the level of the situation model. The sequence with *after* can be interpreted as causal when the reader has knowledge of the causal relation between the events. That is, the reader has to use his knowledge to enrich the information that is explicitly expressed in the text. This process, the enrichment (or elaboration) of the information in the text, is exactly what happens in the construction of situation models (Kintsch, 1998). However, even if readers of the *after* have assigned it a causal interpretation, this does not mean that two connectives *because* and *after*
express the same relation. Rather, the conclusion should be there appears to be a similarity at the level of the situation model.

Another counter-explanation might be that the difference in meaning between *because* and *after* is not of the same magnitude as the difference between *after* and *and* (compare Fletcher & Chrysler, 1990). We believe this point of criticism makes sense indeed, but in stead of a problem for the conclusion that no textbase is constructed, it is in fact a problem for the conclusion that readers construct a situational representation of the relation. The fact that *because* and *after* are very close in meaning cannot be explained in terms of a propositional textbase: different propositions are expressed in the two conditions. We believe that the meaning of *because* and *after* is close, due to the fact that they express the same relation at the level of the situation model. The fact that *and* really seems to be different from the other conditions, does not have any implications for the conclusion that the causal relation is not represented at the level of the textbase. Rather, it may cast doubt on the conclusion that we found evidence for a situational representation of the relation. The possibility remains that low probability of a *yes-* response is not due to *and* expressing a different model, but to differences in plausibility, frequency of occurrence (outside the experiment, that is) and other differences between *and* and the other connectives.

We have concluded that causal coherence relations are represented at the level of the situation model of the text and not part of the textbase representation. We may have to modify this conclusion: the relations are not represented at the level of the textbase, and most likely on the level of the situation model. However, given the result that *because* and *after* are very similar, the lower amount of *yes-* responses on the *and* sentences, combined with the general believe in the literature that causal relations are part of the situation model, makes us think we are on the safe side, even with the stronger conclusion.

As we said above, the conclusion that causal coherence relations are part of the situation model has consequences for text linguistic and cognitive theories of coherence relations. For instance, if it is true that coherence relations are responsible for the coherence of a text – which we believe to be true - then the coherence of a text depends on how well the events in the text are connected at the level of the situation model. It is not hard to imagine why this should be the case, since for any text to be coherent it should be compatible with the readers’ knowledge. In the case of causal relations, the text is coherent when it is compatible with the readers’ knowledge of the causal relation. In other words, a text is causally coherent if the events are related by means of a causal relation in a reader’s knowledge representation. The fact that causal relations are part of the situational representation implies that the interpretation of these relations depends largely on the characteristics of the reader. That is, what exactly the text representation will look like, depends on the knowledge of the reader, his reading goals and or strategies. There is massive evidence for the way in which these reader characteristics indeed play a crucial role in the construction of situation models (see among many others Kintsch, 1998; McNamara & Kintsch, 1996). The consequence for cognitive theories of coherence relations is that the relational structure of the text cannot be properly described without reference to the reader (if the text is analyzed from the readers’ perspective). And indeed, a research program along these lines should include issues of text-
reader interaction. Of course, the situation model is not entirely based on the reader alone; the text provides the ingredients and poses restrictions on the possible situation models.

Under the assumption that coherence relations are necessary for text comprehension, the conclusion that local causal relations are represented at the situation model has consequences for theories of inferential processing, too. Indeed, even such local relations are the product of processes that are involved in the construction of situation models. This casts serious doubts on theories that predict that only automatic inferences are regularly performed during reading (e.g., McKoon and Ratcliff, 1992). Clearly, the results are much more in line with a view of text comprehension in which the reader aims at richly developed mental representations of the text: even at the most local level, enriched representations are constructed. That is, the sentences marked with *after* may be represented as causal relations, while there is in fact no need for additional inferences. And this in turn, leads us to the idea that causality in discourse is such a strong coherence relation that readers seem likely to infer it whenever they can. Ideas like these are discussed in chapter 5.
Chapter 5: The Processing Advantage of Causal Coherence Relations. Causality-by-Default or Schematic Text Structural Expectations?

5.1 Introduction

Chapters 3 and 4 focused on the processing and representation of explicit causal relations. In the present chapter the processing and representation of implicit coherence relations is investigated. This chapter revitalizes a major issue from text processing research in the 80’s and 90’s, which revolved around the question whether text comprehension can be best considered as a bottom-up or top-down process (Van Dijk & Kintsch, 1983; Singer, 1990). In this chapter, this discussion is taken up again by focusing on a particular important type of causal coherence relation: the Problem-Solution relation. As we will see in more detail below, an interesting processing phenomenon is associated with this relation; although text linguistically speaking the relation is one of the most complex relations (Hoey, 1983, 1985; Hoey & Winter, 1986; Sanders, Spooren, & Noordman, 1993), information embedded in it is both processed faster and remembered better than when the exact same information is embedded in a less complex List relation (Sanders & Noordman, 2000). This paradox can be explained from both a bottom-up and a top-down perspective (Sanders, 2005a; Sanders & Noordman, 2000; see also chapter 1). These explanations will henceforth be called the CAUSALITY-BY-DEFAULT (Sanders, 2005a) and the SCHEMATIC (TEXT) STRUCTURAL EXPECTATION hypotheses. The first hypothesis focuses on the special status of causal relations in text processing, which has been confirmed by studies that show the representational advantages (Black & Bern, 1981; Bradshaw & Anderson, 1982; Myers, O’Brien, Balota, & Toyofuku, 1984; Trabasso & Sperry, 1985; Trabasso & Van den Broek, 1985) and the processing effects of causally related information (Golding, Millis, Hauselt, & Sego, 1992; Keenan Baillet, & Brown, 1984; Linderholm, 2001; Myers, Shinjo, & Duffy, 1987; Traxler, Bybee, & Pickering, 1997 Traxler, Sanford, Aked, & Moxey, 1997; Wolfe, Magliano, & Larsen, 2005). The TOP-DOWN hypothesis focuses on the knowledge of the reader. As was argued in chapter 1, processing differences between relations may arise, because knowledge of text structure aids the reader in interpreting the currently processed information (Van Dijk & Kintsch, 1983; Kintsch & Yarbrough, 1982).

The effect of causality on on-line processing is the main focus of this chapter. Two experiments will be presented on the basis of which the validity of the SCHEMATIC STRUCTURAL EXPECTATIONS and the CAUSALITY-BY-DEFAULT hypotheses can be assessed. Before these experiments are described, results from previous processing studies will be discussed, on the basis of which the two hypotheses were formulated.
5.2 The effect of relation type on on-line text processing

In the discourse processing literature, two types of study can be identified, which are directly relevant to our interest in the effect of relation type on on-line processing. In a first type of experiments, researchers investigated the processing of identical sentences that were connected to the previous sentence by a relation that varied in causal strength. The impact of this variation in causal strength was studied. In the second set of studies, the influence of causal coherence relations was investigated by studying identical sentences, which where connected to previous text by different coherence relations, for instance causal and non-causal relations or different kinds of causal relations.

5.2.1 Causal strength

In the first line of studies, researchers focused on the differences between relations within the class of causal relations (Golding, Millis, Hauselt, & Sego, 1992; Keenan Baillet, & Brown, 1984; Linderholm, 2001; Myers, Shinjo, & Duffy, 1987; Wolfe, Magliano, & Larsen, 2005). In these experiments the strength of the causal relation is varied. For instance, Keenan et al. (1984) use materials like the following.

(1) Joey’s brother punched him again and again. The next day his body was covered with bruises.
(2) Racing down the hill, Joey fell of his bike. The next day his body was covered with bruises.
(3) Joey’s crazy mother became furiously angry with him. The next day his body was covered with bruises.
(4) Joey went to a neighbor’s house to play. The next day his body was covered with bruises.

In the examples above, the numbering from (1) through (4) mirrors their ordering with respect to the strength of the causal relation that exists between the first and second sentence. That is, in (1) the causal relation is considered strongest, whereas in example (4) the relation is weakest. The researchers’ intuitions that the relations differ in terms of causal strength is confirmed by participants in a rating task. According to Keenan et al. (1984), the dissimilarity in causal strength arises from a difference in the probability of the consequence. For example, the fact that Joey’s body was covered with bruises is a much more probable consequence of the event of ‘being punched’ than of the event ‘going to play at a neighbor’s house’.

The results from the causal strength studies show that a decrease in causality leads to an increase in reading times. Elaborative processing seems to be the most likely explanation for this pattern of results (Keenan et al., 1984; Myers et al., 1987; Van den Broek, 1994). In the case of the highly connected relations, no elaborate processing is needed, because the events are closely related. At intermediate levels (examples (2) and (3)), the reader will have to use background knowledge in order to be able to construct a causal connection. This use of
The Processing Advantage of Causal Coherence Relations

5.2.2 Different types of causal relations

A second set of studies also focuses on the effect of relation type on the processing of text, but compares different types of (causal) relations. Traxler, Bybee, and Pickering (1997) studied the processing difference of an identical target clause, which is related to a previous clause by either a causal or an argumentative relation. As an example, consider the following sentences from their materials.

(5) Heidi felt very proud and happy, because she won first prize at the art show.
(6) Heidi could imagine and create things, because she won first prize at the art show.

In (5), the relation describes a (semantic) **Consequence-Cause** relation and in (6) a (pragmatic) **Claim-Argument** relation holds. The difference between these types of relation has been described as semantic, content or objective relations versus pragmatic, epistemic or subjective relations (Sanders, Spooren, & Noordman, 1993, Sanders, 1997). Traxler et al. use the terms causal versus diagnostic sentences. In the first sentence, the fact that Heidi felt very proud and happy is caused by the fact that she won first prize. Thus, the event in the first clause is caused by the event in the second clause. In the second sentence, “Heidi could imagine and create things” is a claim made by the author or speaker that is supported by the argument “because she won first prize at the art show”. In an eye tracking experiment, Traxler, Bybee, and Pickering (1997) showed that a processing difference exists between the two types of relation. More precisely, in the **Claim-Argument** condition (6), the total reading times are longer on ‘first prize’ than in the **Consequence-Cause** condition (5).

Traxler, Sanford, Aked, and Moxey (1997) investigated the difference between similar causal and argumentative relations. They used materials that are comparable to that of Traxler, Bybee and Pickering (1997). Using a self paced reading paradigm, Traxler, Sanford et al. showed that simple causal sentences, like the one in (5) are processed faster than the argumentative relations. According to the authors this difference arises from the fact that in the interpretation of (6), additional information has to be added in the representation in order to fully understand the sentence. This additional information is that someone believes that Heidi could imagine and create things. This unstated mental state of the author or speaker remains implicit in sentence (6), and has to be incorporated in the mental representation in order for sentence (6) to be coherent. Therefore, in the processing differences would arise from additional processes involved in the construction of the representation.
These results indicate that relations that differ in terms of their structure may also involve different processes during the construction of their representations. Indeed, the fact that the additional unstated mental state of the author or speaker has to be inferred is a direct consequence of the structural differences between the sentences. In this case it is the structural difference (in terms of the coherence relations in the representation of the reader) between semantic and pragmatic relations. This difference is one of the defining cognitive primitives in the cognitive coherence account of Sanders, Spooren, and Noordman (1993), see chapter 1. A similar conclusion can be based on research reported by Sanders and Noordman (2000) who have shown that another structural difference between coherence relations influences text processing. Sanders and Noordman (2000) manipulated the basic operation (causal versus additive) underlying the relation between a target sentence and its context and thereby the nature of the coherence relation. In their experiment, short newspaper articles are presented. These newspaper articles each contain a target sentence that is either embedded in a Problem-Solution context or in a List context. In the Problem-Solution contexts, the target information provides the reader with the solution to the problem described in the previous context. In the List texts the same information is part of an enumeration. Special care was taken to ensure that there were no additional systematic differences between the two conditions, for instance in terms of the given-new structure of the text.

From a text-linguistic perspective, Problem-Solution relations are more complex than List relations (Hoey, 1983, 1985; Hoey & Winter, 1986; Sanders, Spooren, & Noordman, 1993). Sanders and Noordman (2000; Sanders, 2005) explain where the complexity of the Problem-Solution relation arises from. The Problem-Solution relation

"consists of a negatively evaluated situation (the problem) that leads to an action undertaken to take away this negatively evaluated situation. This action result in a new situation that is evaluated positively" (p. 44).

This implies that there are two causal relations that can be inferred in every Problem-Solution structure. One causal relation exists between the negatively evaluated situation and the action, and a second between the action and the new situation. By contrast, List is an example of an additive relation. These relations are assumed to be less complex than causal relations. The reason for this is that causal relations presuppose additive relations, as is widely acknowledged in logic and linguistics (see Allwood, Anderson & Dahl, 1977; Sanders, 2005b).

The difference in complexity between Problem-Solution and List relations might lead one to expect that information embedded in a Problem-Solution relation will be processed slower than when it is embedded in a List relation. Generally speaking, the more complex the information that has to be processed, the more cognitive energy will be spent by the reader, and hence the longer the processing time will be. This hypothesis can be called the CAUSAL COMPLEXITY hypothesis. Surprisingly, Sanders and Noordman’s (2000) results show a pattern that is exactly the opposite of this CAUSAL COMPLEXITY hypothesis. Reading times collected in a self paced reading paradigm appeared to be shorter in the Problem-Solution condition than in the List
condition. Sanders and Noordman (2000) provide two explanations why the more complex relation is processed faster. According to the CAUSALITY-BY-DEFAULT hypothesis (Noordman & Vonk, 1998; Sanders, 2005) readers will, by default, first try to establish a causal relation. This preference for causal relations originates from a reading strategy to construct a highly connected representation. This highly connected representation is established if the reader has been able to relate events to their causes. That is, if the reader was able to establish meaningful explanations. This view on the reading process is very much in line with some of the processing theories that have been described in CHAPTER 1 (notably, the “effort after meaning principle” from the global inference theories (Graesser, Singer, & Trabasso, 1994; Singer, Graesser, & Trabasso, 1994)).

It can be argued that the advantage of a CAUSALITY-BY-DEFAULT strategy is that it minimizes the effort of establishing causal relations. Indeed, if the reader first tries to connect the events by means of a causal relation, and only considers other relations when this attempt fails, the processing of causally related information will proceed faster than the processing of non-causal information, if indeed the information is causally related. Thus, if readers adopt a strategy in which they seek to relate events to their causes, the CAUSALITY-BY-DEFAULT processing leads to more economic text processing in terms of cognitive energy. For this reason it may explain the finding that the (causal) Problem-Solution relation is processed faster than the List relation.

In the second hypothesis it is assumed that reader schematic expectations guide the processing of text in a top down way. This hypothesis will be referred to as the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis. In the specific situation of Problem-Solution relations, reading the Problem triggers the expectation that a Solution is likely to be presented in the text. This expectation triggering characteristic of Problem-Solution relations has been recognized in both text linguistics (Hoey, 1986; Sanders, 2000) and text processing (Meyer, 1985; Horowitz, 1987). Supposedly, these kinds of expectations are absent (or at least weaker) for additively connected statements. If the new information indeed relates to the previous text in the way that was expected by the reader, this new information will be processed faster than when no expectation was constructed.

The key concept in the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis is “knowledge of text structure” (cf. Van Dijk & Kintsch, 1983; Kintsch & Yarbrough, 1982); it is this knowledge that enables the reader to construct expectations about the text yet to come. If readers indeed have knowledge about text structure, it should be the case that readers have knowledge about Problem-Solution structures as well as knowledge about List structures. This implies that if the presence or absence of expectations explains processing differences, it must be true that knowledge of Problem-Solution structures enables the reader to construct expectations, whereas knowledge of List structures only does so to a lesser extent.

Thus, the question is: which differences between Problem-Solution and List structures explain why expectations are more frequently present in the former structure than in the latter?

The answer to this question is that knowledge of the Problem-Solution structure pertains to a very specific configuration, in which a negatively evaluated situation is followed by an action to take away the evaluation or the situation at hand. The key idea in the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis is that identifying the Problem-part triggers the whole
Chapter 5

**Problem-Solution** structure. That is, reading the problem part activates the **Problem-Solution** structure in the knowledge of the reader. The activated structure functions as a framework against which the representation of the text is built. For this reason, the Solution part is processed with greater ease; after all, the structure is already activated and integration of the Solution in the developing representation will be easier (cf. Van den Broek, 1990).

In the case of a **List** structure, the minimal pattern is that of a conjunction of two situations. If the first situation is identified, the reader has no clue as to what the whole structure of the text will look like. For instance, it may be that the first situation is followed by a Consequence or an Elaboration. Because there are many options for continuation, it is less likely that the **List** structure is activated as a result of identifying the first situation. For this reason, knowledge of **List** is much less likely to lead to anticipation of the information in the text than knowledge of **Problem-Solution** structures.

### 5.2.3 Conclusion from the literature

In conclusion, there is some evidence in the literature that suggests that different coherence relations result in different processing of these relations. Identical information is processed differently as the relation between the information and the context is varied. Evidence from two types of study was reviewed. First, there is evidence from studies in which the strength of the causal relation was varied. Second, we looked at studies in which the nature of the relation was varied so that they differed in type of coherence relation, such as additive (**List**) versus causal (**Problem-Solution**). Note that differences in causal strength can also be viewed as differences in relation type. This can be demonstrated by means of a paraphrase test. For instance, causal relations of the type used in the study by Keenan et al. (1984) are of the kind that can be made explicit with the lexical signaling phrase *as a result*. This is the case for the examples (1) and (2) above:

1' Joey’s brother punched him again and again. *As a result*, the next day his body was covered with bruises.
2' Racing down the hill, Joey fell of his bike. *As a result*, the next day his body was covered with bruises.

In these cases, the sentences marked with *as a result* are perfectly natural. This is not the case in example (3), but it is still possible if a bridging event is constructed of the type “so she must have hit him”:

3' Joey’s crazy mother became furiously angry with him. *As a result*, the next day his body was covered with bruises.

In the case of (4), on the other hand, the marking with *as a result* clearly leads to a *non-sequitur*; an ill-formed text, which leads one to suspect the relation really is not causal at all.
(4') Joey went to a neighbor’s house to play. As a result, the next day his body was covered with bruises.

This shows how differences in causal strength may also be viewed as a difference in relation type.

In conclusion, if exactly the same sentence is connected to the previous context with different relations, it is both processed and represented differently. Several explanations have been presented to account for the processing differences. We will argue that they essentially all boil down to three possible explanations: two bottom-up ones, in terms of the properties of the relation and a top down one, in terms of readers’ expectations.

The first explanation is the COMPLEXITY HYPOTHESIS. It concerns the relative differences in complexity between relations, and the differences in inferential processing resulting from that. Traxler et al.’s (1997) argumentative relations turned out to cost more cognitive energy than the ‘plain’ causal relations. They argued that additional processing is required for the more complex relation, which is processed slower than the less complex relation. This line of reasoning might explain the processing difference found between “simple” causal relations and argumentative relations, but it does not explain for another type of complexity we have discussed: the difference between Problem-Solution and List relations found by Sanders & Noordman (2000). They found that the more complex Problem-Solution relation was processed faster. Although we acknowledge that the notion of complexity may blur the differences between the Traxler et al. and the Sanders & Noordman studies, we can at least conclude here that the processing advantage of the Problem-Solution relation over the List relation cannot be accounted for by a COMPLEXITY hypothesis.

The second explanation is called CAUSALITY-BY-DEFAULT: when interpreting discourse, language users try to interpret relations between events in a causal way, unless they are instructed otherwise (Sanders, 2005a; 2005b). This idea might explain for the processing advantages that causal relations seem to have in all experimental studies discussed so far. Traxler et al. (1997) found that causals were processed fastest, Sanders and Noordman (2000) found how causal Problem-Solution relations were processed faster and the Myers et al. causal strength experiments may be taken as providing evidence for the very same idea: the stronger the causal relation, the faster the processing; additional processing is needed when the causal relation is weakest.

The third explanation is that the Top Down schematic expectations that influence the reading process are directly related to the differences in text structure and relation type. Experienced readers of expository text, as most participants in processing experiments are, have acquired schematic knowledge of text structure. When they read about a Problem, they may expect a Solution to come up next in the text. Under the assumption that expectations are present in the case of Problem-Solution text, but to a far lesser extent in List texts, the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis indeed explains the processing difference between Problem-Solution and List found by Sanders & Noordman (2000). It is unclear, however, to
what extent this hypothesis generalizes to other types of relation. For instance, can the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis explain differences between causal relations of different strength? Keenan et al. (1984) hypothesize that the processing advantage of the strong causal relations originated from reader expectations. This might be the case in example (1) where reference to repeated punching might lead the reader to expect that the one that gets punched will be bruised (see also Linderholm, 2001; Van den Broek, 1990; 1994). This means that the target event will be activated prior to its presentation, and hence interpretation of the target sentence is facilitated. However, as we shall see in the next section, this kind of content expectations are not the kind of expectations our specific TOP-DOWN hypothesis refers to (see also section 6.6).

Both the CAUSALITY-BY-DEFAULT and the SCHEMATIC STRUCTURAL EXPECTATIONS hypotheses account for the results found by Sanders and Noordman (2000). By contrast, the COMPLEXITY hypothesis cannot. For this reason, the COMPLEXITY hypothesis will not be pursued further. In the remainder of this chapter, two experiments will be presented in which these two hypotheses are further tested. Before these experiments are described in more detail, we show how the two hypotheses can be separated empirically.

5.3 The causality-by-default and schematic structural expectations hypotheses compared

In Sanders and Noordman (2000) the CAUSALITY-BY-DEFAULT and the SCHEMATIC STRUCTURAL EXPECTATIONS hypotheses are presented as alternative explanations. A possibility that is not considered by the authors is that both hypotheses might be simultaneously true. It is possible that CAUSALITY-BY-DEFAULT and SCHEMATIC STRUCTURAL EXPECTATIONS exert their influence at the same time. In that case, the reader both benefits from the relation being causal and from the strong expectations elicited by the context. In terms of processing, causally related information is processed faster because of the causal nature of the relation and, if this information was also expected (as it would be in the case of a Problem-Solution relation), an additional processing advantage might result.

The SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis and the CAUSALITY-BY-DEFAULT hypothesis both account for the difference between the processing of Problem-Solution relations and List relations. But in terms of the predictions that can be based upon them, the two hypotheses are quite distinct. The SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis predicts that contexts that elicit strong expectations will lead to a processing advantage if the information in the text matches this expectation. The CAUSALITY-BY-DEFAULT hypothesis focuses on the nature of the relation itself: causal relations are processed faster than non-causal relations.

In terms of experimental design, evidence for the possibility that both hypotheses are true can be obtained by comparing three conditions. In the first condition, information is presented that is both expected from and causally related to the previous context. In the second condition, the same information would be presented, but in this case it is not expected from the context, but ‘only’ causally related to it. In the third condition, the information is both not expected...
The Processing Advantage of Causal Coherence Relations

and non-causally related. The differences between the conditions are summarized in Table 5.1.

**Table 5.1** Conditions needed to distinguish between the SCHEMATIC STRUCTURAL EXPECTATIONS and CAUSALITY-BY-DEFAULT hypotheses.

<table>
<thead>
<tr>
<th>Causal</th>
<th>Expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition 1</td>
<td>+</td>
</tr>
<tr>
<td>Condition 2</td>
<td>+</td>
</tr>
<tr>
<td>Condition 3</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: a '+' indicates presence of a characteristic, '-' the absence of it.

The CAUSALITY-BY-DEFAULT hypothesis predicts that the textual information in the causal conditions will be processed faster than that in the non-causal conditions. Thus, conditions 1 and 2 are processed faster than condition 3, since in conditions 1 and 2 the information is causally related to the context, whereas in condition 3 it is not. Note that on the basis of the CAUSALITY-BY-DEFAULT no prediction can be made with respect to the difference between the two causal conditions, since the hypothesis only concerns the difference between causal and non-causal relations. On the basis of the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis, it is expected that condition 1 is processed faster than conditions 2 and 3. The reason for this is that in condition 1 the information is expected on the basis of the context, whereas in conditions 2 and 3 it is not. If both the CAUSALITY-BY-DEFAULT and the hypothesis is true, condition 1 will be processed faster than condition 2, because condition 1 triggers expectations whereas condition 2 does not, and condition 2 will in its turn be processed faster than condition 3, because condition 2 is causal and condition 3 is not.

The predictions of the hypotheses are summarized in Table 5.2.

**Table 5.2** Predicted reading time patterns according to the SCHEMATIC STRUCTURAL EXPECTATIONS and CAUSALITY-BY-DEFAULT hypotheses.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCHEMATIC STRUCTURAL EXPECTATIONS and CAUSALITY-BY-DEFAULT</td>
<td>Condition 1 &lt; Condition 2 &lt; Condition 3</td>
</tr>
<tr>
<td>CAUSALITY-BY-DEFAULT</td>
<td>Condition 1 &lt; Condition 3; Condition 2 &lt; Condition 3</td>
</tr>
<tr>
<td>SCHEMATIC STRUCTURAL EXPECTATIONS</td>
<td>Condition 1 &lt; Condition 2; Condition 1 &lt; Condition 3</td>
</tr>
</tbody>
</table>

Note: < indicates faster processing of identical information in terms of reading times.

Thus, with three conditions that differ from each other in the nature of the relation (causal versus additive) and the presence of reader expectations, the schematic structural expectations and causality-by-default hypotheses can be tested empirically. How can these conditions be translated into textual materials?

Two of the conditions were already used in the Sanders & Noordman (2000) experiment. Problem-Solution structured texts fulfill the demands of a condition in which the relation is
causal and the reader has high expectations, whereas the List texts are non-causal texts without these strong expectations. In addition, a condition is needed in which the relation is causal and the reader has low expectations. A Cause-Consequence text may serve this purpose. In a text organized like that a situation is described, this situation has a consequence, but this consequence is not likely to be anticipated by the reader.

As an example of the difference between Problem-Solution (7), Cause-Consequence (8) and List (8), consider the following texts.

(7) The Main Street in Maasbracht has been a popular traffic route for years. Not surprisingly, local residents have filed a lot of complaints about discomfort brought about by stench and noise. Also, heavy traffic in the Main Street has led to numerous dangerous situations. According to the City council this situation has gone too far. As of December first, the Main Street will be closed for motor vehicles. The residents have reacted positively on this news.

De hoofdweg in Maasbracht is jarenlang een populaire sluiproute geweest. Het is dan ook niet verwonderlijk dat de bewoners regelmatig hebben geklaagd over stank- en geluidsoverlast. Het drukke verkeer in de hoofdweg leidde bovendien tot tal van gevaarlijke situaties. Voor de gemeente is de maat nu vol. De hoofdweg zal vanaf 1 december gesloten zijn voor motorvoertuigen. De bewoners hebben verheugd gereageerd op het nieuws.

(8) The City Council of Maasbracht announced yesterday that the sewer system under the Main Street will be renewed. The work will start at the end of this year. Local residents of the Main Street will have to reckon with worse accessibility of their houses. Because of the sewer work the street will have to be partly broken open. As of December first, the Main Street will be closed for motor vehicles. BicycLists and Pedestrians will still be able to use the road.

De gemeente Maasbracht maakte gisteren bekend de riolering onder de hoofdweg te zullen vernieuwen. De werkzaamheden zullen aan eind van dit jaar beginnen. Bewoners van de hoofdweg moeten rekening houden met een slechtere bereikbaarheid van hun woning. Vanwege de rioolwerkzaamheden zal de straat gedeeltelijk moeten worden opengebroken. De hoofdweg zal vanaf 1 december gesloten zijn voor motorvoertuigen. Fietsers en wandelaars kunnen gewoon van de weg gebruik blijven maken.

(9) The City Council of Maasbracht will take drastic measures on the area surrounding the Main Street. This is part of a plan to improve the infrastructure of the city. The City Council will be performing several public works around the Main Street. An additional parking garage will be opened in November. As of December first, the Main Street will be closed for motor vehicles. Also, a new bicycle path will be constructed in the direction of the Handelskade.

De gemeente Maasbracht zal de omgeving van de hoofdweg stevig onder handen gaan nemen. Dat staat in een plan ter verbetering van de infrastructuur van de gemeente. De gemeente zal verschillende werkzaamheden verrichten rond de hoofdweg. Er zal een extra parkeergarage worden geopend in
The Processing Advantage of Causal Coherence Relations 109

In the Problem-Solution text (7), closing of the Main Street is the solution to the problems described in the context (stench and noise, dangerous situations). The closing of the Main Street is a volitional action that is a consequence of the problematic situation reported earlier. This text may be analyzed as a Problem-Solution text, because the action (the closing of the Main Street) is performed with the intention to solve the problematic situation. According to the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis, the description of the problem triggers expectations about the text yet to come. Such expectations are likely to be absent in the Cause-Consequence text (8). Here, closing of the Main Street for motor vehicles is a consequence of the fact that the street has to be broken open. This is only one of many possible continuations. Other plausible continuations are, for instance, “the work will start tomorrow”, or “this is bad news for the neighborhood”. Crucially, some of the many possible continuations are connected to the context by means of non-causal relations. This means that because of the fact that the context is not very restrictive, it is not likely that readers expect a continuation that will be connected to the context by means of a Cause-Consequence relation. Thus, according to the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis, there should be a processing difference between the information in the Problem-Solution text and the information in the Cause-Consequence text. In the List text (9), “the closing of the Main Street” is part of an enumeration of public works that are being planned in the city. For the same reasons given in the case of the Cause-Consequence text, it is not likely that readers expect a specific type of relation to be expressed in the remainder of the text. According to the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis, then, the information in the List text will be processed slower than in the Problem-Solution (PS) text. The CAUSALITY-BY-DEFAULT hypothesis, on the other hand, predicts that the List text (LI) will be processed slower in comparison to both the Problem-Solution text and the Cause-Consequence text (CC), because the List is non-causal and the other two are causal relations.

In Table 5.3, the hypotheses and the predictions that can be based upon them are presented in terms of the different relation types.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCHEMATIC STRUCTURAL EXPECTATIONS and</td>
<td>PS &lt; CC &lt; LI</td>
</tr>
<tr>
<td>CAUSALITY-BY-DEFAULT</td>
<td></td>
</tr>
<tr>
<td>CAUSALITY-BY-DEFAULT</td>
<td>PS &lt; LI; CC &lt; LI</td>
</tr>
<tr>
<td>SCHEMATIC STRUCTURAL EXPECTATIONS</td>
<td>PS &lt; CC; PS &lt; LI</td>
</tr>
</tbody>
</table>

Note: < indicates predicted faster processing; PS=Problem-Solution, CC=Cause-Consequence, LI=List.
The hypotheses presented in Table 5.3 and their underlying assumptions were tested in a continuation experiment and in a reading experiment. In the reading experiment, two different tasks were used. First, the participants read short newspaper articles, during which sentence reading times were collected. Each text contained a target sentence that was presented in one of three conditions: Problem-Solution, Cause-Consequence and List (see examples (7) through (9)). Second, a sentence verification task was administered immediately after each text was read. During this task, the participants had to verify a paraphrased version of the target sentence. Both verification latencies and accuracy was registered. Thus, the same experimental paradigm as in Sanders and Noordman (2000) was used, except that in this new experiment no free recall task was administered.

With respect to the reading times, the expectations are in line with the hypotheses presented in Table 5.3. According to the causality-by-default hypothesis, reading times are longer in the List condition, as opposed to the Problem-Solution and Cause-Consequence condition. The SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis predicts that reading times are shorter in the Problem-Solution condition than in the Cause-Consequence or List conditions. If both hypotheses are correct, Problem-Solution will be read faster than Cause-Consequence and the information in the Cause-Consequence will be read faster than when it is part of a List.

As for the verification times and accuracy, it is expected that reactions are faster and more accurate in the causal conditions than in the List condition. This would confirm the representational advantage found for causal relations in earlier studies (Black & Bern, 1981; Bradshaw & Anderson, 1982; Myers, O’Brien, Balota, & Toyofuku, 1984; Sanders and Noordman, 2000; Trabasso & Sperry, 1985; Trabasso & Van den Broek, 1985).

The assumptions underlying the CAUSALITY-BY-DEFAULT and SCHEMATIC STRUCTURAL EXPECTATIONS and their predictions in terms of reading times were tested in two experiments: a continuation experiment and a reading experiment. In the continuation experiment crucial assumptions underlying the two hypotheses were tested. These assumptions are that Problem-Solution texts indeed elicit more and stronger expectations than Cause-Consequence and List texts, as is predicted by the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis, and that there is a general preference for causal relations, which is in line with the CAUSALITY-BY-DEFAULT hypothesis. In the reading time experiment, the on-line predictions based on the two hypotheses were tested. The results of the continuation experiment will be described first.
5.4 Continuation Experiment

In the continuation experiment, each participant read several contexts. Participants were asked to continue the contexts with one or two sentences. With this continuation task we tested the idea that contexts differed with respect to expectations readers have of the text structure that follows. The continuation experiment was designed so that the expectations elicited by the three coherence relations under investigation could be compared. Recall that the Problem-Solution texts were expected to generate more schematic structural expectations than the other two relations.

The reasoning behind the task is as follows. The structural expectations each context elicits should be reflected in the number of different continuations the participants present. That is, if expectations are strong, there will be less variation in continuations than when the expectations are not so strong. One would predict participants to produce a smaller number of different relation types between the context and the continuations in the case of strong expectation contexts, such as Problem context. In a Problem context, for instance, one would expect participants to produce a high number of Solutions and other volitional actions, whereas hardly any continuations are produced that should be related in another way to the context. In the low expectation contexts, there will be more types of relations presented between the continuations and the previous contexts.

Finally, the experimental set-up allows us to investigate the hypothesized preference for causal relations under the CAUSALITY-BY-DEFAULT hypothesis. If this preference is correct, most of the continuations should be causally related to the context.

5.4.1 Method

Materials

The materials were constructed for the reading experiment described later. In that experiment, twenty-four target sentences are embedded in the three different contexts. Since each combination of target sentence (item) and context constitutes a text, a total of 72 texts were constructed. An example of three of these texts was presented earlier in this chapter. Each text was a short news paper article that was minimally 5, and maximally 7 sentences long. The texts were constructed by a team of researchers and advanced students in text linguistics and text processing. Special attention was paid to maintain similarities between conditions. In particular, differences in semantic and lexical overlap between the conditions were avoided (Sanders & Noordman, 2000).

The following fragments illustrate the experimental texts that were presented to the participants.
The Main Street in Maasbracht has been a popular traffic route for years. Not surprisingly, local residents have filed a lot of complaints about discomfort brought about by stench and noise. Also, heavy traffic in the Main Street has led to numerous dangerous situations. According to the City council this situation has gone too far.

The City Council of Maasbracht announced yesterday that the sewer system under the Main Street will be renewed. The work will start at the end of this year. Local residents of the Main Street will have to reckon with worse accessibility of their houses. Because of the sewer work the street will have to be partly broken open.

The City Council of Maasbracht will take drastic measures on the area surrounding the Main Street. This is part of a plan to improve the infrastructure of the city. The City Council will be performing several public works around the Main Street. An additional parking garage will be opened in November.

The text in (10) is an example of a Problem-context, (11) is a Cause-context, and (12) illustrates a context in the List condition.

Participants

18 participants took part in the experiment. They were undergraduate students from Utrecht University. Participants were paid for their participation.

Design and Procedure

For each scenario, such as “The Main Street in Maasbracht”-scenario illustrated in (11)-(12), there were three experimental conditions. Each participant read only one of the conditions for each scenario, defined as the three structures we are interested in: Problem-Solution, Cause-Consequence and List. For instance, one third of the participants would read the Problem context, another third the Cause context and one third the List context. Each participant was presented with 8 contexts per condition, making a total of 24 texts that each participant continued. For each participant group, the design was a nested design with contexts nested under conditions. Participants were instructed to read the texts carefully and continue with a maximum of two sentences. The subjects completed the task in their own pace. Completing the whole task ranged from 45 minutes to one hour.

Analysis of protocols

Each of the continuation protocols was analyzed in terms of the coherence relation that existed between the continuation and the context. The analyses were done by two judges, with basic knowledge of text linguistics. The protocols were analyzed independently by the judges, and their analyses were compared. Differences in judgment were resolved by discussion. The basic categories were analyzed by making use of the taxonomy of Sanders, Spooren, and
Noordman (1992; 1993). Rhetorical Structure Theory (Mann & Thompson, 1988) was used to describe the relationships in more detail. Each of the 72 contexts received a score from 1 to 11, where 1 is strong expectation and 11 weak expectation measured on an ordinal scale. The rationale behind the scoring is based on the possible patterns in the continuation protocols. These patterns range from 6 times the same type of relation (strong expectations), or 6 times a different relation (weak expectations); if all of the participants produce the same type of relation for a given context, this context is given a high expectation score (=1), if every participant produces a different type of relation the context is given a low expectation score (=11).

The possible patterns and their respective scores are presented in Table 5.4.

**Table 5.4 Scoring for continuation patterns.**

<table>
<thead>
<tr>
<th># different relation types</th>
<th>Highest frequency</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>11</td>
</tr>
</tbody>
</table>

The scoring is based on two principles. The first principle is the total number of relation types produced. The higher the number of different relation types, the higher the score is. Here, the lowest rank (11) signifies the weakest expectations. Note that each context was presented to six participants. This means that the maximum number of different relation types produced per context is six. If that is the case, each participant produced a different relation. Under such circumstances, the context clearly does not trigger a specific type of relation, and hence it is assumed that the context triggers only weak expectations. In Table 5.4, the situation in which six different relations are produced is presented in the bottom row. According to the table, the score associated with that pattern equals 11.

The minimal number of relation types produced is one. In that situation each participant produced the same type of relation. If every participant produced a Solution to the problem stated in the context, the context apparently biases towards a specific type of relation, and it is therefore assumed that that specific context triggers strong expectations. According to Table 5.4, a score of 1 is assigned to contexts in which the participants all produced the same relation.

The second principle is the spreading of the scores. It is necessary to take this spreading into account, because the same amount of different relation types may reflect different continuation patterns. For instance, suppose that for a given context two different relation
types are produced, e.g. Problem-Solution and Elaboration. It might be the case that out of the 6 relations produced for that context 5 Solutions were produced and 1 Elaboration, but it might just as well be the case that both relations were produced by 3 participants. The frequency distribution of the relations produced in each context is taken into account in the scoring procedure. According to Table 5.4, if the number of different relations types produced for a given context is 2, and one relation occurs 5 times and the other relation 1 time, the context receives a score of 2, because the highest frequency (i.e. the frequency of the most occurring relation) is 5. By contrast, a score of 4 is assigned to a context in which two relations types are produced and each of the two relations is produced 3 times, since the highest frequency equals 3. The former receives a higher rank, because the continuations are somewhat more centered on one type of relation.

In sum, the contexts are given an expectation score that ranges from 1 (high) to 11 (low). A high expectation score means that the context triggers uniform continuations. This is reflected in a low number of different relation types or, stated differently, a high frequency of a particular relation type. A low expectation score indicates that the context triggers a higher number of different relation types.

To illustrate the scoring procedure, consider the following analysis of two of the continuations produced after the subjects read the cause-context of the “Main Street in Maasbracht”-text (example (13)).

(13) The City Council of Maasbracht announced yesterday that the sewer system under the Main Street will be renewed. The work will start at the end of this year. Local residents of the Main Street will have to reckon with worse accessibility of their houses. Because of the sewer work the street will have to be partly broken open.

a. This will take at least four months.
   Dit zal minstens vier maanden duren.

The first continuation can be analyzed as an Elaboration. The participants elaborates on the time span of either the renewal of the sewer system as a whole, the worse accessibility, or the breaking up of the street.

A second continuation is the following.

b. Local residents will have to reckon with some traffic annoyances.
   Bewoners moeten dan ook rekening houden met enige verkeershinder.

This can be analyzed as a Consequence. The participant describes a consequence of the street being broken open.

In total, two different relations types were produced for this context: Elaboration and Cause-Consequence. Elaborations occurred 4 times and Cause-Consequences 2 times. In terms of Table 5.4, the number of different relations is 2 and the highest frequency is 4. Therefore, this context receives a score of 3.
5.4.2 Results of the continuation experiment

The results of the continuation experiment are presented in two sections. The first section focuses on the assumptions of structural expectations. The second section focuses on the assumption of a preference for causal relations.

Structural Expectations

Because of the ordinal nature of the data, they were analyzed with a Kruskall-Wallis test. The analysis showed a significant effect of condition ($\chi^2(2) = 6.47, p = .039$). Since the mean rank of the Problem-contexts (27.7) was lower than the mean rank in the cause (39.5) and List conditions (42.3), the analysis confirms that the problem contexts elicit more uniform continuations. Thus, the continuation task shows that the problem-contexts lead to stronger expectations than the other contexts.

A few general observations can be made on the basis of the data. The first observation is that there is a large amount of variation within each of the conditions. This becomes apparent if a cross-tabulation of condition and score is considered.

Table 5.5 Distribution of expectation scores per condition (scores in the table are frequencies; $N=72$)

<table>
<thead>
<tr>
<th>Expectation Score</th>
<th>Problem</th>
<th>Cause</th>
<th>List</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (high)</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>-</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>11 (low)</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>72</td>
</tr>
</tbody>
</table>

The pattern of results that emerges from this cross-tabulation is, that it is by no means true that all the Problem contexts elicit high expectations. A total of 9 contexts have a score of 1,2 or 3, which may be regarded as a scores associated with high expectations. Also, 8 contexts may be considered to elicit only weak expectations. Thus, the amount of problem-contexts with high expectations is almost equal to the number of problem-contexts with low
expectations. This means that overall, the problematic situation itself does not guarantee high expectations.

Still, there is a difference in the distributions of the scores if we compare the problem-contexts with the other types of contexts. This difference is situated in the high expectation range. If we look at the range from 1 to 4, there are 9 observations in the Problem contexts, but only 4 in the Cause contexts and 2 in the Lists contexts. The magnitude of this difference ranges from 5 to 7. This same difference is found in the range of the low expectation scores. For the Problem contexts there are 8 observations, but in the Cause and List contexts there are 13 occurrences. For the middle range, the distributions within the three conditions are more equal. Thus, on average, the data confirm our assumption that Problem contexts elicit stronger expectations than the other two types of contexts. This is mirrored in the larger frequency of high expectation scores and lower frequency of low expectation scores.

As a further test of the assumptions, a qualitative analysis was performed on those contexts that had received high expectation scores (scores 1 through 4). The reason for performing this additional analysis is that the quantitative analysis does not consider the type of relation produced per se. The scoring is based on the number of different continuations in terms of relation types. This means that if a Problem context elicits only Solution continuations a score of 1 would be assigned to this context, indicating strong expectations, but when only Elaborations are produced, the context would also be given a score of 1. The question is, then, whether eliciting Elaborations, is really an instance of eliciting strong expectations. We might also argue that this is really an instance of weak expectations, since elaborations are considered as one of the weaker organizing types of relation (Meyer & Freedle, 1984; Sanders, Spooren, & Noordman, 1992).

There are 9 Problem contexts that received a high expectation score. As each context was presented to 6 participants, the total number of relations produced equals 54. The qualitative analysis showed that in 46 out of the 54 relations the continuation provided a Solution to the Problem stated in the context. A total of 4 Cause contexts received a high expectation score. In 20 out of the 24 relations produced in these contexts the continuations were consequences of the situation described in the context, these were either of a volitional (11 cases) or non-volitional (9 cases) nature. In the List condition, 2 contexts received a high expectation score. From the 12 relations produced, 9 were List continuations.

Thus, with respect to the Problem condition, the results are in line with the assumption that the description of the Problem elicits the expectation of a Solution. Most importantly, the high expectation scores are not an artifact of the scoring procedure. Interestingly, although there are only few contexts in the Cause and List conditions that received a high expectation score, these contexts show that they might also trigger structural expectations.

Preference for causality

The second analysis focused on the nature of the relation that was produced. According to the causality-by-default hypothesis, there is a preference for detecting causal relations. It may be expected that this preference will be reflected in the continuations: the frequency of causal relations produced will be higher than the frequency of other relations.
In terms of the nature of the relation, a total of six different relations has been observed. In Table 5.6, the results are presented.

**Table 5.6** Nature of the relations produced per condition (Scores are percentages, N= 430).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Causal</th>
<th>Contrastive</th>
<th>Elaborative</th>
<th>Additive</th>
<th>Temporal</th>
<th>Restatement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem (N=144)</td>
<td>76.4</td>
<td>2.8</td>
<td>19.4</td>
<td>0.7</td>
<td>0.0</td>
<td>0.7</td>
</tr>
<tr>
<td>Cause (N=143)</td>
<td>58.0</td>
<td>3.5</td>
<td>32.9</td>
<td>4.2</td>
<td>1.4</td>
<td>0</td>
</tr>
<tr>
<td>List (N=143)</td>
<td>31.5</td>
<td>3.5</td>
<td>25.2</td>
<td>39.2</td>
<td>0.7</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>55.3</td>
<td>3.3</td>
<td>25.8</td>
<td>14.7</td>
<td>0.7</td>
<td>0.2</td>
</tr>
</tbody>
</table>

The first thing to notice from the table is the strong preference for causal relations. A total of 55.3% of the continuations is causally related to the context. The difference in frequencies of the causal relations as opposed to the other relations is significant ($z = 4.8, p <.001$)\(^1\). Within the Problem condition the preference for causal relations is most prominent; here too, the frequency of causal relations is significantly different from the frequencies of other relations ($z = 9.4, p <.001$). In the Cause-condition there is also a preference for causal relations ($z = 5.8, p <.001$). In the List condition the frequencies are more equally distributed.

A preference for causality cannot be demonstrated in this condition ($z = .219, p = .41$).

5.4.3 Conclusion

The continuation experiment clearly indicates that the Problem-contexts elicit stronger expectations than the other two types of context. In the Problem condition, the frequency of high expectation scores is higher and the frequency of low expectation scores is lower than in the other conditions. Therefore, the continuations were more uniform in the problem condition than in the other two conditions. This leads to the conclusion that the problem contexts that were constructed indeed lead to stronger expectations than Cause and List contexts.

Furthermore, the continuation protocols show evidence in favor of the preference for causal relations that was hypothesized. Overall, the frequency of causal relations is higher than the frequency of other types of relations.

In general, then, it may be concluded that the pattern of results found in the continuation experiment is in line with both the SCHEMATIC STRUCTURAL EXPECTATIONS and the CAUSALITY-BY-DEFAULT hypothesis. This means that the construction of the materials is successful to a large degree, even though there are still a few Problem-contexts that do not behave like expected, since there is a large amount of variance between texts. This implies

---

\(^{10}\) Because of the low frequencies across conditions, the temporal and restatement relations are not included in the statistical analyses.

\(^{11}\) The statistical analysis proposed by Van den Bergh (1990) was performed.
that some texts may be regarded as less prototypical examples than others. Still, in the reading time experiment, texts in the whole range (from not-prototypical to prototypical) were included.

5.5 The reading experiment

In the reading experiment the texts were presented on a computer screen in a self paced reading paradigm. After each text a statement verification task was administered in which the participants had to verify the target information.

5.5.1 Method

Materials

As was explained before, the texts were developed by a team of researchers and advanced students in the field of text linguistics. Special care was taken to prevent other properties of the text from varying systematically with factors manipulated in the experiment (Sanders & Noordman, 2000). The information in the target sentences was equally given across conditions (Chafe, 1976; Haviland & Clark, 1974). Therefore, there were no differences in the availability and or accessibility of concepts in the target sentences that might otherwise explain differences in the processing times between conditions.

During the verification task, the information in the target sentence was presented again. A difference between the current study and the experiment by Sanders and Noordman (2000) is that the target information was paraphrased in stead of presented literally. In this way, the verification task was less superficial; if the statement is presented literally, it is possible that differences between conditions can be attributed to the surface code representation see chapters 1 and 4) only. This would result in a very superficial assessment of differences in the quality of the representation across conditions.

Design and Procedure

The materials were presented in a Latin Square confounded factorial design (Kirk, 1982). This means that each participant was presented with items in all conditions, and that all items were presented in all three conditions. Each subject read 8 items per condition. The order of the items was randomized per participant. Besides the experimental texts, 24 fillers were presented. These were minimally edited short news articles found on the internet.

The first statement after the experimental texts were always correct, so the subject had to press the button indicating ‘yes’. The first statement after the filler texts was always incorrect. In total, 50 percent of the statements was correct, and 50 percent incorrect.

The participants were instructed to read the texts as normal a possible (like reading a newspaper). They were explicitly instructed not to memorize the texts and to react as fast a
possible in the verification task. They were also told to judge the truth of the statement on the basis of the information in the text.

After this instruction, a practice session was run with two texts. On the basis of this practice session, the experimenter gave a few more directions to the subjects, and there was also an opportunity for the subjects to ask questions regarding the procedure.

The texts were presented sentence by sentence in a self-paced reading paradigm. The presentation of the materials was controlled by the program E-prime. The sentences were vertically centered on the screen. The participants pressed the space-bar after a sentence was read, after which a new sentence appeared.

Each text was followed by a warning that the statements were about to be presented, and on the same screen, the subjects were reminded to keep their fingers on the ‘F’ and ‘J’ buttons. These buttons were marked red (‘F’) and green (‘J’). Each text was followed by two statements.

The average time to read all of the 48 texts was 20 minutes.

Participants

33 participants took part in the experiment. They were undergraduate students from the Faculty of Arts of Utrecht University. 12 participants were male and 21 were female. Participants were paid for their participation.

5.5.2 Results

Reading Times

Prior to the analysis, the data were screened for outliers using the procedure outlined in chapter 3. This resulted in the detection of 4 outliers. These outlying scores were entered in the model as fixed parameters.

In Table 5.7 the mean reading times per condition are presented.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Reading time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem-Solution</td>
<td>2.49 (.168)</td>
</tr>
<tr>
<td>Cause-Consequence</td>
<td>2.72 (.168)</td>
</tr>
<tr>
<td>List</td>
<td>2.73 (.168)</td>
</tr>
</tbody>
</table>

Variance Estimates

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Between items</td>
<td>.493 (.149)</td>
</tr>
<tr>
<td>Between participants</td>
<td>.202 (.054)</td>
</tr>
<tr>
<td>Between individual measurements</td>
<td>.718 (.035)</td>
</tr>
</tbody>
</table>
Table 5.7 shows that the reading times in the Problem-Solution condition is shorter than in the other conditions. The significance test showed that this difference is significant ($\chi^2(1) = 15.5, p < .001$). There is no difference between Cause-Consequence and List ($\chi^2(1) < 1$). Hence, the results indicate faster processing in the Problem-Solution condition as opposed to the Cause-Consequence and List conditions.

**Statement Verification**

The analysis of the statements focused on proportion correct and reaction time. The reaction times were checked for outliers. This resulted in the detection of two outliers. Prior to the analysis of the reaction times, these scores were normalized, in order to meet the assumptions of multi-level modeling. The reaction times were modeled with the outlying scores included in the fixed part. The proportion correct is modeled with a binary logistic model. The results of the verification task are presented in Table 5.8.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Reaction Time</th>
<th>Proportion Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem-Solution</td>
<td>2.590 (0.187)</td>
<td>.89</td>
</tr>
<tr>
<td>Cause-Consequence</td>
<td>2.699 (0.187)</td>
<td>.88</td>
</tr>
<tr>
<td>List</td>
<td>2.817 (0.187)</td>
<td>.82</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between items</td>
</tr>
<tr>
<td>Between participants</td>
</tr>
<tr>
<td>Between individual measurements</td>
</tr>
</tbody>
</table>

As we can see in Table 5.8, the proportion correct is high in all conditions. Nevertheless, the proportion correct is higher in the Problem-Solution and Cause Conditions than in the List conditions ($\chi^2(1) = 6.16, p < .025$). There is no significant difference between Problem-Solution and Cause-Consequence ($\chi^2(1) = .281$).

The pattern of reaction times is comparable to that of the proportion correct. The reaction times are shorter in the Problem-Solution and Cause-Consequence condition than in the List condition ($\chi^2(1) = 6.13, p < 0.025$). There is no difference between Problem-Solution and Cause-Consequence ($\chi^2(1) = 1.85, p = 0.17$). These reaction time results are irrespective of the correctness of the statement verification. Although participants reacted slower to statements that were incorrectly verified ($\chi^2(1) = 31.7, p < .001$), the differences between the conditions remain the same after incorporating the reaction time difference between correctly and incorrectly verified statements in the model (Problem-Solution and Cause-Consequence versus List: $\chi^2(1) = 3.89, p < 0.05$; Problem-Solution versus Cause-Consequence: $\chi^2(1) = 1.61, p = 0.20$).

The variance estimates show that, in contrast to the reading times, participant variability is more important than item variability. That is, the random variance between participants is
higher than the variance between items. Apparently item characteristics are far more important during reading than after reading. For the verification task individual differences between participants (which may involve issues like memory capacity and reaction speed) are more important than characteristics of the text.

Data exploration

A preliminary conclusion is that the difference in reading times between the conditions is in line with the STRUCTURAL EXPECTATIONS hypothesis, as the reading times in the condition with strong expectations (Problem-Solution) are shorter than in the conditions were only weak expectations are expected. In order to directly explore the relationship between structural expectations and reading times, a preliminary analysis was conducted with the expectation scores from the continuation experiment. Although this analysis did not show significant results ($\chi^2(3) = 6.0, p = .11$), it did indicate that even List relations, which were characterized as the relations about which readers have the least structural expectations, are processed faster in the exceptional case (2 out of 24 contexts) that they do trigger clear expectations. In these cases the List relations were processed faster.

5.5.3 Conclusion

The most important result of this experiment is that information that is connected to the context by means of a Problem-Solution relation is processed faster than when it is connected to the context by means of a Cause-Consequence relation or a List relation. This pattern of results was expected under the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis. Indeed, according to this hypothesis, Problem-Solution structure generates more structural expectations than the other two do. For this reason, a sentence connected to the previous context in a Problem-Solution relation is easier to process than the identical sentence connected in a Cause-Consequence or List relation.

It may be concluded that this experiment provides evidence for the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis. This implies that the experiment shows that readers construct expectations about the structure of the text yet to come. Additional support for this conclusion was obtained in the continuation experiment: the Problem context elicited more uniform continuations than the Cause and List contexts.

Interestingly, the on-line processing difference between Problem-Solution and Cause-Consequence does not translate to a difference in the representation. Neither verification latencies nor the accuracy differs between Problem-Solution and Cause-Consequence. There is, however, a difference in latency and accuracy between these two causal conditions on the one hand and the List condition on the other hand. This result confirms the special status of causal relations in the representation, as has been repeatedly shown in the literature.
5.6 General Discussion

The experiments reported in this chapter were designed to further investigate the special status of causal coherence relations in text processing and memory. As was observed in the beginning of this chapter, information connected to the context by means of a causal relation is both remembered better and processed faster than when it is connected by means of an additive relation. In the majority of studies, the special status of causal relations was demonstrated for narrative text. The results of the present study replicate the findings of Sanders and Noordman (2000) who showed that the special role of causal relations also extends to the processing and representation of expository text.

With respect to the cognitive representation, this study indicates that the causal nature of the relation increases its quality, and as a result information connected to the context by means of a causal relation is remembered better than when the same information is part of an additive structure. The Problem-Solution and Cause-Consequence relations did not differ in terms of verification latency and accuracy, but the statements were processed more quickly and more accurately in these conditions than in the List condition. This confirms the more strongly organizing quality of causal relations in the mental representation readers construct on the basis of the text.

Yet, the main focus of this study was on the on-line processing of causal relations. To be more precise, the experiments were primarily designed to differentiate between two explanations for the processing advantage of Problem-Solution versus List relations found by Sanders and Noordman (2000). These two explanations were the SCHEMATIC STRUCTURAL EXPECTATIONS and CAUSALITY BY DEFAULT hypotheses. According to the first hypothesis, the processing advantage can be explained by the text-structural expectations readers have about the remaining text. If the information in the text fits in with these expectations, the processing of the information is easiest. The CAUSALITY-BY-DEFAULT hypothesis on the other hand, explains the processing advantage in terms of the nature of the relation: readers try, by default, to establish a causal relation and will only construct a non-causal relation if a causal relation is not possible.

The results of the reading time experiment are in line with the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis. The Problem-Solution relations are processed faster than the other two relations. Since the difference between the Problem-Solution condition and the Cause-Consequence and List conditions lies in the presence of expectations, the results provide support for the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis. The fact that Problem contexts elicit stronger expectations than Cause-Consequence and List was confirmed in the continuation experiment.

Hence, the results of Sanders and Noordman (2000) were replicated in this experiment. The Problem-Solution texts were read faster than the List relations and the performance on the verification task was better in the Problem-Solution condition. In addition, we have gained much new information. We are now able to answer a crucial remaining question: how can this processing difference be explained? The answer to that question is that the processing
difference results from the fact that the context elicits stronger expectations in the Problem-Solution condition than in the other conditions.

Are there any possible counter explanations for our findings? One could be that the context triggers the activation of certain concepts more strongly in one condition than in the other, as could, for instance, be expected on the basis of accessibility theory (Ariel, 2001; Sanders & Gernsbacher, 2004). If the information in the target sentence contains these concepts, we expect a processing advantage to arise as opposed to a situation where the new information does not contain such concepts. This explanation is only problematic for the SCHEMATIC STRUCTURAL EXPECTATIONS account if the distribution of concepts is unequal across conditions. That is, only if there is systematic variation in concepts between conditions (both in terms of frequency and accessibility), processing differences found between conditions may be attributed to content expectations.

There are two reasons why it is unlikely that this is the case in the present study. The first is that all materials were controlled for differences in the “newness” of the concepts. As was explained in the materials section, the lexical and semantic overlap between concepts in the target sentences and the context were equal across conditions. The second reason is that across experimental texts the contents of ‘scenarios’ were varied. That is, the concrete topics and themes in the different scenarios were very diverse. This makes it unlikely that, in this experiment, there was systematic variation in the concepts between conditions.

A second alternative explanation for the current findings focuses on content expectations regarding causal consequences. Strong causal relations have strong restrictions on the consequences that may follow in the text. Therefore, it is easier to predict what the content of these text fragments may be. For instance, reading about a large car crash might trigger the expectation that cars will be damaged. In general, predictive inferences are likely to be made when there are only few alternatives that need to be considered by the reader on the basis of the context (Calvo & Castillo, 2001; Graesser, Swamer, Baggett, & Sell, 1994; Linderholm, 2001). With respect to causal relations, there is evidence that a high degree of causal sufficiency (Van den Broek, 1990; see also chapter 2) is one of the restrictions the context has to meet for it to be able to elicit expectations about causal consequences (Linderholm, 2001). In the case of strong causal relations there are only few alternative plausible consequences and (therefore) the cause is sufficient for the consequence to occur. Thus, all conditions seem to be met that enable the construction of predictive inferences.

The crucial difference between the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis and such a predictive inference account is that, according to the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis, the expectations concern text structure, whereas the predictive inference account assumes that specific content is expected. Thus, adopting content rather than structural expectations would account for the processing difference between Problem-Solution and Cause-Consequence and List. The fact that in this experiment the texts were manipulated across many different content domains (“scenarios”) renders it unlikely that specific consequence predictions were created consistently by the participants. Rather, since not the content but the structure of the text was systematically manipulated, a SCHEMATIC
STRUCTURAL EXPECTATIONS account provides a more plausible explanation for the results than a predictive inference account which focuses on the content of the text. The SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis does not rule out that under specific circumstances content predictions will be created. But note that in any case, content predictions also entail structural expectations: predicting a Solution means that the reader also predicts the text to follow a Problem-Solution structure. For instance, if on the basis of reading a text about a dangerous traffic situations in a street readers expect that road will be closed, they also predict that the text will describe a Solution to the Problem stated earlier. Thus, causal strength explanations focusing on causal prediction are to a large extent in line with the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis. An interesting question for future research is to determine which of the types of predictions are temporally prior: structural or content predictions.

A third alternative explanation might be inspired by ideas on elaborative processing (see chapter 1); the relation that takes less processing time should then be hypothesized to be the relation that takes less elaborative processing. Is it conceivable that Problem-Solution relations indeed require less elaborative processing than Cause-Consequence? It is impossible to answer that question at this point. However, we may speculate against such a conclusion. Looking at the text-linguistic analyses, one would predict Problem-Solution to be more complex (see section 5.2.2) and therefore to require more elaborative processing than Cause-Consequence. The only way to view upon the processing of Problem-Solution structures as requiring less elaborative processing is to assume that the crucial point of a Problem-Solution relation is in the volitional causality: something has to be done to take away an undesirable state-of-affairs. Such a relation of volitional causality can be considered very basic and strong and – still speculating – this might prevent readers from much elaborative processing.

As was explained in section 5.2.1, stronger causal relations require less elaborative processing because the events are more closely related. This implies that the events in the Problem-Solution relation should be more closely related. The question is, how this notion of 'closeness' of the relation should be defined. One possibility might be a definition in terms of causal sufficiency (Van den Broek, 1990; 1994). A challenge for future research is to ascertain whether there is a systematic difference between Problem-Solution and Cause-Consequence in terms of causal sufficiency.

In conclusion, the reading time data are in line with the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis that states that the information embedded in the Problem-Solution relation is processed faster because the Problem triggered a structural expectation. Causal strength can be seen as a challenge for the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis. This challenge has to be taken up in further research, since it cannot be answered on the basis of the current experimental data.

A final question that will be addressed is to what extent the CAUSALITY-BY-DEFAULT hypothesis needs to be abandoned. Strictly speaking, the results are still inconclusive with respect to the CAUSALITY-BY-DEFAULT hypothesis. Evidence for this hypothesis would have been obtained, had there been a processing difference between Cause-Consequence and List. In the experiment no such pattern of results was found. In fact, there was no difference
The Processing Advantage of Causal Coherence Relations

between *Cause-Consequence* and *List*. It can be concluded that causality itself is no guarantee for faster processing of the information in the text.

A second conclusion that can be drawn on the basis of the current experiments is that the results open the possibility of refinement of the original causality-by-default hypothesis. A fourth explanation (besides structural expectations and the two views on causal strength) of why the events in the *Problem-Solution* relation are processed faster than in the other conditions is the fact that the *Problem-Solution* relation is a volitional causal relation whereas in the other causal condition the relation is non-volitional. It might therefore be the case that it is volitional causality that is the default relation.

The main difference between volitional and non-volitional relations is intentionality (Stukker, 2005). Sanders (2005b) argues that (evidence from) child development theories leads to the hypothesis that volitional causal relations are acquired first by children learning their first language. For instance, Piaget (1961) claims that children develop their understanding of causality through intentional interaction with the surrounding world. Also, young children are able to make a distinction between actions that are performed to reach a certain goal from actions that are not goal related (Woodward & Sommerville, 2000). This indicates that volitionality is indeed a very basic cognitive notion, that might explain why humans have a preference to detect volitional actions. If this idea is extended to text comprehension it may be hypothesized that interpreting volitional consequences proceeds quicker than non-volitional relations.

In conclusion, this study provides additional support for a cognitive view on coherence relations. Readers expect a specific structure in terms of coherence relations like *Problem-Solution*. If this structure of the text they are reading matches this expectation, there is a processing advantage. This means that a processing advantage will arise only if the interpretation of the text in terms of its coherence relations structure matches the expectation regarding that structure. That the nature of the expectations was indeed text structural was shown in the continuation task: in the *Problem-* contexts the continuations were more uniform in terms of text structure than in the other conditions. Both for the construction of structural expectations and the interpretation of text structure, the reader will draw on knowledge of coherence relations. Thus, the fact that information embedded in a problem context is processed faster is an indication that readers are indeed sensitive to differences in relation type during discourse processing. This is exactly what a cognitive theory of coherence relations would predict (see Sanders et al., 1992; Sanders, 2005a, 2005b).

One of the remaining questions is what exactly the origin of the expectations is. If the expectations are text structural then this implies that knowledge of coherence relations is part of the readers’ knowledge base. It may indeed be script-like knowledge about text structure. Under such an interpretation, text comprehension can be viewed as a process in which the text activates this knowledge structure and the structure serves not only as a source for expectations but also as a framework for interpreting the incoming text (Kintsch, 2005). A parallel may be seen in the construction of situation models, where scripts are regarded as the types and the model of the current text as the tokens (Zwaan & Radvansky, 1999). In terms of coherence relations, abstract knowledge of types of relations may function as the framework
that enables the construction of expectations and helps the reader interpreting the incoming text.
Chapter 6: Summary and Conclusions

6.1 Introduction

In chapter 1, it was explained that the major aim of this thesis is to contribute to the psychological validity of a cognitive theory of coherence relations. Psychological validity entails that the theory should be able to make precise statements regarding the representation and processes involved in the interpretation of coherence relations. It was argued that focusing on causal coherence relations is a logical first step in the investigation of the psychological status of coherence relations, since these types of relation are one of the most extensively studied phenomena in the field of discourse processing.

It was explained that in cognitive theories of coherence relations the relations are distinguished from their surface indicators. Coherence relations are not part of the surface code of the text, but are part of the meaning representation readers construct on the basis of the text. Their linguistic indicators (such as connectives and lexical signaling phrases) were characterized as processing instructors: linguistic elements that instruct readers how to interpret the relation between discourse segments. In this dissertation, connectives were mainly studied as providing a window on the way in which causal relations were processed and represented. Especially in chapter 3 and 4, connectives were manipulated to create implicit and explicit text versions, in order to investigate the processing and representation of the relations.

This thesis set out to investigate four research questions regarding the representation and processes involved in understanding causal coherence relations. These four questions were:

- RQ1: What is the representational status of causal coherence relations?
- RQ2: What is the representational status of connectives as linguistic indicators of causal coherence relations?
- RQ3: How can the process of causal coherence relation interpretation be characterized in psycholinguistic terms?
- RQ4: How can the process of the interpretation of the linguistic markers of causal coherence relations be characterized in psycholinguistic terms?

In chapter 1, a tentative proposal was presented that provides a preliminary answer to these questions. This cognitive processing account (CPA) was based on hypotheses put forward by Noordman and Vonk (1997) and processing evidence obtained by Cozijn (2000), Singer et al. (1992), and Haldorson & Singer (2002). The CPA involves the assumption that the interpretation process of implicit causal relations involves a text connecting inference that establishes the link, and an extra textual inference (ETI) that consists of encoding the extra textual information underlying the causal connection. When the relation is explicitly marked, the coherence relation is established by means of a text connecting operation (TCO), which
involves encoding the type of coherence relation as specified by the connective. In addition, the CPA assumes that the ETI is made in the presence of a connective too. As for the representation of causal relations, the CPA assumes that causal relations are represented at both the textbase and the situation model level of representation. As was explained in chapter 1, the textbase and the situation model form the meaning representation of the text. The difference between these two cognitive structures is that the textbase contains the representation of concepts and the relations between them that are explicitly mentioned in the text, whereas the situation model contains both text and knowledge derived concepts. It was hypothesized, that the TCI and TCO result in the encoding of the coherence relation at the level of the textbase and the ETI is represented at the level of the situation model.

In chapters 3 through 5 different aspects of the CPA were investigated. The main objective in chapter 3 was to show that the two different inferences are made during the processing of causal relations. These inferential processes were expected to be differentially influenced by the presence of causal connectives. According to the CPA, the process of establishing the causal relation between the clauses (integration) should be faster when a connective is present, and hence a speeding-up effect of the connective is expected. The reason for this is that in the implicit case (without connective) readers have to perform a TCI, whereas in the explicit case a TCO suffices. In this case, integration of the sentences consists of encoding the type of relation signaled by the connective. The speeding-up effect of the connective was referred to as the integration effect.

Besides the integration effect, the CPA also predicts a slowing-down effect of the connective to occur. This slowing-down effect reflects the influence of the connective on an extra textual inference process, which consists of activating and encoding the knowledge underlying the causal relation. Crucially, this process is predicted to be faster when no connective is present, since this knowledge has already been activated during the integration phase. That is, it is assumed that the TCI requires the reader to access the knowledge underlying the causal relation, whereas this is not necessary for the TCO in the explicit case. Therefore, the presence of the connective will lead to a slowing-down effect on the extra textual inference process. This slowing-down effect of the connective is called the extra textual inference effect.

The results obtained in chapter 3 and the conclusions that these results allow, will be presented in section 6.2 more thoroughly, but the main conclusion will be given here. The results of the experiment indicate that the presence of the connective does not necessarily speed up processing. This means that we were unable to detect the integration effect of the connective. There was some evidence, however, for the extra textual inference effect of connectives, but this effect was only detected for the connective want (since), not for omdat (because).

Chapter 4 dealt with the representational status of causal coherence relations. It focused on the question whether separate textbase and situation model levels of representations of the causal relation should be distinguished. The results indicate that no separate textbase could be demonstrated and it was therefore concluded that even explicit causal relations are only represented at the level of the situation model. The experiment and the conclusions will be discussed further in section 6.3.
Both chapters 3 and 4, primarily focused on the processing and representation of explicit causal relations. In chapter 5, the processing and representation of implicit causal coherence relations was investigated. The main objective was to examine whether processing results related to the causal nature of the relation can be explained by assuming bottom-up processing or by means of top-down processing in terms of schematic text structural expectations. The results of the experiment indicate that schematic text structural expectation may be the most important predictor of differences in the processing of coherence relations. The results of and conclusions based on this experiment will be reviewed in section 6.4.

The first four research questions define the substantive issues in this thesis. The fifth and final question pertains to the statistical analysis of experimental data. This question was as follows.

**RQ5:** What is the preferred statistical technique for the analysis of experimental data from common designs used in discourse processing studies?

In chapters 1 and 2, it was argued that the currently widespread use of separate F1- and F2-analyses are not appropriate for the analysis of experimental data of studies in which both multiple items and multiple participants are used. For this reason, the experimental data in this thesis were analyzed with multilevel modeling (MLM).

In the remainder of this chapter, the results of the experiments reported in this dissertation will be summarized, and after accounting for the limitations of the studies, a general conclusion that combines the results of the individual experiments is drawn. The discussion ends with the consequences of this study for cognitive accounts of coherence relations and with directions for future research.

### 6.2 Text connecting inferences and extra textual inferences

In chapter 1, it was explained that the processing of text relies heavily on bridging inferences that create conceptual links between the information expressed in the text. With respect to causal relations, two types of inferences were distinguished. A text connecting inference is involved in the construction of the causal link between the sentences, and an extra textual inference encodes the extra textual information underlying the causal link. On the basis of previous studies it was hypothesized that the presence of connectives has different effects on the text connecting process (the integration process) and the extra textual inference process. To be more precise, the connecting process was assumed to proceed faster as a result of the presence of the connective and the extra textual inference process was assumed to slow down as a function of the presence of the connective. The reasoning was as follows. In the implicit case, the readers use the information in the text and the underlying extra textual information to infer the causal link. In this case, the causal link is established by means of a text connecting inference (TCI). By contrast, in the explicit case the connection does not have to be inferred, but is established in accordance with the processing instructions provided by the connective. In chapter 1, the term text connecting operation (TCO) was used to denote the...
Connecting process that takes places in the processing of explicitly marked causal relations. Since establishing the causal connection requires a connecting inference when the relation is implicit, readers will take longer to process the part of the sentence where the integration of the sentences takes place, as opposed to when the relation is explicitly marked with a connective. This speeding up effect of the connective is referred to as the integration effect of the connective. While the presence of the connective speeds up the integration process, it also slows down the extra textual inference process. The reason for this is that readers of implicit relations can benefit from the fact that the underlying extra textual information was activated during the integration phase. As a result, the extra textual inference process proceeds slower when the relation is explicit than when the relation is implicit. This slowing down effect of the connective is the extra textual inference effect. If experimental materials are used that allow for the separation of the integration and extra textual inference effect of the connectives, a speeding up and a slowing down effect of the presence of the connectives should be observed.

In chapter 3, an experiment was presented that tested these hypotheses. In a self-paced reading the experiment, causal relations were either marked with the Dutch causal connectives *want* or *omdat*, or the relations remained implicit. The results indicated that the presence of connectives did not speed up the processing of the segment where the integration of the sentences (by means of a TCO or a TCI) was hypothesized to occur. As was expected, however, a slowing down effect was observed, but only for one of the connectives that were used: *want*. For the connective *omdat* a slowing-down effect was observed too, but in this case the effect was observed in the first segment of the sentence following the target sentence, and not at the end of the second target-clause. These results lead to the conclusion that the presence of the connectives did not lead to an integration effect in this experiment, but they did seem to affect the extra textual inference process. For the verification data no effect of the presence of the connectives was observed. This finding is in line with the idea that the extra textual inference is made in both explicit and implicit conditions.

Why was no integration effect found? It was argued that this may be mainly due to methodological problems in the experiment. The underlying causal coherence relation had to be compatible with the meaning of both of the connectives that were used. This may have resulted in an ambiguity problem for the connective *omdat*, since this connective can be used to express both objective (semantic) and subjective (pragmatic) causal relations. This problem is smaller for *want*, because this connective has a clear tendency to express subjective causal relations. We can speculate that the ambiguity problem for *omdat* has lead the reader to access the extra textual information underlying the connection between the sentences. As a result, the processing times in the middle and at the end of the sentence did not differ from the implicit condition. This explanation of the absence of the integration-effect still assumes the existence of integration (TCI and TCO) and extra textual inference processes to occur during the interpretation of text. It goes without saying, that much more convincing evidence would have been obtained, had clear speeding-up and slowing-down-effects been detected on the critical regions in the second clause.

With hindsight it may be concluded that the experiment in chapter 3 may have suffered from methodological problems that prevented the detection of an integration effect of the
connectives and also from the failure to detect a clear inference effect in both connective conditions. One aspect of this problem has already been discussed: the ambiguity of *omdat* with respect to the relation it signals. Another aspect of the problem is the assumption that a causal relation is inferred in the implicit condition. This assumption is present in all processing experiments in the literature in which marked relations (e.g. with *because*) are compared to implicit relations. The assumption is crucial if the comparison between the processing of implicit and explicit causal relations is used to investigate inferential processes. Indeed, if readers do not establish a causal coherence relation in the implicit condition – as they do in the explicit condition – the comparison between implicit and explicit conditions will have little to say about causal processing in the implicit case. Instead of assuming that readers of implicit relations infer a causal connection, it should be empirically determined whether or not a causal relation is inferred. In retrospect, both problems could have been avoided by pretesting the materials and checking reader’s representations, so that it would have been clear beforehand what kind of relation was established in the *omdat* and *implicit* conditions. This latter aspect will be discussed further in section 6.7.3.

A second methodological problem in this experiment is the fact that a statement verification method was used to detect whether or not the extra textual inference was made during reading. The use of off-line data to investigate the on-line generation of inferences is in fact insufficient. Although the verification times and question answering times on statements/questions that contain the extra textual information should vary as a function of whether or not the extra textual inference is made during reading, these tasks also reflect the representation readers construct on the basis of the text. Therefore, the verification times also reflect retrieval phenomena. As we do not know how large the contribution of the representation and the on-line processes is, this entails that the verification results obtained in earlier studies (Cozijn, 2000; Halldorson, & Singer, 2002; Singer, et al., 1992, and many others) as well as the results in the current study may not (only) reflect the making of an inference.

Now that we have established that the results should be interpreted with great care, let us assume that the observed differences between implicit and explicit condition do permit conclusions about the underlying causal inference processes. What kind of conclusions may be drawn on the research presented in chapter 3?

The results show that the inference effect of the connective seems to be generalizable over text genres. Previous studies used simple stories or simple sentence pairs. Clearly, the present research enables us to generalize the inference effect to naturalistic newspaper articles. Furthermore, the results indicate that the presence of connectives does not necessarily speed up processing. This may be the result of the fact that subjective relations (marked with *want*) were used in stead of objective relations, but it may just as well indicate that readers in the implicit condition did not infer a causal relation at all. (Further implications and conclusions are presented in section 6.5 and 6.7)
6.3 Textbase versus situation model representations

The experiment reported in chapter 4 directly tested the assumption that explicit causal relations are represented at both the textbase level and the situation model level of representation. Investigating the representation of explicit causal relations provides us with a base line for hypotheses on the representation of implicit causal relations. The need to investigate the representational status of causal coherence relations became apparent in the literature review in chapter 1 in which it was observed that many assumptions are present in the literature, but decisive evidence for the representational status of causal coherence relations is lacking.

The literature review in chapter 1 indicated that the question as to what level of representation causal relations are represented is open for several answers. This is especially true for the representation of implicit causal relations. The literature can be summarized as follows. One line of reasoning is that since the causal connection is established by means of background knowledge, the resulting causal link can be best viewed as a property of the situation model. The other line of reasoning is that the textbase contains the representation of explicit text, but also inferences that preserve local text coherence. Therefore, causal relations that establish local coherence should be considered part of the textbase representation.

It was assumed that explicit causal relations are represented at both levels of representation. This assumption was based on the integration and inference function of the connective *because* (Noordman & Vonk, 1997). The assumption is in line with the idea that the textbase contains the meaning representation of all the explicit information in the text, and the situation model is a further elaboration of the textbase representation. As was argued in chapter 1, this idea is not entirely uncontroversial because researchers have hypothesized that connectives serve as processing instructions for the construction of the situation model. According to this latter hypothesis, there is no need to construct an intermediate textbase level of representation. The literature review therefore suggests that there are two possibilities for the representation of explicit causal relations. The first option is that explicit causal relations are represented both at the level of the textbase and at the level of the situation model. The other option is that the causal relation indicated with the connective is only represented at the level of the situation model.

The recognition experiment in chapter 4 was designed to discriminate between these two possibilities. In this experiment, the relation signaled by the connective varied in the recognition task, so that the target sentences differed from the original sentence at the level of the surface code, textbase or situation model. For instance, an original sentence that contained *omdat* (*because*) was paraphrased by substituting *omdat* by *doordat* (“resulting from the fact that”). This substitution only yields a surface code change: the textbase and situation model of the paraphrase are the same as in the original sentence. In the meaning change condition the target sentence contained *nadat* (*after*) in stead of *omdat*. This introduces a change at both the surface code and textbase levels of representation in comparison to the original *omdat* sentence, while the situation model remains intact. In the model change condition *omdat* was replaced by *en* (*and*). This leads to a change on all three levels of representation. Crucially, if
we rank the conditions with respect to the amount of change from the original, each condition only differs from a previous one by one level of representation. Indeed, originals and paraphrases only differ at the level of the surface code, paraphrases and meaning changes only at the level of the textbase, and meaning changes and model changes only at the level of the situation model. If readers construct all three levels of representation of the original sentence, we expect the probability of participants to give a yes-response in the recognition task to show a decreasing pattern as a function of the amount of change from the original. Thus, if the surface code representation of the sentence is retained during recognition, the proportion of yes-responses should be lower in the paraphrase condition than in the original condition. Furthermore, if a textbase representation is constructed during reading, the probability of a yes-response will be lower in the meaning change condition than in the paraphrase condition. Finally, if readers construct a situational representation of the relation, the probability of a yes-response will be lower in the model change condition than in the meaning change condition.

The results discussed in chapter 4 indicate that only the model change sentences significantly decreased the probability of a ‘yes’ response. Surface changes and explicit propositional changes were not detected by the reader. It was concluded that no evidence for a separate textbase representation of the relation was detected, and that explicit causal coherence relations are only represented at the level of the situation model.

This conclusion has several important implications. The first implication is that the integration and inference effects found for causal connectives cannot be interpreted as evidence for textbase and situation model construction. On the contrary, the experimental results obtained in chapter 4 lead to the conclusion that both integration and extra textual inference processes are involved in the construction of the situation model. This means that the CPA that was presented in chapter 1 makes the wrong assumptions about the levels of representation readers construct of causally related events. Clearly, the results are much more in line with an account that assumes that causal connectives serve as processing instruction for situation model construction.

The second implication of the results is that we are now able to hypothesize about the representational level of implicit causal relations. If readers of explicit relations do not construct a separate textbase representation of the relation, it is highly unlikely that readers of implicit causal relations do construct an intermediate propositional representation of the causal connection. This leads to the idea that background knowledge of the reader, which is crucial for situation model construction, may be the most important source of information for the interpretation of coherence relations. The implications for cognitive accounts of coherence relations and discourse processing theories will be discussed in section 6.7.

Of course, the conclusion that explicit causal relations are only represented at the level of the situation model is based on the failure to detect an effect of the meaning change condition. As was argued in chapters 4 and 5, acceptance of the null-hypothesis is always suspicious, because proving the null-hypothesis is virtually impossible and failure to detect significant effects may be the result of lack of statistical power. For this reason, one should replicate the results of this experiment before a final conclusion about the absence of a separate textbase
representation of the causal connection can be drawn. Notwithstanding these cautionary notes and even though several counter-explanations were discussed in chapter 5, we believe that the application of the sentence recognition paradigm has provided innovative insights in the issue of the levels of coherence relation representation.

6.4 Causality-by-default or schematic structural expectations?

The results in chapter 4 indicate that explicit causal coherence relations are represented at the situation model level of representation. This led to the conclusion that it is likely that implicit relations are also only represented at the situation model. As the knowledge of the reader is a crucial factor for situation model construction, this opens the possibility that the reader’s knowledge plays an important role in the interpretation of implicit causal connections. One kind of knowledge that may play a role was investigated in chapter 6: schematic knowledge of text structure.

Earlier investigations have indicated that causal relations are processed faster than additive relations. An intriguing result was obtained by Sanders and Noordman (2000) who have shown that Problem-Solution relations are processed faster than List relations, even though Problem-Solution relations are, (text-) linguistically speaking, much more complex than List relations in terms of their text linguistic properties. Two explanations were given for these results. The first explanation is the CAUSALITY-BY-DEFAULT hypothesis, which assumes that readers by default try to connect statements by means of a causal relation and will only arrive at a different interpretation when a causal relation is not possible. The second explanation refers to schematic text structural expectations reader create during reading. The latter explanation involves the assumption that the specific configuration of the Problem-part of the text triggers the readers’ schematic knowledge of the Problem-Solutions structure. Once this schematic knowledge is activated, the reader expects that a Solution will be presented. If this expectation is substantiated by the text, the integration of the solution sentence in the developing representation requires less processing resources in comparison to situations were schematic structural expectations are not constructed. Supposedly, the List structure evokes weaker expectations. Therefore, the reader will not benefit from these expectations during text processing.

The experiment reported in chapter 5 set out to distinguish between these two explanations. Identical target sentences were connected to the context by means of three different coherence relations: Problem-Solution, Cause-Consequence, and List. These three coherence relations differ in two respects. First, they differ with respect to the nature of the relation, namely causal (Problem-Solution and Cause-Consequence) versus additive (List). Second, they differ with respect to the strength of schematic structural expectations, namely strong (Problem-Solution) versus weak (Cause-Consequence and List). According to the CAUSALITY-BY-DEFAULT hypothesis, a difference is expected between the causal conditions and the non-causal conditions. According to the SCHEMATIC STRUCTURAL EXPECTATIONS hypothesis a processing difference is expected between the strong expectations condition and the weak expectations conditions. The assumptions underlying both hypotheses were confirmed in a
continuation experiment. In this continuation experiment the contexts prior to the target sentences was presented to the participants and they were asked to complete the text in one or two sentences. The results of the continuation task indicate that in the Problem-context the coherence relations produced where more uniform than in the Cause- and List-contexts, which corresponds to the assumption that these contexts are more likely to generate schematic expectations. Also, over contexts participants showed a strong preference for producing causal coherence relations.

The results of the reading experiment strongly favored the structural expectations hypothesis: a processing difference was only found for Problem-Solution versus Cause-Consequence and List and not between Cause-Consequence and List. Also, in a sentence verification task, performance was quicker and better in the causal conditions than in the non-causal condition. On the basis of these results it was concluded that schematic expectations guide the interpretation of text in a top-down manner, and that the causal nature of the relation per se, does not necessarily speed up processing. With respect to the representation it was concluded that the causal nature of the relation leads to a better quality of the representation.

Several counter explanations of the results were addressed in chapter 5. The most challenging counter explanation is repeated here. The pattern of reading results may also be explained in terms of causal strength. Since the relations in the Problem-Solution structure seem to be stronger than in the Cause-Consequence and List structures, the speeding up effect associated with the Problem-Solution structure may be a direct reflection of this difference in strength. It was argued, however, that since we need additional assumptions that explain how causal strength might influence processing, the observation that the structures differ in terms of causal strength in itself does not explain very much.

Two additional explanations with respect to causal strength were discussed. The first is that in the case of strong causal relations readers are able to predict the upcoming event in terms of the contents of the text. Yet, the contents of the texts in the experiment in chapter 5 were so diverse that it is unlikely that readers were able to consistently expect specific events to be described in the text. Still, the schematic structural expectations account does not rule out that in specific circumstances content expectations are created by the reader. Still, if a specific Solution is expected after reading a Problem, it certainly is the case that a text structural expectation is created too. In any case, more research is needed to disentangle schematic expectations and content predictions.

The second explanation why strong causal relations are processed faster is that weak causal relations require more elaborative processing (see also chapter 1). These additional elaborations result in an increase in processing times. It was argued in chapter 5, that the interpretation of the entire Problem-Solution structure requires a large amount of elaborative processing, because the Problem-Solution relation consists of two causal relations. One relation holds between the Problem that causes the Solution (the volitional action) to occur, the other causal relation is the relation between the Solution and the Problem being solved (Sanders, 2005; Sanders, Spooren, & Noordman, 1993). On the basis of an elaborative processing account it could therefore be expected that the Solution to a Problem is processed slower in comparison to a situation in which the exact same sentence is part of a Cause-Consequence or List structure. The obtained results clearly contradict this expectation.
In sum, the results of the experiment reported in chapter 5 lead to the conclusion that schematic expectation created by the reader may explain the processing differences between coherence relations. It could not be shown that the nature of the relation per se (causal versus non-causal) necessarily results in processing differences. With respect to the representation, however, the nature of the relation does seem to be a crucial factor for predicting memory phenomena: both causal relations (Problem-Solution and Cause-Consequence) were again shown to be represented better. Where future research is concerned, it should focus on disentangling the content and structure expectations readers may generate during reading and should investigate whether there is a systematic difference between Problem-Solution and Cause-Consequence structures in terms of causal strength.

6.5 The role of surface code and meaning representation in the processing of causal coherence relations

In chapter 1, two predictions based on a fundamental assumption underlying a cognitive account of coherence relations were presented. This fundamental assumption is that coherence relations are not properties of the text but a characteristic of the meaning representation readers construct on the basis of the text. The straightforward predictions that were based on this fundamental assumption were that both the linguistic marking of relations and the nature of the relation influence the processing and representation of text. On the basis of existing studies, a more detailed cognitive processing account (CPA) was formulated. Two hypotheses were derived from this CPA. These hypotheses were that linguistic marking of causal relations leads to two different effects on the processing of causal relations. A speeding-up effect should occur at the time readers integrate the sentence by means of constructing a causal link between the representations of the individual sentence, and a slowing down effect should occur at the time the knowledge underlying the causal relation is encoded in the representation of the text. It was also hypothesized that these effects reflect two underlying processes involved in the construction of the textbase and the situation model. The integration of the sentences results in a representation of the causal link at the level of the textbase, and an extra textual inference process encodes the underlying knowledge at the level of the situation model. As was explained in chapter 1, these processing hypotheses follow a bottom-up view of text processing, where inferential processes (text connecting and extra textual) are needed to connect the current sentence to the previous context.

The CPA gives a preliminary answer to the first four research questions. These research questions are as follows.

RQ1: What is the representational status of causal coherence relations?
RQ2: What is the representational status of connectives as linguistic indicators of causal coherence relations?
RQ3: How can the process of causal coherence relation interpretation be characterized in psycholinguistic terms?
Now that we have reviewed the results and conclusions of the research described in chapters 3, 4, and 5, we are able to give an answer to these research questions. It can be concluded that the CPA is not supported by the empirical evidence that was obtained. With respect to the speeding-up and slowing-down effects of causal connectives, it has been concluded that the experiment has some methodological problems that prevented the detection of a speeding up effect. The detection of a slowing-down effect for the want condition, however, indicates that extra textual inference is indeed made, irrespective of the particular causal relation type and that making this inference is not restricted to the processing of simple unnatural narratives. An acceptable conclusion based on chapter 3 is that the speeding-up effect was not properly tested in the experiment and therefore it does not allow for conclusions with respect to the integration effect of connectives. It has been demonstrated, though, that the presence of connectives may lead to a slowing-down effect. This effect is most likely to pertain to encoding of the knowledge underlying the causal relation in the representation of text.

The conclusion of chapter 4 is that causal coherence relations can be best considered as represented at the level of the situation model. Therefore, a combined conclusion on the basis of chapters 3 and 4 is that the effects detected (and not detected) should be explained with reference to the construction of the situation model. The integration effect of connectives can therefore be considered as faster construction of a model representation of the causal connection between the events and the extra textual inference effect as slower encoding of the underlying knowledge in the model. This position is compatible with a surface-code-as-processing-instruction view on the effect of connective and other lexical signaling devices (Britton, 1994; Gernsbacher, 1990; Gernsbacher, & Givón, 1995; Sanders, & Spooren, 2001; see also Graesser, Millis, & Zwaan, 1997). In such a view, the explicit indicators in the text specify how the situation model representation of the text must be constructed. In this case, it indicates that the situation model should represent the causal connection between the events described in the text. The model can be constructed without the need of constructing an intermediate textbase representation. The conclusion that causal coherence relations are represented at the level of the situation model is compatible with many theories of discourse processing and representation (Fletcher, 1994; Zwaan & Radvansky, 1998; Graesser, Singer, & Trabasso, 1994; Singer, 1993; Halldorson & Singer, 2002).

Although we must be very careful in assuming that causal connections are inferred without the presence of a marker (see section 6.2) the results obtained with explicit causal relations, allow for predictions about the representation of implicit causal relations. If explicit causal relations are not represented at the textbase, it is implausible to assume that implicit relations are indeed represented at the level of the textbase. It may therefore be predicted that the role of background knowledge is the most important predictor for the interpretation of causal connections between sentences. This follows from the importance of background knowledge in the construction of situation model representations.

One type of background knowledge that may be important for situation model construction was investigated in chapter 5: readers’ schematic knowledge of text structure. On the basis of
this experiment it was concluded that knowledge of text structure guides the interpretation of text. Under the assumption that coherence relations are represented at the level of the situation model, we are now able to conclude that the activation of the appropriate schematic text structural knowledge as part of situation model construction may function as a guide for the interpretation of the incoming text.

In other words, the combined results of the experiments in the previous chapters strengthen the conclusion that coherence relations are represented at the level of the situation model. Knowledge structures that are activated during the construction of the situation model guide the interpretation of the text in a top-down manner. This entails that the CPA presented in chapter 1 underestimates the top-down influences during text processing.

A view that is more compatible with the results presented in this thesis was formulated by Noordman and Vonk (1998). This view adopts both bottom-up and top-down influences in text processing. An interpretation of the results presented in this thesis that is compatible with that account may be as follows. The explicit text triggers knowledge structures in long term memory in a bottom-up fashion. The mechanisms that are involved in this process may be envisioned as the spreading of activation from concepts in the explicit text to concepts in memory along the lines proposed for instance by Kintsch’s (1988, 1998) construction-integration model. These knowledge structures guide the processing of the explicit text yet to come in a top-down manner. For instance, the presence of a connective triggers the activation of schematic structural knowledge of the relation type indicated by the connective, and the clause introduced by the connective is subsequently interpreted in accordance with this knowledge structure. This results in a representation at the situational level that combines both explicit text and background knowledge.

When the relation is implicit, the identification of the appropriate schematic knowledge structure depends on how well the explicit text allows for the recognition of the appropriate schema. For texts with clearly identifiable structures, such as Problem-Solution texts, this is more likely to be the case than for other structures, such as Cause-Consequence and List (see chapter 5). If the appropriate text structural schema is identified, the interpretation of the incoming text is guided by this schema. The end result is a situational representation of the text in which the explicit text and background knowledge are integrated.

From this point of view, integration of the sentences, either by means of a text connecting inference or a text connecting operation, leads to a situational representation in which the individual sentences are combined with reference to schematic knowledge of text structure. The extra textual inference that was detected in previous studies, however, consists of general background knowledge underlying the causal relation. What may be the role of this general background knowledge? First, it may be the case that general knowledge is used to ascertain whether the schematic knowledge that is activated by the connective or by the context is appropriate for the current text. Second, for implicit cases, general knowledge about the relation between the events described in the text may be needed to identify the appropriate text schematic knowledge, when the appropriate schema was not activated by the context. Third, a general knowledge base may be needed to check whether the coherence relation signaled in the text does indeed make sense.
In sum, the experiments reported in this thesis, allow for the reformulation of the original hypotheses in two respects. First, integration (text connecting inference and text connecting operation) and extra textual inference can be interpreted in terms of situation model construction. Second, the bottom-up view on text processing may not be entirely appropriate. We are in need of an account that incorporates both bottom-up and top-down processes and makes reference only to the situation model. Such an account would be compatible with many aspects of discourse processing, including the occurrence of text connecting inferences (integration) and extra textual inferences in the processing of causal relations as well as the top-down influence of text structural knowledge on the processing of text.

### 6.6 Implications for cognitive theories of coherence relations

The general conclusion drawn on the basis of all the experiments described in the current thesis has several implications for cognitive theories of coherence relations. As was explained in chapter 1, a fully developed cognitive theory of coherence relations should be able to predict and explain processing and memory phenomena associated with the interpretation of coherence relations. The results of this study contribute to the development of such a theory. With respect to the representation of causal coherence relations it was concluded that these relations are represented at the level of the situation model. This is compatible with the main assumptions underlying cognitive accounts of coherence relations as these theories claim that coherence relations are part of the meaning representation of the text. On the basis of the present study, this assumption can be further specified. A text is coherent when in the situation model of the text the information is conceptually connected by means of coherence relations.

This conclusion has important implications, because the situation models that readers construct on the basis of the text rely heavily on their background knowledge. As readers may differ dramatically in the knowledge they have about the contents of the text, the kinds of relations readers construct on the basis of the exact same text may also differ. This poses a potential problem for the descriptive adequacy of coherence relation theories, since the type of relation assigned by the text linguist may not match the type of relation the reader will construct in his representation. A similar point is made by Kintsch (1998), who claims that

> “The mental representation of a text a reader constructs includes the textbase (not necessarily complete and veridical) plus varying amounts of knowledge elaborations and knowledge-based interpretations of the text – the situation model. Neither the micro- nor the macrostructure of the situation model is necessarily the same as the micro- and macrostructure of the textbase, for the reader may deviate from the author’s design and restructure a text both locally and globally according to his or her own knowledge and beliefs.” (p. 50).

This is of course problematic if coherence is defined as a property of the meaning representation readers construct on the basis of the text. From this perspective, descriptive
adequacy of a theory of text structure can only be achieved with reference to an idealized reader who is assumed to possess the knowledge necessary to construct a situational representation that mirrors the exact same text structure as analyzed by the text linguist. This is not a trivial problem, because assumptions about idealized readers may have negative consequences for the psychological validity of the respective theory. Of course, both descriptive adequacy and psychological validity are crucial for cognitive theories of coherence relations (Sanders, Spooren, & Noordman, 1992; 1993).

A second important conclusion from the present study is that knowledge of coherence relations (text structure) influences the processing of text. This underlines the psychological validity of coherence relations, because it shows that readers do not only have this knowledge, but can also use this knowledge during the on-line interpretation of text. The processing evidence presented in this dissertation further supports the psychological validity of the taxonomy presented in Sanders, Spooren and Noordman (1993). Most importantly, the results imply that the distinction between causal and additive relations (the basic operation) is a fundamental cognitive distinction. For instance, as the results in chapter 5 indicate, information from a sentence is verified quicker and more accurately when it is connected by means of a causal relation (Cause-Consequence or Problem-Solution) than when the exact same information relates to the context by means of an additive relation (List). This representational advantage of causal relations shows that the distinction between causal and additive operations has a psychological counterpart.

In short, the conclusion that coherence relations are represented at the level of the situation model confirms the assumption that coherence relations are part of the meaning representation of the text, and the processing results indicate that knowledge of text structure is used by the reader during reading and that the distinctions between the cognitive primitives Basic Operation are reflected in representation results. At the same time the conclusion that coherence relations are presented at the level of the situation model poses a challenge for the development of cognitive theories of coherence relations.

6.7 Implications for processing theories and experimental research on discourse processing

The studies reported in this thesis and conclusions based upon them have several implications for processing theories, both in theoretical and methodological terms. These implications will be presented below.

6.7.1 Implications for theories of discourse processing

To my knowledge, this study is the first to directly investigate the representational status of coherence relations. Thus the major contribution to processing research is probably the conclusion that causal relations are best considered to be part of the situation model. Even when the relation is explicitly marked, an intermediate textbase representing the causal
relation is not constructed. This conclusion is in line with a constructionist view of text processing (Bransford, & Franks, 1978; Graesser, Singer, & Trabasso, 1994), which assumes that readers aim at constructing deep levels of representation that are both locally and globally coherent.

Another implication of the present study is that it further corroborates the idea that text comprehension is a mixture of bottom-up and top-down processes (Kintsch, 1998; 2005; Noordman & Vonk, 1990, 1998; Van den Broek, Rapp, & Kendeaou, 2005). Results obtained in this dissertation clearly show that reader’s schematic knowledge can influence processing (chapter 5), but also that the interpretation of causal relations involves extra textual inferences that are generated bottom-up.

A third important contribution of the present study is that some fundamental processing findings reported in earlier studies are also present during the processing of non-narrative discourse. As was explained in chapter 1, earlier processing studies have primarily used short stories as their materials and one of the aims of this study was to investigate to what extent the experimental results of these studies can be generalized to non-narrative discourse. The present study shows that extra textual inferences and readers’ schematic knowledge play a role in the processing of naturally occurring short news paper articles. Therefore, the present study clearly indicates that the role of extra textual inferences and readers’ schematic knowledge is not only important in the processing of simple stories but also in naturalistic non-narrative text. Because naturalistic texts were used, the conclusions in this study have a large ecological validity. This shows how fruitful it can be do experiments with naturalistic texts.

6.7.2 The statistical analysis of experimental data

In Chapter 3, it was argued how important it is to use statistical models in which the fixed effects of the experimental manipulations are tested against the random variance associated with items and persons in a single analysis. It was explained that the current practice of F1- and F2-analyses is not appropriate for the analysis of experimental data. The main reason is that these analyses do not give insight in the question whether the variance created with the experimental manipulations outweigh both the item variance and person variance simultaneously.

A significant F1-ratio, means that the experimental manipulation is more important than the random variance associated with persons. This enables the conclusion that is likely that the variance caused by the manipulation is not restricted to the specific sample of persons but is indicative of a population effect of the manipulation over similar samples. However, the estimate of the population effect only holds for other samples of persons that read the exact same text. In other words, a significant F1-ratio enables the researcher to conclude that a population effect is likely to exist for the specific sample of texts that were used. This is a direct consequence of the fact that in the F1-test the texts are considered to constitute a fixed effect. Likewise, a significant F2-test allows the researcher to conclude a population effect is likely to exist for the specific sample of persons that were used in the experiment. Crucially,
however, separate F1- and F2-analyses do not give any information about the question whether a population effect of the manipulation is likely to exist when both a new sample of participants and a new sample of texts are used. For this reason, using F1- and F2-tests restricts the generalizations researchers can make with respect to the experimental effects. The conclusions researchers draw about the population effect are always bound by either the specific texts that were used or by the specific sample of persons that was used in the experiment.

In this thesis, multilevel modeling was used to estimate the population effect of the experimental manipulations with respect to items and persons simultaneously. This means that the estimates of the population effects in this thesis are not restricted to the specific samples or persons and texts, but can be generalized to comparable samples of other persons reading other texts. This does, of course, not entail that the estimated population effects applies to every single text and person in the real world, or every single text and person in the samples that were used, for that matter.

A drawback of the technique is that it is hard to provide a measure of effect size. And this is the reason why no measures of effect size were presented in the present study. Of course, on the basis of the estimated means and standard errors a confidence interval, which gives an indication of how large or small the population difference may be, can be constructed. However, a general measure of effect size, such as a measure that indicates the percentage of variance explained (like R²), that is easily interpretable is not available. For instance, Snijders and Bosker (1999) propose a R²-like measure, but this measure can take negative values depending on how it is calculated. It is, of course, a highly unpleasant state of affairs to have an effect of an experimental manipulation and have a negative measure of effect-size at the same time.

Most importantly, however, the technique clearly has great advantages. Two important advantages of using multilevel regression techniques in the analysis of experimental data are the following. First, the multilevel models enable the inspection of text level residuals. Inspection of these residuals gives a good indication of the extent to which the experimental manipulation generalizes to specific texts. In all of the experiments reported in the previous chapters, the experimental manipulation clearly interacted with the individual texts (although these analyses were not reported). This offers a good starting point for follow-up research as this research begins with trying to explain why one text behaves differently than another text. This may lead to the formulation of new hypotheses that may be tested in new experiments.

Second, not only the individual residuals are worth investigating, but also the residual variances. For instance, a substantive interpretation of the absence of an integration effect was possible after the residual variance was shown to differ between the conditions. Note that despite the fact that the means did not differ significantly the difference in residual variance is clearly a result of the experimental manipulations. For this reason, it is worthwhile not only to try and hold on to trying to obtain significant differences between the mean reading times, but also to investigate the effect of the experimental manipulations on the variance estimates.
6.7.3 Implications for the design of experimental studies on coherence marking

In section 6.2, it was explained that one of the methodological problems in the experiment reported in chapter 3 was that it was assumed that readers of implicit causal relations do in fact infer the causal connection. It was argued that this assumption is crucial for the investigation of causal inference processes. For instance, a speeding up effect of a connective can be expected if readers of an implicit relation infer a causal connection, while in the explicit case readers encode the relation signaled by the connective by means of a TCO. Thus, it is by virtue of the comparison of the processing of implicit and explicit relations that studies obtain insight in causal inferences.

The assumption that readers construct a causal relation when the relation is implicit is actually a more general problem that can be said to apply to many experiments in which the presence of the connective is manipulated and in which it is assumed, and not assessed, that in the implicit condition an identical relation is inferred as the one indicated by the causal connective. The crucial point is that it is by no means certain that the same relation is actually interpreted by readers in the implicit and explicit versions of the relation.

In methodological terms, construct validity is at stake in experiments in which the presence of connectives is manipulated and in which it is assumed, but not established that implicit and explicit relations receive an identical interpretation. The doubts about construct validity are that it is questionable that the reading times that are measured in implicit conditions reflect the construction process of causal relations. As long as it is not established with high probability that implicit relations are indeed interpreted in the same way, the construct validity of these experiments will remain problematic.

For example, many experiments have shown a speeding-up effect of connectives and other lexical signaling devices, and it has been concluded in the literature that connectives speed up processing because in the implicit condition readers have to infer the causal connection all by themselves. However, we cannot be sure that this processing pattern is indeed due to the inference of this causal relation. Readers might just as well have inferred an additive relation, a temporal relation between the events, or they could have inferred a claim-argument relation. In either case, the presence of the connective may have resulted in shorter reading times, but it does not give any information on the actual processes involved in the interpretation of the implicit relations.

We could also argue the other way around. As long as we do not know what kind of relation readers construct in the implicit conditions, there is no way of predicting what kind of processing results are likely to occur as a result of the inclusion of the connective. For instance, readers might infer a non-causal connection. Should inferring such a connection take longer or is it more likely that such a connection is established quicker? Only if we know what happens in the implicit condition, we are able to predict the kind of processing pattern we are likely to observe.

These considerations provide a major challenge for discourse processing studies that investigate the influence of coherence marking. In order to fully appreciate the point that is made here, let us assume for a moment that no causal connection is inferred at all in implicit
conditions. However unattractive the assumption is for coherence relation theories, it might be in line with earlier processing evidence. It might explain, for instance, why readers react quicker to the extra textual information in the connective condition (Cozijn, 2000). The reason for this is that when the causal connection is not inferred, the extra textual information that is needed to infer the causal connection is probably not activated during reading either. In other words, if the causal relation is not inferred, the extra textual information is only likely to be activated in the explicit condition and hence readers of explicit versions react faster to this information. Also, when the connection is not inferred in an implicit version of a relation, it is likely that the quality of the representation is better after explicit relations than after implicit relations. Indeed, experiments in which memory measures were used, have repeatedly shown a representational advantage for explicit causal relations (Degand, Lefèvre, & Bestgen, 1999; Degand, & Sanders, 2002; Millis, & Just, 1994). This whole line of reasoning demonstrates how important it is to directly assess whether or not the causal connection is inferred during reading of implicit causal sentences.

Hence, experiments in which the presence of connectives and other signaling devices are manipulated should provide evidence what exactly readers infer in implicit conditions. One possibility of obtaining such evidence is to use a version of the recognition paradigm that was used in chapter 5. For instance, implicit causal relations could be included in an on-line processing experiment, while the explicit versions of the relations were used in a recognition task immediately after reading. It may be expected that readers who indeed infer the causal TCI during reading will not perform differently on tests with an explicit text or with an implicit (original) text: they have already inferred the connection on-line. Hence, no difference in the proportion of yes-responses would be expected.

Of course, the recognition paradigm is only one potential way to investigate whether the causal relations are being inferred during reading in the implicit version. Other possibilities include the use of think-aloud-protocols to expose the inferences being generated during reading (e.g. Trabasso, & Suh, 1993) or question-answering methods (e.g. Graesser, 1981). In any case, the results obtained in these preliminary tasks should also be connected to on-line measures such as reading times or word naming latencies to verify that the inferences are actually generated during reading (Graesser et al., 1996). Using experimental tasks like these to corroborate crucial assumptions about the causal inference underlying the processing of text, will lead to more firmly established evidence with respect to the processing of causal coherence relations. This will ultimately lead to a better understanding of the role of causal coherence relations in the processing and representation of text. This brings us one step closer to the formulation of a psychologically valid cognitive theory of coherence relations. This is a goal that is certainly worthwhile and to which this study has contributed new results and insights.
References


Glenberg, A.M., & Sashi, M. (1992). When Minimalism is not enough: Mental Models in Reading Comprehension. *PSYCHOLOQUY* 3(64) reading-inference.1.2


References


References


References


Samenvatting in het Nederlands

Dit proefschrift richt zich op de cognitieve processen en representaties die betrokken zijn bij de interpretatie van causale coherentierelaties in tekst. Coherentierelaties zijn betekenisrelaties zoals Oorzaak-Gevolg, die worden gelegd tussen tekstdelen, bijvoorbeeld zinnen. Cognitief georiënteerde tekstlinguïstische theorieën claimen dat coherentierelaties onderdeel zijn van de betekenisrepresentatie die lezers creëren op basis van een tekst. Deze theorieën nemen aan dat lezers een tekst pas volledig hebben begrepen wanneer zij de (representaties) van tekstsegmenten onderling hebben verbonden door middel van coherentierelaties. Bij het afleiden van coherentierelaties kunnen lezers gebruik maken van expliciete markeringen van coherentierelaties zoals connectieven (bijvoorbeeld omdat of doordat). De aanwezigheid van dergelijke linguïstische indicatoren voor coherentierelaties is echter niet noodzakelijk voor het afleiden van de coherentierelaties. Met andere woorden, aangenomen wordt dat tijdens het verwerken van onderstaand zinspaar (1) de lezer een causale relatie zal afleiden, hoewel er geen causaal connectief gebruikt wordt om het verband tussen de zinnen expliciet te maken, zoals in (2).

(1) Pieter zat onder de blauwe plekken. Hij was van zijn fiets gevallen tijdens het spelen.
(2) Pieter zat onder de blauwe plekken, omdat hij van zijn fiets was gevallen tijdens het spelen.

In hoofdstuk 1 wordt nagegaan hoe bovengenoemde cognitieve claims over coherentierelaties zich verhouden tot de psychologische en psycholinguïstische inzichten in de manier waarop lezers een tekst en in het bijzonder causale relaties verwerken. Allereerst wordt in het hoofdstuk duidelijk dat spreken over dé betekenisrepresentatie in feite een simplificatie is wanneer we recente psychologische inzichten in de mentale representatie in ogenschouw nemen. In het hoofdstuk wordt beschreven dat de mentale representatie bestaat uit verschillende niveaus van representatie. In psychologische theorieën over tekstbegrip wordt doorgaans een onderscheid gemaakt tussen de volgende niveaus: surface code, textbase en situatie-model.

De surface code is een niveau van representatie dat bestaat uit de letterlijke bewoordingen in de tekst. De textbase is een representatie van de betekenis van de tekst per se. Dit niveau van representatie wordt vaak gezien als een netwerk waarin de propositionen die uitgedrukt worden in de afzonderlijke zinnen onderling zijn verbonden. Het situatie-model is een niveau van representatie waarin informatie uit de tekst wordt aangevuld met informatie uit de kennis van de lezer. De textbase en het situatie-model vormen samen de betekenisrepresentatie van de tekst. Aansluitend bij cognitieve benaderingen van coherentierelaties wordt in het hoofdstuk bearugmenteerd dat linguïstische markeringen van coherentierelaties (zoals connectieven en signaalzinnen) worden gerepresenteerd op het niveau van de surface code, en de coherentierelaties zelf op het niveau van de textbase en het situatie-model.
In hoofdstuk 1 wordt verder betoogd dat het interpretatieproces van causale coherentierelaties kan worden beschreven in termen van inferentiële processen. Bij afwezigheid van causale markeringen is het waarschijnlijk dat de lezer twee inferenties uitvoert om de causale relatie af te leiden, die beide een beroep doen op de achtergrondkennis van de lezer. De lezer voert zowel een tekstverbindingen (text connecting) inferentie (TCI) als een extra-tekstuele inferentie (ETI) uit. De TCI bestaat uit het leggen van het causale verband. De relatie wordt afgeleid uit de expliciete tekst en de achtergrondkennis van de lezer. Bijvoorbeeld, bij het afleiden van de causale relatie in het voorbeeld (1) activeert en gebruikt de lezer de kennis dat van een fiets vallen kan leiden tot blauwe plekken.

De achtergrondkennis van de lezer die gebruikt wordt bij het bepalen van de coherentierelatie wordt in de betekenisrepresentatie opgenomen door middel van de ETI. Wanneer causale markeringen wel aanwezig zijn, dan wordt alleen de ETI door de lezer gemaakt. Dat houdt in dat de kennis die ten grondslag ligt aan de causale relaties door de lezer geactiveerd wordt en opgenomen wordt in de mentale representatie van de tekst. De TCI hoeft niet te worden uitgevoerd omdat de markering specificeert van welk verband sprake is. In plaats daarvan wordt de door de markering gespecificeerde relatie in de representatie opgenomen door middel van een tekstverbindingen (text connecting) operatie (TCO).

Deze uitspraken over de inferentiële processen die betrokken zijn bij het interpreteren van causale relaties zijn gebaseerd op eerder onderzoek dat aantoont dat tijdens het lezen twee effecten optreden als gevolg van de aanwezigheid van connectieven. In de eerste plaats leidt het connectief tot een versnelling van leestijd, omdat het interpreteren van de causale coherentie relatie minder moeite kost wanneer een connectief het verband expliciet aanduidt. De reden voor dit versnellingseffect is dat in de impliciete conditie (zonder connectief) een TCI wordt uitgevoerd, terwijl in de expliciete conditie (met connectief) het leggen van het verband neerkomt op het uitvoeren van een TCO. Het versnellingseffect van het connectief wordt ook wel het \textit{integratie-effect} van het connectief genoemd.

In de tweede plaats leidt de aanwezigheid van het connectief tot een vertragingseffect op de leestijden. De reden hiervoor is dat het uitvoeren van de ETI meer tijd kost wanneer een connectief aanwezig is. Op het moment van de ETI wordt namelijk achtergrondkennis geactiveerd en opgenomen in de betekenisrepresentatie. Verondersteld wordt dat deze achtergrondkennis in de impliciete conditie al in een vroeger stadium geactiveerd is en dat daardoor het uitvoeren van de ETI minder tijd zal kosten dan in de expliciete conditie. Het vertragingseffect van connectieven wordt het (extra-tekstuele) \textit{inferentie-effect} genoemd.

In hoofdstuk 1 wordt op basis van de literatuur gerapporteerde empirische resultaten een verband gelegd tussen de niveaus van representatie en de inferentiële processen die betrokken zijn bij de interpretatie van causale coherentierelaties. De TCI en TCO zijn processen die leiden tot het representeren van het causale verband op het niveau van de \textit{textbase} en de ETI is onderdeel van het constructieproces van het situatie-model. De bovenstaande aannames vormen samen een \textit{cognitive processing account} (CPA) van de interpretatie van causale coherentierelaties. De validiteit van deze CPA wordt nader onderzocht in de hoofdstukken 3 en 4.
Samenvatting

In hoofdstuk 2 wordt de statistische techniek beschreven waarmee de meeste van de experimentele resultaten in het proefschrift geanalyseerd zijn. Deze techniek staat bekend als Multilevel analyse (Multilevel Modeling (MLM)). In het hoofdstuk wordt betoogd dat deze techniek een aantal voordelen biedt boven variantieanalyse (ANOVA) bij de analyse van de resultaten van experimenten waarin zowel meerdere items (woorden/teksten, etc.) als meerdere personen worden gebruikt. Wanneer onderzoekers meerdere items in het onderzoek gebruiken, dan moeten zij kiezen of zijn in de analyse van de afhankelijke variabele deze item-variabele als fixed of random effect opnemen. Deze keuze heeft gevolgen voor de reikwijdte van de conclusies die de onderzoeker mag trekken op grond van zijn analyse. Wanneer item als fixed effect in de analyse wordt betrokken dan gelden de conclusies over het effect in de populatie alleen voor een populatie personen die exact dezelfde items krijgen aangeboden als in het onderzoek. Wanneer item als random effect wordt opgenomen dan mag de onderzoeker generaliseren naar een populatie personen en een populatie items.

Wanneer de onderzoeker item als een random factor wil beschouwen, dan is de analyse van de resultaten een stuk lastiger met ANOVA dan met MLM. De belangrijkste reden daarvoor is dat voor het uitvoeren van een ANOVA een quasi-F moet worden uitgerekend. Het uitrekenen van die quasi-F moeten variantieschattingen handmatig gecombineerd worden. Welke variantieschattingen gebruikt dienen te worden, is afhankelijk van het specifieke design dat wordt gebruikt in het onderzoek. Dat betekent dat er geen algemeen recept is dat onafhankelijk van het design kan worden toegepast. Het toepassen van F₁- en F₂-analyses, zoals dat momenteel gebruikelijk is, biedt geen oplossing voor het probleem, omdat in beide analyses telkens één factor als fixed factor wordt beschouwd. In de F₁-analyse is het de item-factor en in de F₂-analyse de proefpersonen-factor. De consequentie daarvan is dat experimentele effecten altijd gebonden zijn aan één van de twee steekproeven: het effect is generaliseerbaar naar een populatie personen die de gebruikte steekproef van items krijgt aangeboden (F₁) of naar een populatie items die worden aangeboden aan dezelfde steekproef van personen (F₂). Generaliseren naar beide populaties tegelijkertijd is niet mogelijk. MLM lost de bovengenoemde problemen op. Hetzelfde basismodel kan worden toegepast ongeacht het experimentele design dat wordt gebruikt en er hoeven geen variantieschattingen handmatig gecombineerd te worden. MLM is om die reden van grote waarde voor experimenteel taalwetenschappelijk onderzoek. Het belangrijkste voordeel is dat de techniek statistische uitspraken mogelijk maakt die de specifieke steekproeven van items en proefpersonen overstijgen.

Hoofdstuk 3 richt zich op de inferentiële processen die betrokken zijn bij het interpreteren van causale coherentierelaties. Het doel van het leesexperiment dat hier wordt gepresenteerd is om eerder gevonden integratie- en inferentie-effect van connectieven te replisseren. Het bestaan van deze effecten is immers cruciaal voor de CPA die gepresenteerd wordt in hoofdstuk 1. In het experiment lazen proefpersonen korte nieuwsberichten. Elke bericht bevatte 1 causale relatie. De relatie bleef impliciet of werd gemaakte met de connectieven want of omdat. Na elke tekst kregen proefpersonen een bewering aangeboden die de informatie bevatte die verondersteld werd te zijn geactiveerd tijdens het interpreteren van de
zin. Proefpersonen werd gevraagd om zo snel mogelijk aan te geven of de informatie waar
was op grond van wat ze in de tekst gelezen hadden.
De resultaten laten zien dat er geen integratie-effect optrad en dat ook op de verificatietijden
geen effect van de connectieven gedemonstreerd kon worden. Er was wel een indicatie van
een inferentie-effect, maar dit effect deed zich alleen voor bij het connectief want.
De conclusie die in het hoofdstuk wordt getrokken is dat er blijkbaar niet per definitie een
versnellingseffect van connectieven optreedt. Het uitblijven van een vertragingseffect voor
omdat zou kunnen worden verklaard door de ambiguïteit van omdat wat betreft de relatie die
wordt gemarkandeerd. Het connectief kan immers gebruikt worden voor zowel objectieve als
subjectieve causale relaties. Door deze ambiguïteit is het mogelijk dat de lezers van de omdat-
zinnen net als lezers van de impliciete versies tijdens het leggen van het causale verband de
extratekstuele informatie hebben geactiveerd. Daardoor zou de ETI net zo snel kunnen
worden uitgevoerd als in de impliciete conditie. De variantieschatten in de omdat-conditie
lijken in overeenstemming met deze verklaring.

In hoofdstuk 4 wordt de aanname van de CPA wat betreft de niveaus van representaties nader
onderzocht: (expliciete) causale coherentierelaties zijn op zowel textbase als op situatie-model
niveau gepresenteerd. Deze aanname is experimenteel onderzocht door het connectief te
manipuleren dat gebruikt wordt om de relatie aan te duiden.
De proefpersonen lasen nieuwsberichten die causale relaties bevatten. De volgende zin is één
van de zinnen die proefpersonen lasen.

(3) Bashir was in het ziekenhuis opgenomen, omdat zijn hartklachten ernstig waren
toegenomen.

Tijdens een recognitietaak die na het lezen werd afgenomen kregen de proefpersonen de
originele zin voorgelegd of één van drie mogelijke manipulaties van de originele zin. De
eerste manipulatie betrof een parafrase van de originele relatie. Hier werd het connectief
doordat voor gebruikt.

(4) Bashir was in het ziekenhuis opgenomen, doordat zijn hartklachten ernstig waren
toegenomen.

De achterliggende gedachte is dat met deze parafrase wel de surface code van de
oorspronkelijke zin verandert, maar niet de expliciete proppositie die wordt uitgedrukt (het
causale verband) en het situatie-model dat geconstrueerd kan worden op basis van de tekst.
De tweede manipulatie betrof een betekenisverandering op propositioneel niveau. In deze
conditie werd het originele causale connectief vervangen door nadat.

(5) Bashir was in het ziekenhuis opgenomen, nadat zijn hartklachten ernstig waren
toegenomen.
Met deze manipulatie worden zowel de *surface code* als de *textbase* van het origineel veranderd. Het situatie-model is echter nog steeds in overeenstemming met dat van het origineel.

In de laatste conditie is het causale connectief vervangen door *en* en daarmee werd een verandering van het situatie-model geconsttrueerd.

\[(6)\] Bashir was in het ziekenhuis opgenomen, en zijn hartklachten waren ernstig toegenomen.

Door het causale connectief te vervangen door *en* wordt een verandering op alle drie de niveaus van representaties gecreëerd. In deze laatste conditie correspondeert het onderliggende model niet meer met het origineel, omdat een natuurlijke causale interpretatie van (6) is dat Bashir’s hartklachten ernstig waren toegenomen doordat hij in het ziekenhuis was opgenomen.

In de onderstaande tabel worden de verschillen tussen de condities weergegeven.

<table>
<thead>
<tr>
<th>Tabel 1. Niveaus van representatie die een testzin deelt met de originele zin.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface code</td>
</tr>
<tr>
<td>Origineel (omdat)</td>
</tr>
<tr>
<td>Parafrase (doordat)</td>
</tr>
<tr>
<td>Betekenisverandering (nadat)</td>
</tr>
<tr>
<td>Modelverandering (en)</td>
</tr>
</tbody>
</table>

Noot: een + geeft aan dat de testzin het betreffende niveau van representatie deelt met het origineel.

De tabel laat zien dat opeenvolgende condities telkens slechts op één niveau van representatie dat zij delen met het origineel van elkaar verschillen. Zo deelt de parafrase-conditie de *textbase* en het situatie-model met het origineel, maar de betekenisverandering-conditie deelt alleen het situatie-model van het origineel. Doordat de condities telkens maar op één niveau dat zij delen met het origineel van elkaar verschillen, kunnen we de resultaten van proefpersonen op de recognitietaak gebruiken om inzicht te verkrijgen in de niveaus waarop de causale relatie geregistreerd wordt. Met betrekking tot de proportie ‘ja’-reacties op de verschillende zinnen kunnen we verwachten dat de afnemende proporties ‘ja’-reacties zal het hoogst zijn wanneer een origineel wordt aangeboden en het laagst wanneer er sprake is van de zin waarin het model is veranderd.

In het experiment werd de recognitietaak afgenomen nadat de proefpersonen 20 teksten hadden gelezen die in totaal 32 causale relaties bevatten. De resultaten laten zien dat de proportie ‘ja’-reacties in de originele, parafrase en betekenisconditie vrijwel gelijk blijft en dat pas bij de modelverandering-conditie een significante daling van de proportie ‘ja’-reacties plaatsheeft.

Dit betekent dat er wel een indicatie is gevonden voor het bestaan van een situatie-model representatie van de relatie en niet van een *textbase* representatie. Het feit dat de proporties tussen originele zinnen en parafrases nauwelijks afwijken, is te verwachten op grond van de
relatief lange tijd tussen het lezen en de recognitietak. Van de surface code is bekend dat deze doorgaans niet lang in het geheugen bewaard blijft.

In hoofdstukken 3 en 4 stond de CPA centraal die werd ontwikkeld in hoofdstuk 1. Het onderzoek in deze hoofdstukken richt zich op de verwerking en representatie van expliciete causale relaties. In hoofdstuk 5 staat de verwerking van impliciete causale relaties centraal. In het bijzonder worden in dit hoofdstuk twee hypothesen getoetst die verklaren waarom informatie wanneer deze causaal verbonden is met de context sneller wordt verwerkt dan wanneer de informatie met een niet-causale relatie is verbonden met de voorafgaande tekst.

De eerste hypothese is de causality-by-default hypothese. Volgens deze hypothese proberen lezers standaard eerst een causale verbinding te leggen tussen een zin en de voorafgaande context. De lezer probeert pas een andere relatie te leggen wanneer een causale relatie niet mogelijk blijkt. De tweede hypothese is de structural expectations hypothese die veronderstelt dat tekststructuren bij de lezers verwachtingen oproept over de relatie die de inkomende informatie zal hebben met de voorafgaande tekst. Deze verwachtingen zijn gebaseerd op kennis over tekststructuur.

Een voorbeeld van een structuur die sterke verwachtingen oproept is een Probleem-Oplossing structuur, waarbij het lezen van de Probleem-context bij de lezer verwachtingen genereert over de structuur van de rest van de tekst. In dit verband zou de lezer kunnen verwachten dat de tekst een oplossing zal beschrijven. Als de inkomende informatie in overeenstemming is met de tekststructurele verwachtingen van de lezer, dan is het voor de lezer eenvoudiger om de inkomende informatie te verbinden met de voorafgaande context.

In een leesexperiment worden de twee hypothesen van elkaar onderscheiden door identieke testzinnen in drie condities aan te bieden: Probleem-Oplossing (PO), Oorzaak-Gevolg (OG) en Lijst (LI). Een eerste verschil tussen de condities is dat in de PO conditie lezers sterkere tekststructurele verwachtingen genereren dan in de OG en LI condities. De validiteit van deze aanname is voorafgaand aan het leesexperiment in een continueringsexperiment empirisch vastgesteld. Het tweede verschil tussen de condities is dat PO en OG causale relaties zijn en LI een additieve (en dus niet-causale) relatie. Met deze drie condities kunnen de hypothesen worden onderscheiden omdat de causality-by-default hypothese voorspelt dat de informatie in de testzin in de causale condities (PO en OG) sneller zal worden verwerkt dan in de niet-causale conditie (LI), terwijl de structural expectations hypothese voorspelt dat de verwerkingsstijd in de conditie met structurele verwachtingen (PO) korter zal zijn dan in de condities zonder tekststructurele verwachtingen (OG en LI).

Naast de leestijden werden in het experiment ook snelheid en accuratesse op een beweringverificatietaak geregistreerd: de informatie in de testzin werd in geparafraseerde vorm aan de proefpersonen aangeboden en zij moesten beoordelen of deze juist was. De resultaten laten zien dat de informatie in de targetzin het snelst wordt gelezen wanneer deze verbonden is met de context door middel van een PO-relatie. Er werd geen verschil gedetecteerd tussen OG en LI-relaties. De resultaten op de leestijden zijn dus in overeenstemming met de structural expectations hypothese. De resultaten laten verder zien dat de informatie in de targetzin beter wordt onthouden in de causale condities (PO en OG) dan in de niet-causale conditie (LI). Dit blijkt zowel uit de verificatietijden als uit de...
accuratesse. Deze resultaten laten het belang zien van causaliteit voor de representatie die lezers construeren op basis van de tekst.

In hoofdstuk 6 worden de resultaten van de afzonderlijke experimenten samengevat en wordt er, nadat rekenschap is gegeven van de beperkingen van het onderzoek, een algemene conclusie getrokken die de resultaten van de afzonderlijke experimenten met elkaar verbindt. Daarnaast wordt ingegaan op de consequenties die de resultaten van het onderzoek kunnen hebben.

Geconcludeerd wordt dat de CPA die in hoofdstuk 1 werd ontwikkeld slechts gedeeltelijk ondersteund wordt door de experimentele resultaten. In de eerste plaats kon een integratie-effect van connectieven niet gedemonstreerd worden. Er werd echter wel een indicatie gevonden voor een vertraagings-effect van connectieven. Dit effect heeft hoogstwaarschijnlijk te maken met het in de representatie opnemen van de extra-tekstuele informatie die aan de causale relatie ten grondslag ligt. In de CPA wordt aangenomen dat dit proces betrokken is bij de constructie van het situatie-model.

In de tweede plaats werd er geen evidentie gevonden voor een textbase representatie van de causale relaties. Wat betreft het niveau van representatie van causale relaties kan daarom worden geconcludeerd dat causale relaties gerepresenteerd worden op het niveau van het situatie-model. De aanname is in overeenstemming met verschillende psychologische theorieën over tekstverwerking en –representatie.

De conclusie dat causale coherentierelaties op situatie-model niveau gerepresenteerd worden is ook in overeenstemming met de resultaten die gerapporteerd worden in hoofdstuk 5. Uit dat hoofdstuk komt naar voren dat lezers kennis over tekststructuren gebruiken bij het interpreteren van het type coherentierelaties. De interpretatie van de relatie komt tot stand door een combinatie van expliciete tekst en achtergrondkennis van de lezer. Dit integreren van expliciete informatie in de tekst en de achtergrondkennis van de lezer is een cruciaal onderdeel van de constructie van situatie-modellen. De nieuwe resultaten en inzichten die in dit proefschrift worden gerapporteerd zijn daarmee van belang voor een cognitieve theorie van de verwerking en representatie van causale coherentierelaties.
### Appendices

**A. Example Data and SPSS syntax used in chapter 2**

<table>
<thead>
<tr>
<th>Condition 1</th>
<th>Condition 2</th>
<th>Condition 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP</td>
<td>T1</td>
<td>T2</td>
</tr>
<tr>
<td>1.00</td>
<td>1.27</td>
<td>-.91</td>
</tr>
<tr>
<td>2.00</td>
<td>.77</td>
<td>1.08</td>
</tr>
<tr>
<td>3.00</td>
<td>1.24</td>
<td>-.59</td>
</tr>
<tr>
<td>4.00</td>
<td>2.46</td>
<td>-1.03</td>
</tr>
<tr>
<td>5.00</td>
<td>.56</td>
<td>-.88</td>
</tr>
<tr>
<td>6.00</td>
<td>.83</td>
<td>.71</td>
</tr>
<tr>
<td>7.00</td>
<td>2.45</td>
<td>1.41</td>
</tr>
<tr>
<td>8.00</td>
<td>.22</td>
<td>-.12</td>
</tr>
<tr>
<td>9.00</td>
<td>.16</td>
<td>-.59</td>
</tr>
<tr>
<td>10.00</td>
<td>1.55</td>
<td>.40</td>
</tr>
<tr>
<td>11.00</td>
<td>-.124</td>
<td>-.30</td>
</tr>
<tr>
<td>12.00</td>
<td>.77</td>
<td>-.24</td>
</tr>
<tr>
<td>13.00</td>
<td>1.57</td>
<td>.09</td>
</tr>
<tr>
<td>14.00</td>
<td>.46</td>
<td>1.43</td>
</tr>
<tr>
<td>15.00</td>
<td>.14</td>
<td>-.59</td>
</tr>
</tbody>
</table>

**SPSS SYNTAX**

```spss
MANOVA t1 t2 t3 t4 t5 t6 t7 t8 t9 t10 t11 t12 t13 t14 t15
  /WSFACTOR = treat (3) item (5)
  /contrast (item)=difference
  /PRINT = cellinfo (means)
  /WSDESIGN = treat item treat*item.

MANOVA t1 t2 t3 t4 t5 t6 t7 t8 t9 t10 t11 t12 t13 t14 t15
  /WSFACTOR = treat (3) item (5)
  /contrast (item)=difference
  /PRINT = cellinfo (means)
  /PRINT = signif (efsize)
  /WSDESIGN = item within treat.
```
Profielschetsen verdachten aanslag Bali

De Indonesische politie heeft profielschetsen opgesteld van drie mogelijke verdachten van de bomaanslag op Bali, die bijna 200 mensen het leven heeft gekost. De politie maakt de profielschetsen niet openbaar. Het is alleen bekend dat het om drie Indonesische mannen gaat. De profielschetsen konden worden opgesteld nadat de politie enkele getuigen had verhoord. De profielschetsen dienen in eerste instantie alleen voor intern gebruik bij de politie. Onderzoeksleider Pastika benadrukte dat de mannen nog niet als verdachte zijn aangemerkt. Hij weigerde de religieuze achtergrond van de mannen bekend te maken, omdat dat “het onderzoek zou kunnen schaden”.

De politie vermoedt dat de militante moslimgroepering Jamaah Islamiyah betrokken is bij de aanslag op Bali. De Indonesische politie wacht nog steeds op de gelegenheid om de vermoedelijke leider van deze groep, Abu Bakar Bashir, te verhoren. Bashir werd afgelopen weekeinde aangehouden in het ziekenhuis van Solo, op Java. Bashir was in het ziekenhuis opgenomen, omdat zijn hartklachten ernstig waren toegenomen. De politie wil Bashir horen over de aanslag op Bali. Hij werd al verdacht van terroristische activiteiten en die verdenking werd des te waarschijnlijker na de aanslag op Bali. Zo vermoedt de politie dat Bashir betrokken is bij het beramen van een aanslag op Megawati Soekarnoputri in de tijd dat zij vice-president was. Ook wordt hij verdacht van betrokkenheid bij aanslagen op kerken in Indonesië in 1999. Bashir ontkent tot nu toe elke betrokkenheid bij Jemaah Islamiyah en bij elke vorm van terrorisme.

Zeker honderd doden bij brand in Vietnam

Bij de brand van gisteren in een gebouwencomplex in het Zuid-Vietnamese Ho Chi Minh-Stad zijn meer dan honderd mensen om het leven gekomen. Aanvankelijk werd nog over vijftig doden gesproken. Onder de slachtoffers zouden volgens diplomaten twee Britten en een Amerikaan zijn.

In het gebouwencomplex waar de brand uitbrak, waren een internationaal zakencentrum, een discotheek en een aantal winkels gevestigd. De oorzaak van de brand is nog onduidelijk. Ooggetuigen vermoedden dat de brand op de bovenverdieping ontstond, doordat er kortsluiting was opgetreden.

Van de honderd doden zijn er nog maar een paar geïdentificeerd. De Vietnamezen autoriteiten vrezen dat het dodental nog zal oplopen, omdat er 500 mensen in het gebouwencomplex aanwezig waren toen de brand uitbrak. De brandweer had de grootste moeite het vuur onder controle te krijgen, nadat de watervoorraad was opgeraakt. Tientallen mensen konden het gebouw pas verlaten, nadat de brandweer een geschikte ladder had gevonden.
Curriculum Vitae

Gerben Mulder was born in Ede on March 11 1975. He attended ‘Het Wagenings Lyceum’ in Wageningen, where he graduated in 1993. In 1999, he received his MA (cum laude) Algemene Letteren (Liberal Arts) at Utrecht University. In 2000, after being employed as a research assistant at the Utrecht Institute of Linguistics OTS, he started working as a PHD-student at the same institute. In 2004, he started working as a lecturer in Methodology and Statistics at the Faculty of Arts of the Free University (VU) in Amsterdam, where he is currently employed as assistant professor.