

Coherence Marking, Comprehension and Persuasion

On the processing and representation of discourse

Published by
LOT
Janskerkhof 13
3512 BL Utrecht
The Netherlands

phone: +31 30 253 6006
fax: +31 30 253 6406
e-mail: lot@let.uu.nl
<http://www.lotschool.nl>

Cover illustration: 1914 advertisement for Ford Canada. Reproduced with kind permission of Ford Nederland BV.

ISBN 978-90-78328-29-2
NUR 616

Copyright © 2007: Judith Kamalski. All rights reserved.

Coherence Marking, Comprehension and Persuasion

On the processing and representation of discourse

Coherentiemarkering, begrip en overtuigingskracht

Over het verwerken en representeren van discourse
(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. W.H. Gispen, ingevolge het besluit van
het college voor promoties in het openbaar te verdedigen op vrijdag 31
augustus 2007 des middags te 4.15 uur

door

Judith Maria Helena Kamalski

geboren op 2 september 1979
te Geldrop

Promotor: Prof. T.J.M. Sanders

Co-promotor: Dr. L. R. Lentz

Part of the research reported in this dissertation was supported by a travel grant of the Netherlands Organization for Scientific Research (N.W.O).

Table of contents

Acknowledgements	11
1 Introduction: coherence marking	13
1.1 Coherence relations.....	14
1.1.1 Basic operation, polarity, order of the segments, source of coherence	16
1.2 Linguistic marking of coherence relations.....	18
1.2.1 Linguistic coherence marking.....	19
1.2.2 Non-linguistic coherence marking.....	21
1.3 Functions of coherence marking.....	23
1.4 Coherence marking in this dissertation.....	24
1.5 Main research question.....	25
2 Effects of coherence marking and prior knowledge on the situation model representation: previous research.....	27
2.1 Text representation	28
2.1.1 Levels of mental representation	28
2.1.2 Coherence marking at the situation model level	32
2.1.3 Effects of coherence marking on the mental representation	33
2.1.4 Methodological options to measure the situation model representation	36
2.2 Extending the view on mental representation.....	45
2.2.1 Appraisal	46
2.2.2 Effects of coherence marking on appraisal.....	48
2.2.3 Meta-cognition	49
2.2.4 Effects of coherence marking on meta-cognition.....	50
2.3 Persuasion.....	52
2.3.1 Persuasion.....	52
2.3.2 Coherence marking in persuasive contexts	53
2.3.3 Distinction between informative and persuasive genre.....	55
2.3.4 O’Keefe’s common features of persuasion.....	55
2.3.5 Biber’s textual dimensions.....	56
2.3.6 Other prototypical characteristics of persuasive texts.....	57

6 – TABLE OF CONTENTS

2.3.7 Pragmatic relations and pragmatic argumentation	59
2.4 Innovative aspects of this dissertation.....	61
3 A Pilot Experiment: comprehension, appraisal and persuasion	63
3.1 Variables in Pilot Experiment 1	64
3.1.1 Coherence marking and prior knowledge.....	64
3.1.2 Comprehension.....	65
3.1.3 Appraisal and Feeling of Knowing.....	65
3.1.4 Persuasion.....	66
3.2 Pilot Experiment 1	67
3.2.1 Materials	67
3.2.2 Pretest.....	69
3.2.3 Dependent measures.....	70
3.2.4 Participants	73
3.2.5 Distinction informative-persuasive.....	73
3.2.6 Design.....	74
3.3 Results	75
3.3.1 Comprehension.....	75
3.3.2 Text appraisal	76
3.3.3 Feeling of Knowing.....	77
3.3.4 Persuasion.....	78
3.4 Conclusion and Discussion	79
3.4.1 Conclusion.....	79
3.4.2 Discussion.....	80
4 How to measure situation model representations.....	83
4.1 The Validation Experiment 2.....	85
4.1.1 The selection of four situation model tasks	86
4.1.2 Hypotheses	86
4.1.3 Materials	88
4.1.4 Situation model task construction	88
4.1.5 Participants and procedure.....	90
4.1.6 Design.....	91
4.1.7 Scoring.....	92
4.2 Results	92
4.2.1 Method of analysis.....	93

TABLE OF CONTENTS - 7

4.2.2 Descriptives and reliability of the operationalizations.....	93
4.2.3 Convergent validity.....	96
4.2.4 Known group validity	96
4.2.5 Divergent validity.....	98
4.2.6 Post-hoc analysis of the cloze tasks.....	99
4.3 Conclusion and Discussion	102
4.3.1 Question asking	102
4.3.2 Cloze task.....	103
4.3.3 Sorting task	103
4.3.4 Mental Model task.....	103
4.4 Discussion	104
5 Effects of coherence marking comprehension, appraisal and FOK revisited.....	107
5.1 Operationalization of prior knowledge.....	108
5.2 Experiment 3: comprehension, appraisal, FOK	109
5.2.1 Hypotheses	110
5.2.2 Materials	111
5.2.3 Participants	113
5.2.4 Design.....	114
5.2.5 Results on comprehension.....	114
5.2.6 Results on appraisal.....	115
5.2.7 Results on FOK.....	115
5.2.8 Conclusion for comprehension.....	116
5.2.9 Conclusion for appraisal and FOK	117
5.3 Experiment 4: Comprehension reconsidered.....	117
5.3.1 Materials	118
5.3.2 Participants	119
5.3.3 Design and procedure.....	119
5.3.4 Results	119
5.4 Conclusion.....	121
5.5 Additional analyses.....	122
5.6 Comprehension: general discussion	123
6 Coherence marking and persuasion: forewarning.....	127
6.1 Markers of coherence in persuasive texts: summary of the pilot	129

8 – TABLE OF CONTENTS

6.2 Forewarning	130
6.3 Step 1: Detection of subjectivity	131
6.4 Step 2: Recognizing and identifying the author’s intent.....	133
6.5 Step 3: Reacting to the identified author’s intent	135
6.6 Step 4: Maintaining the original attitude and making it stronger.....	137
6.7 Conclusion.....	139
7 The forewarning effect: evidence from persuasion and processing ...	141
7.1 The forewarning hypothesis for coherence marking.....	142
7.2 Experiment 5: off-line effects of coherence marking on persuasion.....	143
7.2.1 Materials	143
7.2.2 Participants	146
7.2.3 Procedure.....	147
7.2.4 Design.....	147
7.3 Results	147
7.3.1 Effects of coherence marking on persuasion.....	147
7.3.2 Effects of coherence marking on detection of persuasive intent	149
7.4 Discussion	149
7.5 Experiment 6.....	151
7.5.1 Hypotheses	151
7.5.2 On-line hypotheses.....	151
7.5.3 Off-line hypotheses.....	153
7.5.4 Materials	153
7.5.5 Procedure.....	155
7.5.6 Participants	156
7.6 Results	156
7.6.1 On-line effects of coherence marking.....	156
7.6.2 Off-line effects of coherence marking.....	159
7.7 Discussion	164
7.8 General Discussion	166
7.9 The four steps of the forewarning process	168
8 Conclusion and Discussion	171
8.1 Situation model representations and prior knowledge.....	172
8.1.1 Scope of the situation model representation.....	173
8.1.2 Situation model representation methodology	173

TABLE OF CONTENTS - 9

8.2 Coherence marking	176
8.2.1 Definition of coherence marking.....	176
8.2.2 Implications for coherence marking theory	183
8.3 Forewarning	184
8.4 Implications for Document Design	188
8.4.1 Educational context	188
8.4.2 Marketing context.....	190
8.5 Discussion	191
8.5.1 Implications for situation model theory	191
8.5.2 Limitations	193
8.5.3 Indications for future research	194
8.6 Epilogue.....	199
References	200
Samenvatting in het Nederlands	211
Appendix 1: (Part of) Materials from Pilot Experiment 1	221
Appendix 2: (Part of) Materials from Experiment 2.....	222
Appendix 3: (Part of) Materials from Experiments 3, 4 and 5.....	223
Appendix 4: (Part of) Materials from Experiment 4 in Chapter 7	225
Curriculum Vitae	227

Acknowledgements

“Wetenschap is suspense. Je start je onderzoek op basis van ideeën, maar je weet vantevoren niet wat je tegenkomt”.

Peter Hagoort, De Volkskrant, May 5, 2005.

The life of a junior researcher sometimes resembles that of a detective: always trying to solve mysteries and discover new things. I thoroughly enjoyed doing so, mostly because of the inspiring people I was fortunate enough to meet. Many people have contributed in some way to this dissertation and I thank them all. However, I would like to say a special thank you to:

- Ted Sanders and Leo Lentz, for their insightful feedback and motivational guidance;
- Rolf Zwaan, for giving me the opportunity to spend a great semester in his lab in Tallahassee and learn so much from him;
- Bernard Bichakjian, Margot van Mulken, Piet-Hein van de Ven, and Bregje Holleman, for being true mentors to me;
- my parents, for encouraging me to learn and to enjoy the opportunities that life has to offer;
- Thomas, for simply everything;
- my research friends, specifically Daphne, Judith, Jos, Lisa, Cari, Tres and Johanna, for making it so much fun to discuss our work (and more); I am especially grateful to Cari, Lisa and Tres who have been wonderful ‘Englishness editors’;
- my non-research friends, specifically Marjan, Marie-Laure and my sister Digna, for their friendship and support;
- the teachers at the high schools who allowed me to conduct the large experiment in chapter 4: Gonnie Poell, Joke Methorst, Danielle van Dijk and Ton Jansen;
- the bachelor students who collected the data for experiment 4;
- N.W.O. and UiL-OTS, for financially supporting my international research experiences in Tallahassee, Vancouver, Minneapolis and Toulouse.

12 – ACKNOWLEDGEMENTS

Some say that writing a dissertation is a solitary, predominantly individual activity. While writing, you are supposed to vanish from the face of the earth to do your work in isolation. This has definitely not been the case for me. I am grateful to have met so many inspiring people who have helped me uncover some of the mysteries of coherence marking, comprehension and persuasion. In a way, this dissertation represents a full circle: the communication with so many people resulted in this dissertation about how people effectively understand, persuade, and communicate.

Judith

April 2007

DOVE- SOFT TOUCH

Your skin's natural oils keep it silky and supple. **But** as you age, **your skin** becomes less elastic and the production of oil slows down. **That is why** aging can cause dull, dehydrated skin. **So** it's essential to replenish the lost moisture with a natural, soothing alternative. Dove leaves your body clean and smooth, **because** Dove contains ten times more natural oil than regular shower gels.

DOVE- SOFT TOUCH

Your skin's natural oils keep it silky and supple. As you age, **it** becomes less elastic and the production of oil slows down. Aging can cause dull, dehydrated skin. It's essential to replenish the lost moisture with a natural, soothing alternative. Dove leaves your body clean and smooth. **It** contains ten times more natural oil than regular shower gels.

14 – CHAPTER 1

These two short text passages are based on an advertisement for Dove shower gel. The left column obviously differs from the right column, but these differences seem to be very small: in fact, the only difference is the use of words like *but*, *this is why* and *so*. These so-called **coherence markers** do occur in the left column, but they are left out in the right one.

Which text makes it easier to understand why this product is good for your skin? Which text do readers prefer? Which text is considered to be more appealing? Which text results in better understanding of Dove's advantages? And last but not least: Which text is more likely to make the reader want to buy and use Dove shower gel?

These are a few examples of the questions that are central to the work presented in this dissertation. Small textual characteristics such as coherence markers can have clear effects on readers. Not only on what they understand of the information in a text, but also on their opinions, and maybe even on their behavior. In this dissertation, I investigate different effects that coherence marking may or may not have on the reader.

1.1 Coherence relations

The questions that were raised in the previous section are typical for discourse studies, and some areas of document design. One of the major goals in these disciplines is to develop theoretical and empirical knowledge on the relation between discourse characteristics, such as coherence marking, and the effects they may have on the reader. Several methods are commonly used to investigate influences of discourse characteristics on the reader: by means of an (corpus) analysis of discourse, by studying effects on text processing (during reading) and by studying effects on the reader's mental representation (after reading). A cognitive approach to discourse studies, such as the approach in this dissertation, aims at providing an account for the relations between discourse and its object, its cognitive representations and the cognitive processes of production and interpretation that form the interface between the linguistic characteristics and the cognitive representations (Sanders & Spooren, 2001). A cognitive approach focuses on the question of how specific textual characteristics can influence the reader's mental representation. Processing effects will also be discussed, but the focus will be on the mental representation.

When we study the relation between text and reader, the notion of *coherence* is crucial. Understanding discourse means that readers construct a mental representation of the information in the text. Such a mental representation needs to be coherent, meaning that representations of the segments in the discourse are linked to one another (cf. Hobbs, 1979; Sanders, Spooren, & Noordman, 1992).

In order for a reader to construct a coherent mental representation of the information in the Dove text from the beginning of this chapter, the representations of several text segments need to be linked. Text segments are meaning units, which can be larger discourse units such as paragraphs (on a global text level), but can also be sentences or minimally clauses (on a local level). If we look for instance at the sentences: ‘But as you age, your skin becomes less elastic and the production of oil slows down. That is why aging can cause dull, dehydrated skin’, four different meaning segments can be distinguished: segment 1, the clause [*s₁ as you age*], the second text segment, the clause [*s₂ your skin becomes less elastic*], the third text segment [*s₃ the production of oil slows*] and the fourth segment [*s₄ aging can cause dull, dehydrated skin*]. What readers have to do in order to construct a coherent representation, is to link the concept of [*aging*] to the consequences that this can have for the skin, namely [*less elasticity*] and [*less oil production*]. Ultimately, this results in dry skin. Schematically, readers have to link these representations of information in the text in a way that resembles the following causal chain:

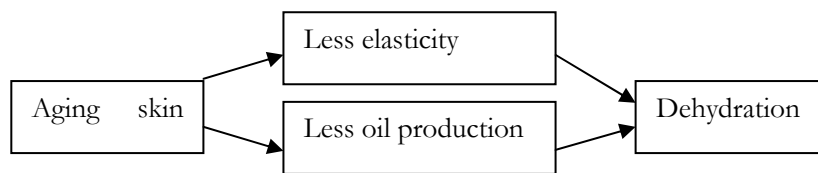


Figure 1.1: *Schematic representation of causal chain in the Dove example*

If readers do not see that these concepts are linked, they have failed to understand (part of) the information in the text. This is why coherence is so crucial for understanding and comprehension: it is what makes a whole string of separate sentences into a *text*. Or, as Graesser, McNamara and Louwerse (2003: 82) put it, coherence is commonly considered ‘the cornerstone of comprehension’.

Generally speaking, we distinguish two levels of coherence in discourse. The first level is that of referential coherence, where units are connected by repeated reference to the same object. In the Dove text, we can see an example of referential coherence in the beginning of the text, where we see the reference to *your skin*, in the right column replaced by *it*, in the left column repeatedly referred to as *your skin*. The sentences in the text are related, because they refer to the same object.

The second level of coherence in discourse is that of relational coherence, where text segments are connected by establishing coherence relations between them (Sanders & Spooren, 2001). If we have another look at the causal chain represented in Figure 1.1, a reader who constructs a coherent mental representation has to link these concepts in a causal relation: there is CAUSE-CONSEQUENCE relation between [s₁] and [s₂] and a comparable causal relation between [s₁] and [s₃]. Or, in other words, the representation of segment 1, the clause [s₁ *as you age*] needs to be related to the representation of the second text segment, the clause [s₂ *your skin becomes less elastic*] and the third text segment [s₃ *the production of oil slows*].

Different types of coherence relations exist, such as CAUSE-CONSEQUENCE, LIST or PROBLEM-SOLUTION. Much work has been done on the categorization of coherence relations (among many others Knott & Dale, 1994; Mann & Thompson, 1988; Sanders, Spooren, & Noordman, 1992). The categorizations that these studies propose differ considerably from one another. I will make use of the classification by Sanders et al. (1992) in this dissertation. Sanders et al. (1992) argue for the classification of coherence relations by means of four cognitive primitives: *basic operation*, *polarity*, *order of the segments*, and *source of coherence*. I will briefly introduce these four primitives in the next section.

1.1.1 *Basic operation, polarity, order of the segments, source of coherence*

The first primitive, *basic operation*, distinguishes between additive and causal relations. An additive relation (example 1) exists if there is a relation of logical conjunction between the two discourse segments ($P \ \& \ Q$). A causal relation (example 2) exists if an implication relation can be deduced ($P \rightarrow Q$).

1. I went to the hockey game. My friends also went to the game.
2. I decided to go to the hockey game, so my friends decided to come along.

The second primitive, *polarity*, distinguishes between positive and negative relations. A relation is positive (example 3) when both discourse segments occur in the basic operation as *P* and *Q*. When these segments do not occur, but their negative counterparts do, the relation is negative (example 4).

3. There were many fights, so it was an interesting hockey game.
4. Although there were no fights, it was an interesting hockey game.

The third primitive, *order of the segments*, distinguishes between basic order and non-basic order. The order is basic (example 3) when segment 1 expresses *P* and segment 2 expresses *Q*, and vice versa for the non-basic order (example 5).

5. It was an interesting hockey game, because there were many fights.

Source of coherence, the fourth primitive, distinguishes between semantic and pragmatic relations. A relation is semantic when discourse segments are related because of their propositional content. A relation is pragmatic when discourse segments are related because of the language user's goal oriented communicative act; the connection is at the illocutionary level. Example 6 shows a semantic relation. The relation between a hockey game being interesting and the occurrence of many fights, is a pragmatic one (7). One way in which pragmatic relations differ from semantic relations is that whereas semantic relations, such as in example 6, are a matter of *facts*, whereas pragmatic ones, such as in example 7, are a matter of *opinion*.

6. There was a huge fight. One of the players was injured.
7. There was a huge fight. It was a very interesting game.

In the specific context of this dissertation, *source of coherence*, is the most relevant one, for reasons that I will explain at a later stage in this dissertation. Several researchers have made the same pragmatic-semantic distinction, although often making use of slightly different definitions. The semantic-pragmatic distinction resembles Halliday and Hasan's (1976) internal versus external relations, Sweetser's (1990) content versus epistemic and speech act relations, or Pander Maat's objective versus subjective relations (Pander Maat & Degand, 2001; Pander Maat & Sanders, 2001).

In Chapters 2 to 5 in this dissertation, I will generalize over different coherence relations, focusing on the way these relations are marked linguistically and how this influences the reader. Hence, the focus throughout the dissertation is on the effects that a specific textual characteristic, namely coherence marking, can have on the reader. My working hypothesis is that there is no immediate reason to assume that the marking of different coherence relations has different effects on the mental representation. They are therefore investigated collectively. In Chapter 6, the distinction between different types of coherence relations will be addressed again, and I will return to the distinction between semantic and pragmatic relations.

1.2 Linguistic marking of coherence relations

We have seen in the previous section that coherence is a crucial notion for discourse studies. We have also seen that different types of coherence relations exist in discourse, such as CAUSE-CONSEQUENCE. Such a coherence relation can be made linguistically explicit in a text by means of a coherence marker. This is what we have seen in the Dove example: the left column makes use of coherence markers to explicitly mark coherence relations, whereas the right column does not use these markers. An example is the following passage: ‘As you age, it becomes less elastic and the production of oil slows down. *That is why* aging can cause dull, dehydrated skin.’ The coherence marker *that is why* marks the CAUSE-CONSEQUENCE relation between aging and dehydrated skin.

Linguistic coherence markers such as *that is why* or *because* can be taken as instructions for readers and listeners to know how to connect the new discourse segment with the previous one (Britton, 1994; Sanders & Noordman, 2000). Again, segments can be larger discourse units such as paragraphs, but also sentences and clauses. Coherence markers explicitly guide the reader in interpreting the text and in connecting ideas with other ideas. Coherence marking can therefore play an important role when readers construct meaning from discourse. Evidence for this role of coherence markers comprises both the effects on the mental representation (see Sanders & Spooren, 2006 for a recent overview) and the effects on text processing (recent examples are Cozijn, 2000; Maury & Teisserenc, 2005). These empirical studies on the effects of coherence marking on the processing and representation will be discussed in detail in Chapter 2.

What exactly are the linguistic tools that can mark coherence relations in a text? There are many different views on this issue, depending on how strict or how free the interpretation of the definition of coherence marking is. The definition used in this dissertation is: 'linguistic markers that instruct readers how to connect the new discourse segment with the previous one'. Even when coherence marking is not present in the text, the relation between the discourse segments can still be inferred by readers. In the following sections, I will further elaborate on different aspects of this definition. I will list the coherence markers that fall within the scope of this dissertation, and the ones that do not fall within my scope.

1.2.1 Linguistic coherence marking

Generally, linguistic markers can signal global or local coherence. Local coherence relates discourse segments such as clauses and sentences. Or, in Murray's (1995: 107) terms, local coherence concerns 'connecting the phrase or sentence that is currently being read with the contents of the immediately preceding sentence.' Establishing local coherence can involve factors such as argument overlap, causal inferences and co-reference mechanisms. Global coherence relates segments such as paragraphs and sections. This type of coherence concerns organization and hierarchical relations of the propositions available in the text (Kintsch, 1998). Both types of coherence can be explicitly marked by linguistic tools.

Murray (1995) gives examples of explicit marking of local coherence. Local coherence can be marked explicitly by sequence markers (*first, next, finally*), relevance indicators (*in summary*) and connectives (*and, therefore, but*). A repeated reference instead of using an anaphor also marks textual coherence. In addition, lexical cue phrases (*that's why, the reason for this is*) can also mark local coherence.

Global coherence relates larger discourse units, such as paragraphs and passages. Global coherence can be marked by headings and by organizers such as *in the following paragraph, we will discuss x* (Lorch & Lorch, 1986; Lorch, Lorch, Ritchley, McGovern, & Coleman, 2001).

The main criterion for coherence marking that I use throughout this dissertation is that a linguistic coherence marker explicitly signals a coherence relation, but otherwise does not add any extra content to the text. Or, in other words, when the marker is absent, the relation between the segments can be

20 – CHAPTER 1

inferred by readers when they understand the text, even though it remains unmarked. The relation can be present in the mental representation that readers construct, even though it is not explicitly marked in the text. The following short text passage on the history of the French Eiffel Tower provides examples for all categories of linguistic coherence marking (coherence markers are italicized). The first text passage is the explicit one, where coherence markers are more often used than in the second text passage, which we refer to as the implicit text. Table 1.1 (on the next page) provides an overview and categorization of different types of markers.

8. History of the Eiffel Tower

The Eiffel Tower was erected for the Paris Exposition of 1889. Of the 700 proposals submitted in a design competition, one was unanimously chosen. *The proposal* was a radical creation from the French structural engineer Alexandre Gustave Eiffel. The controversial tower elicited some strong reactions, *for different reasons*. *The first reason* was that some people— including Maupassant and Zola— found the Eiffel Tower useless and monstrous. *Second*, nature lovers thought that it would interfere with the flight of birds over Paris. *For these reasons*, it was almost torn down in 1909, *but* was saved because of its antenna - used for telegraphy at that time. *Still*, the Tour Eiffel is now completely accepted by French citizens, *and* is internationally recognized as one of the symbols of Paris itself.

9. The Eiffel Tower was erected for the Paris Exposition of 1889. Of the 700 proposals submitted in a design competition, one was unanimously chosen. *It* was a radical creation from the French structural engineer Alexandre Gustave Eiffel. The controversial tower elicited some strong reactions. Some people— including Maupassant and Zola— found the Eiffel Tower useless and monstrous. Nature lovers thought that it would interfere with the flight of birds over Paris. It was almost torn down in 1909. It was saved because of its antenna - used for telegraphy at that time. The Tour Eiffel is now completely accepted by French citizens. It is internationally recognized as one of the symbols of Paris itself.

Again, the types of coherence markers in Table 1.1 are only distinguished when they are expected to have a different effect on the reader. For now, the assumption is that effects do not differ from connectives to lexical cue phrases, or headings from organizers. All these markers make coherence relations in the text explicit, thereby making it supposedly easier for readers to infer these relations.

Category	Explicit text	Implicit text
Headings	<i>History of the Eiffel Tower</i>	- (no heading)
Organizers	<i>For different reasons</i>	- (no organizer)
Connectives	It was almost torn down in 1909, <i>but</i> was saved because of its antenna.	It was almost torn down in 1909. It was saved because of its antenna.
Lexical cue phrases	<i>The first reason was that</i> some people— including Maupassant and Zola— found the Eiffel Tower useless and monstrous.	Some people— including Maupassant and Zola— found the Eiffel Tower useless and monstrous.
Referential coherence	Of the 700 proposals submitted in a design competition, one was unanimously chosen. <i>The proposal</i> was a radical creation...	Of the 700 proposals submitted in a design competition, one was unanimously chosen. <i>It</i> was a radical creation...

Table 1.1 *Examples of linguistic marking of coherence*

1.2.2 *Non-linguistic coherence marking*

In the previous section, I listed examples of coherence markers that are part of the scope of this dissertation. In this section, I will discuss examples of manipulations of coherence marking that are *not* within the scope of this dissertation. Often, when coherence marking is being investigated, many more different textual aspects are varied at the same time, such as added descriptions or explanations. In a very broad sense, these textual aspects might be considered to mark coherence, because they help the reader in linking two text segments. For instance, an additional explanation can in some cases help to understand why two text segments are related. The following examples, taken from McNamara, Kintsch, Songer and Kintsch (1996), illustrate this type of manipulation. First, we have a look at the explicit, fully coherent version (10). McNamara et al italicized the marked differences with the implicit version (11).

10. A congenital heart disease is a defect that *a baby* is born with. Most babies are born with perfect hearts. *But* one in every 200 *babies is born with a bad heart. For example, hearts have flaps, called valves, that control the blood flow between its chambers.* Sometimes a valve develops the wrong shape. It may be too tight, or fail to close properly, *resulting in congenital heart disease.*

11. A congenital heart disease is one that a person is born with. Most babies are born with perfect hearts. In about one in every 200 cases, something goes wrong. Sometimes a valve develops the wrong shape. It may be too tight, or fail to close properly.

Additions to the text like the connective *but*, which make the coherence relation explicit, are considered linguistic manipulations of coherence relations. The same contrastive relation exists between *babies with perfect hearts* and *babies with bad hearts*, whether we use the connective *but* to mark this relation or not.

However, many of the other manipulations that we see here in this example, such as the sentence *For example, hearts have flaps, called valves, that control the blood flow between its chambers*, add extra propositional information to the text, thereby changing the content. One can choose to consider such a manipulation as a coherence marker, because it explains a notion that a reader needs to understand in order to establish coherence. Still, I do not consider this additional content a linguistic tool to mark coherence relations. An explanation adds additional content to the text and therefore, this type of markers is not part of the coherence markers studied in this book.

Another example of the same broadness of the notion of coherence marking can be found in a study by Long, Wilson, Hurley, & Prat (2006). In their experimental conditions, the high coherence version differs considerably from the low-coherence version, in more ways than simply in the linguistic way that coherence is marked. The experimental materials in this study concern a specific Star Trek episode. Some of the many aspects that are manipulated are names (*Enterprise* versus *Starship Enterprise*), explanations about interpersonal relationships (*Lwaxana* versus *Lwaxana Troi, Deanna's mother*), vocabulary (*come aboard* versus *beam aboard*) and so forth. Again, these manipulations explain, elaborate and add information, instead of simply linguistically marking coherence relations. The linguistic criterion that I use in this dissertation (a linguistic coherence marker explicitly signals a relation, but otherwise does not add any extra content) clearly shows that these examples are not part of the category of markers that I investigate.

Distinguishing between these two operationalizations of coherence marking is crucial in order to develop the existing theory on coherence marking, because it is the only way to systematically study the effects that the textual characteristic coherence marking may have on the mental

representation. When the content of a text is varied between conditions, it becomes impossible to decide whether the effects we observe on the mental representation are caused by a textual characteristic, or by the additional content of the text. This leads to a possible confounding effect which makes it more difficult to draw any conclusions on the effects that coherence marking has on the reader.

1.3 Functions of coherence marking

In the previous paragraphs, we have seen what coherence marking *is*. But how does it function? What effects does it have on the reader? Coherence marking is known to have two distinct functions that I will briefly describe in this section.

The first function of coherence marking is the *integration* function of coherence marking (Noordman & Vonk, 1997). When a connective like *because* signals a causal relation, readers do not have to integrate two different text segments themselves, because the linguistic marking of this relation guides them. This demands less cognitive effort and leads to faster processing (Britton, Glynn, Meyer, & Penland, 1982; Haberlandt, 1982; Maury & Teisserenc, 2005; Sanders & Noordman, 2000). As we have seen in this chapter, in the absence of a coherence marker, the text segments can still be integrated by the reader. This takes more effort, but it is still possible.

Aside from an integration function, coherence markers are assumed to have an *inference* function, causing the reader to check the relation against their knowledge base. This process usually occurs towards the end of the sentence and elevates reading time (Noordman & Vonk, 1997, 1998). Again, the inference can also be made in absence of the coherence marker, but this is assumed to be less likely (see also Chapter 7). There are different types of inferences. Graesser, Singer and Trabasso (1994) listed 13 types of inferences. Examples of these inferences on the level of local coherence are *causal antecedent inferences* and *referential inferences*. At the level of global coherence, inferences such as *thematic inferences* (the main point or moral in the text) and *superordinate goal inferences* are made. Examples of inferences at a more pragmatic level are *author's intent* and *emotions*.

Evidence for both the integration and the inference effect of coherence marking was provided by Cozijn (2000). He concluded that the integration effect occurs early in the sentence and speeds up processing of the words immediately followed by the marker. The inference effect occurs late in

24 – CHAPTER 1

the sentence and slows down the processing of the final words. An example of both integration and inference effects is illustrated by examples 12a and b. Cozijn (2000) found a speeding up effect (integration) on segments 1 and 2, and a slowing down effect (inference) on segment 3. The inference effect was consequently confirmed with verification latencies. When participants were presented verification statements, containing the inference in question, they were faster in their true/false judgments after the explicit version than after the implicit version. This confirms that in the explicit version, readers make the inference on-line, whereas they make the inference off-line in the case of the implicit condition.

12a. / Hij ondervond / een flinke vertraging / omdat / [s1] er / [s2] een snelheidscontrole was / [s3] op de snelweg. / Hij....

12b. / Hij ondervond / een flinke vertraging. / [s1] Er / [s2] een snelheidscontrole was / [s3] op de snelweg. / Hij....

Other researchers have tried to replicate this pattern. However, Maury and Teisserenc (2005) found only evidence for an integration effect (speeding up the reading times) and no evidence for an inference effect (slowing down). This same pattern was found by Millis and Just (1994). They conclude that the slowing down at the end of the second clause is due to reactivation of the first clause, not to making inferences. They suggest that the presence of a marker causes *fewer* inferences to be made, because the text in itself is already explicit enough.

In this dissertation, both the integration and the inference functions of coherence marking are studied, as well as other possible functions that coherence marking might have (see Chapter 2). In short, there are indications for coherence marking to influence not only integration and inference, related to comprehension, but also people's beliefs and opinions. Therefore, these types of effects will also be studied in this dissertation.

1.4 Coherence marking in this dissertation

In this first chapter, I raised several important points concerning the notion of coherence marking and how I will be using the term *coherence marking* throughout this book. I will briefly repeat them here.

First, I take a broad perspective on coherence marking in the sense that I generalize over different types of coherence relations (1.1) and different types

of markers (1.2.1). On the other hand, I take a more limited perspective in the sense that I include only *linguistic* marking of coherence in my definition (1.2.2), which means that coherence marking should *only* mark the relation in question with an adequate marker without adding elaborative information.

Both effects of coherence marking on text processing (integration and inference effects) and effects on the mental representation are important in this dissertation, although the effects on the mental representation will receive some more consideration than processing effects. Processing theory and empirical data are used to gain more insight in effects on the mental representation. Effects of coherence marking can only be fully understood by combining off-line and on-line data. In other words, we will investigate what effect coherence markers have on the mental representation that a reader has constructed after reading a text, but also how this happened and what processes occurred.

At the start of this chapter, I distinguished three different ways of investigating effects that textual characteristics may have on the reader (analyzing the discourse, studying processing effects and studying the mental representation). This dissertation focuses on the mental representation and also includes discourse processing. Therefore, the work reported here is experimental in nature.

1.5 Main research question

After having defined coherence marking and discussed its role in text processing and representation, we can now formulate the main research question in this dissertation as follows:

How does linguistic coherence marking influence the reader's mental representation?

A first aspect in which the research in this dissertation differs from previous research is the definition of the mental representation. In many studies on text effects on the mental representation, this is defined as text comprehension. Also, mental representations are mostly investigated in the context of one single text type, often informative texts. This approach, as I will explain in more detail in the following chapter, is too narrow. We need to consider other aspects of the mental representation than comprehension alone. Chapter 2 will explain in detail what types of aspects I want to include in the definition of mental

representation and why. In short, I propose that mental representations that readers make include not only comprehension, but also persuasion and appraisal. If opinions and beliefs are used when readers construct a mental representation of the information in the text, it makes sense to expect the text to have a possible influence on these factors as well. This will subsequently lead to new concrete research questions.

A second new factor in this dissertation is the inclusion of an important reader characteristic: prior knowledge. When readers construct a mental representation of the information in a text, they have to integrate the textual information with their own prior knowledge. Chapter 2 will provide more information about this interaction. The research in this dissertation systematically includes prior knowledge as an experimental variable.

Third, the research in this dissertation investigates coherence marking and its effects in two different text genres. Do the effects of coherence marking differ if they are used in another text genre? Do coherence markers in an informative text play the same role as coherence markers in a persuasive text?

To close this first chapter, I want to stress the importance of coherence marking. One might wonder why coherence marking is of such particular interest to reading researchers. We have seen in this chapter how crucial coherence is for comprehension. Marking the coherence of a text is expected to make the understanding of the information in the text easier. By studying the effects that coherence marking has on readers, we develop knowledge on how to write effective texts. This is one of the major goals in research in the area of document design. But these insights are not only useful in document design. In an educational context, it may help to increase or decrease the text's coherence in order to help pupils learn new information. Another practical application of studies like the ones in this dissertation is marketing: marketing can benefit from knowledge about the persuasive effects of textual characteristics such as coherence marking in order to promote certain ideas in a more effective way. And finally, in the context of cognitive science, studying cognitive processes of reading and understanding provides insight in the way these processes result in the mental representation of the textual information. Since coherence plays a crucial role in discourse processing and discourse representation, the studies reported in this dissertation contribute to these essential issues in cognition and communication.

2 Effects of coherence marking and prior knowledge on the situation model representation: previous research

In chapter 1, I defined the notion of coherence marking and explained its importance for discourse. I also introduced the main research question: what are the effects of coherence marking on the mental representation?

In the current chapter, I argue that the situation model representation includes more than comprehension alone. I present an overview of previous studies concerning the effects of coherence marking. The research that I propose sets out to add to existing knowledge about coherence marking and its effects in three ways: (1) by including an important reader characteristic, prior knowledge, (2) by including more than one aspect of text representation and (3) by comparing effects of coherence marking in two different text genres. This leads to six concrete research questions.

In the previous chapter, I introduced the main research question:

How does linguistic coherence marking influence the reader's mental representation?

But what exactly is 'the mental representation'? Is there such a concept as *the* mental representation? What aspects are included in such a representation? In the current chapter, I argue that the current views on the cognitive representation of textual information are often too narrow. Mostly, when the mental representation is the focus of attention, text comprehension is the only aspect that is investigated. I will show why this view on representation needs to be enlarged. In short, I show that representations are formed by integration of textual information with prior knowledge, beliefs, experiences, feelings, and so on. This means that it is very plausible to expect all these aspects to be part of the mental representation: if they are used to construct the mental representation, they must be included in some form. I propose to include at least three related aspects in the mental representation: text appraisal (2.3), meta-cognition (2.4) and persuasion (2.5).

2.1 Text representation

Many studies of coherence marking focus on 'the mental representation' that readers construct of a text. This term is a simplification: *the* mental representation does not exist. In fact, psycholinguistic research has shown that readers construct multiple representations of a text (Kintsch, 1998; Singer, 1990). In the next paragraph, I will explain the levels of representation that are generally defined and which one is most appropriate when we focus on the effects of coherence marking.

2.1.1 Levels of mental representation

Generally, three levels of mental representation are distinguished in discourse studies. The most superficial level is called *surface code*. It concerns grammar and wording. This level represents the exact linguistic material in the sentence. The second level is referred to as the *text base* and contains the propositions from the sentences. The third form of representation is called the *situation model representation*. This is the deepest level of representation, where readers integrate the linguistic material from the text with their own background knowledge.

The following example illustrates the three levels of representation. Suppose a reader comes across the sentence *The prince holds the princess*¹. The surface code representation contains information about the grammar of this sentence, for instance *prince* and *princess* are nouns, *holds* is a verb, *the prince* the NP, *holds the princess* the VP, etc (see A in the Figure 2.1 below). The text base representation contains representations of the concepts of a prince and a princess and the act of holding (see B in Figure 2.1). Finally, the situation model representation can differ considerably from one person to another, from one moment to another, from one context to another.

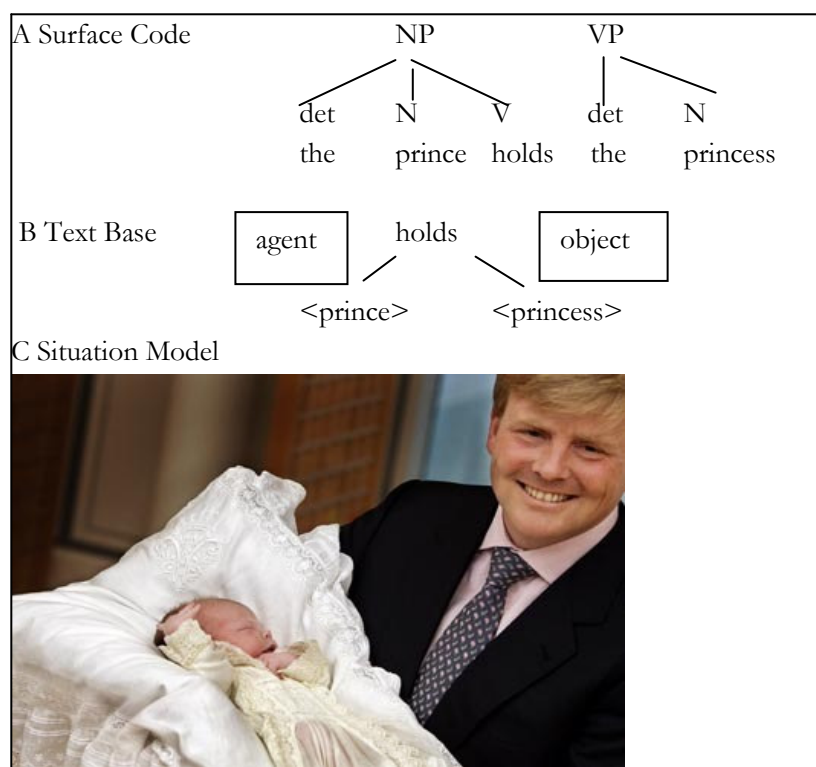


Figure 2.1: *Three levels of representation for 'The prince holds the princess'.*

Depending on the context, a reader could, for instance, form this image of the Dutch Crown Prince Willem Alexander, holding one of his daughters, Princess Alexia. This example shows how flexible situation model representations are;

¹ Example taken from paper presented at the Fifteenth Annual Meeting of the Society for Text and Discourse (Kamalski, van den Bergh, Lentz, & Sanders, 2005).

for each reader, the mental image that they form of this sentence can differ. Situation model representations can be influenced by the text, by the situation described, and by the prior knowledge a reader has at his disposition. They are therefore very rich and elaborate representations. Kintsch (1998:198) stresses the flexibility in situation model representations by stating that ‘it should be clear from everything that has been said so far that there is no single type of situation model and not a single process for the construction of such models’.

Several theoretical definitions of situation model representations are used, in all of which the text is linked to the reader. Zwaan (1999b:15) defines situation model representations as ‘mental representations of the state of affairs described in a text rather than the text itself’. Kintsch (1998:103) describes situation models by linking comprehension specifically to the reader’s prior knowledge: ‘various sources of knowledge must be used in construction of situation models: knowledge about the language, about the world in general, and about the specific communicative situation’. Graesser, Millis and Zwaan (1997) refer to situation models as the micro world of a text. Such a micro world is constructed inferentially by integrating the information from a text with the appropriate background world knowledge. What all these definitions have in common is that situation model representations exceed the text itself, and include extra-textual information.

Empirical evidence for the existence of situation model representations was offered by Schmalhofer and Glavanov (1986) and Fletcher and Chrysler (1990) in a *sentence recognition paradigm*. First, participants read a text and then a target sentence was presented to them. These target sentences possibly differed from the original sentence that the participants had read in the text on several levels of representation. The target sentence could be an original (same surface code as the text), a paraphrase (same text base as the text, but different surface code), a changed-meaning (same situation model, but different text base), or a changed-model (the situation model had been changed). The following example is taken from Fletcher and Chrysler (1990). It is important to keep in mind that in the short text passage they used, the vase was the most expensive item, the necklace somewhat less, and the carpet and the painting were least expensive (vase > necklace > carpet and painting). The participants had read this information before they read the target sentences. The target sentences in the experiment were the following:

EFFECTS OF COHERENCE MARKING: PREVIOUS RESEARCH - 31

1. original	George says that his wife was angry when she found out that the necklace cost more than the carpet.
2. surface test	George says that his wife was angry when she found out that the necklace cost more than the rug.
3. text base test	George says that his wife was angry when she found out that the necklace cost more than the painting.
4. situation model test	George says that his wife was angry when she found out that the necklace cost more than the vase.

Then, participants were asked to judge whether or not the target appeared in the passage they just read. The results indicate that participants can discriminate between sentences they have actually seen (1) and meaning preserving alternatives (2). This provides evidence for the existence of the surface code representation: rug and carpet are synonyms, so both statements are true in the context of the story, but the participants had only read the word carpet before. They correctly recognized that the sentences were not identical. Participants can even discriminate better between originals (1) and new meaning alternatives (3). This is evidence for a text base representation: although the meaning of the sentence has been changed, it is still consistent with the original situation. The distinction between originals and changed meaning sentences (4) was the easiest distinction to make for the participants. The least mistakes were made when the target sentence differed on the situation model level from the original text. Replacing rug with vase violates the original situation. The fact that participants' performance on this test exceeded their performance on the text base test is a clear indication for the situation model representation and its importance. These results from Fletcher and Chrysler (1990) show why the situation model representation is such an important concept in text comprehension; readers clearly recognize situation model representations that are inconsistent with the information that they just read².

² However, there are many studies that report how readers fail to notice inaccuracies, such as the Moses-illusion (Erickson & Mattson, 1981; Sanford & Garrod, 1994; van Oostendorp & De Mul, 1990). When asked 'How many animals of each kind did Moses take on the Ark,' most people respond 'two,' even though they know that it was Noah who took the animals on the Ark. When a term is replaced with a semantically similar but incorrect term, people have difficulty in detecting the distortion. The surface code is not correctly represented, caused by the readers' top-down processing and the place of Moses in the sentence that causes readers to take this information as given. Therefore, readers do not check the term Moses and assume that Noah is correct.

These studies provide evidence for the existence of situation model representations and their importance, but they do not give information on the construction process of these representations. When processing a text, how do readers construct their situation model representation? Or, what particular process takes place that result in this deep level of comprehension, where the representation of the information in the text is integrated with previous knowledge?

Kintsch (1988, 1998) described this process in his Construction-Integration Model. Constructing a mental representation occurs in a cyclical process focusing on the current sentence and how it relates to the preceding discourse. In general, this yields a mental representation that is not completely coherent. It needs to be connected to the reader's personal knowledge in order to be coherent. Various sources of knowledge are used in this process: knowledge about the text topic, about the world in general, about this specific situation, and so on. Also, readers use personal experiences in the construction of the mental representation. All these factors combined (text, knowledge, experiences, etc) interact when a mental representation is constructed. There are no general rules for this process: every reader has other experiences and different background knowledge to use for situation model representation construction. The situation model representation also depends on the text itself and on reader characteristics, such as the reader's goals, motivation, and resources.

2.1.2 Coherence marking at the situation model level

In the previous sections, I introduced the three levels of representation that are commonly assumed to exist: the surface code, the text base and the situation model representation. I also explained how situation model representations are constructed. An important question is: At which level do we expect the textual characteristic coherence marking to play a role?

Of course, the surface code changes when a coherence marker is used. But the representation levels that are most interesting are the text base and situation model representation. The surface code representation is short-lived, whereas the situation model representation has a longer lasting effect on the reader. Do coherence markers influence surface code, text base or situation model representations?

EFFECTS OF COHERENCE MARKING: PREVIOUS RESEARCH - 33

Sanders and Noordman (2000) provide evidence for the position that coherence marking influences the surface code representation. They tested the influence of type of *coherence relation* on comprehension and processing as well as the influence of *marking* on comprehension and processing. Their results show that a sentence in a problem-solution relation was read faster, caused better and faster results on verification statements and better recall than the same sentence in a list relation. These are indications for a text base or situation model influence of the type of relation. When they focused on coherence marking, Sanders and Noordman found that by adding a lexical signaling sentence, explicit sentences were read faster than implicit sentences, where the coherence relation was left implicit. However, there was no effect on recall. They concluded that the coherence *relations* influence text base and situation model representations, but the *marking* of these coherence relations seems just to have a surface code effect.

However, many studies hypothesized a more profound effect of coherence marking at the level of the situation model representation. Some of these studies have shown that the effect of coherence marking interacts with prior knowledge. Prior knowledge only plays a role at the level of the situation model representation. Therefore, the fact that coherence marking has been reported to interact with prior knowledge clearly provides evidence for the position that coherence marking acts on the situation model level. This is the level at which we can expect coherence marking to play its most important role. The fact that Sanders and Noordman found no evidence for this in their recall task is probably related to methodological issues. Free recall and other experimental methods that are used to measure situation model representations do not always seem precise enough to measure an effect of coherence marking. This is clearly a matter of concern, which will be addressed separately in section 2.1.4 and in Chapter 4 of this dissertation.

2.1.3 Effects of coherence marking on the mental representation

In general, research on the effect of coherence marking on the mental representation after reading a text does not provide clear-cut answers to the question of whether or not markers are beneficial for the quality of the mental representation that readers construct from a text (see for recent overviews: Ben-Anath, 2005; Sanders & Spooren, 2006). On the one hand, some studies show that markers of coherence improve the mental representation that readers

build from a text. Coherence marking has been shown to cause better answers on text comprehension questions (Boscolo & Mason, 2003; Degand, Lefevre, & Bestgen, 1999; Degand & Sanders, 2002) and better recall performance³ (Gilabert, Martínez, & Vidal-Abarca, 2005; Lorch & Lorch, 1986; Lorch, Lorch, Ritchley, McGovern, & Coleman, 2001; Meyer, Brandt, & Bluth, 1980; Sanchez, Lorch, & Lorch, 2001). On the other hand, some researchers did not find any effects of coherence marking on text comprehension questions (Spyridakis & Standal, 1987) or on information recalled (Britton, Glynn, Meyer, & Penland, 1982; Sanders & Noordman, 2000).

Some researchers have included the reader characteristic prior knowledge in their study of coherence marking, as Roller (1990) suggested. Prior knowledge, also referred to as background knowledge, topic knowledge, content knowledge, domain knowledge or world knowledge, is defined by Anderson (1984: 243) as ‘organized knowledge of the world that provides much of the basis for comprehending, learning and remembering ideas from stories and texts’. Some studies reported an interaction between the linguistic marking of text structure and prior knowledge (McNamara, Kintsch, Songer, & Kintsch, 1996; McNamara & Kintsch, 1996): readers with prior knowledge benefit optimally from a non-marked or implicit text, readers who lack this knowledge perform better on text comprehension questions after a marked or explicit text. The same interaction effect has been reported by McNamara (2001), Voss and Ney Silfies (1996) and Long, Wilson, Hurley and Prat (2006). Others found a main effect of prior knowledge and a main effect of the marking of text structure, but no interaction (Birkmire, 1985; Gilabert, Martínez, & Vidal-Abarca, 2005; McKeown, Beck, Sinatra, & Loxterman, 1992). A related study is that by Linderholm, Everson, van den Broek, Mischinski, Crittenden and Samuels (2000), who have investigated differences between difficult and easy texts instead of high knowledge and low-knowledge readers. The difficult texts were harder to understand because readers did not have the necessary prior

³ This finding seems to be in conflict with my earlier remarks that the recall tasks from Sanders & Noordman (2000) were not sensitive enough to measure an effect of coherence marking at the situation model representation level. There is no clear explanation for their absence of effect and the fact that other researchers did find such an effect with a recall task: it could be due to the type of text that was used (short informative texts in Sanders and Noordman, or longer schoolbooktexts such as the ones in Gilabert et al or Sanchez et al). In Chapter 4, I will look for a valid method to assess situation model representations in order to solve these contradictions.

knowledge, whereas they *did* have the necessary knowledge to interpret the easy texts. Linderholm et al. found that for difficult texts (which we could compare to the 'low-knowledge' condition) readers benefit from explicit coherence.

Not all of these studies on effects of coherence marking on the mental representation are equally reliable (see Degand & Sanders, 2002). The studies vary among other things in 1) quality of the manipulations, including 2) the variable prior knowledge, and 3) methodology used to assess text comprehension. To clarify the sometimes contradictory results, these three factors are systematically investigated in the first set of experiments in this dissertation.

Another clarifying factor concerns the exact definition of coherence marking. Often, when coherence marking is being investigated, many more different textual aspects are varied at the same time. In Chapter 1, we have seen how some studies manipulated coherence in a way that not only increased text coherence, but also added extra information to the text content. It is of great importance to investigate the interaction between coherence marking and prior knowledge, found by McNamara and Kintsch (1996), McNamara, Kintsch, Songer and Kintsch (1996) or more recently by Long et al. (2006), but to do so with experimental texts that only differ in linguistic markers of coherence. Hence, Research Question 1 can be formulated as:

RQ 1a: In what way does linguistic coherence marking influence the situation model representation of a text?
--

The exact nature of the influence of prior knowledge on the situation model representation effect of coherence marking is not clear yet. It becomes evident from the overview above that some studies included this reader variable and other studies did not. But even when this variable *was* taken into account, this did not lead to clear-cut results: in some cases an interaction was observed and in other cases it was not. Therefore, in the experiments in this dissertation, the reader characteristic prior knowledge is systematically included. I hypothesize that there is indeed an interaction between prior knowledge and coherence marking, in the sense that coherence marking is beneficial for comprehension of low-knowledge readers, but not for comprehension of high knowledge readers. Hence, we can formulate Research Question 1b as follows:

RQ 1b: Does the reader characteristic prior knowledge influence the effect of linguistic coherence marking on the situation model representation?

In the research question formulated so far, it was assumed that the situation model representation can be assessed in a reliable and valid way. However, showing the hypothesized interaction effect of coherence marking and prior knowledge seems to depend on what methodology is used to assess the crucial level of text representation: the situation model representation. A logical second question is therefore:

RQ 2: What methodology is most suitable when assessing situation model representations?

Almost all of the studies above have used different methodologies to determine the effect of coherence marking on text representation. In the following paragraph, I will discuss some of these different methodologies, discuss several examples and conclude what adaptations could be made in order to make the method a valid situation model representation task.

2.1.4 Methodological options to measure the situation model representation

Researchers can choose from many options when measuring situation model representations. This can be problematic when we want to compare experimental studies and their outcome: which study leads to reliable measurements and conclusions and which does not? In this section, some of the methodologies and recent examples of practical empirical applications⁴ are discussed, as well as advantages and inconveniences of each method. I do not intend this presentation of methodologies to be exhaustive; however, the most often used methods in coherence marking research are included.

First, we will look at the traditional method of question asking, followed by scaling or rating methods, recognition methods, mental model tasks and cloze tasks. The following short text passage (7) on the Parisian Eiffel

⁴ The methodologies were selected from overview papers such as Royer, Cisero & Carlo (1993), Hoffman, Shadbolt, Burton & Klein (1995), Ericsson & Smith (1991) and Olsen & Biolsi (1991). Also, Kintsch (1998) provides a useful list of methods and strong criticisms.

Tower, identical to Chapter 1, will serve as the basis for the example questions for all these methods.

7. The Eiffel Tower was erected for the Paris Exposition of 1889. Of the 700 proposals submitted in a design competition, one was unanimously chosen, a radical creation from the French structural engineer Alexandre Gustave Eiffel. However, the controversial tower elicited some strong reactions, for different reasons. The first one was that some people— including Maupassant and Zola— found the Eiffel Tower useless and monstrous. Second, nature lovers thought that it would interfere with the flight of birds over Paris. It was almost torn down in 1909, but was saved because of its antenna - used for telegraphy at that time. Still, the Tour Eiffel is now completely accepted by French citizens, and is internationally recognized as one of the symbols of Paris itself.

Question asking: problem solving and bridging inference

The traditional method to measure text comprehension is *question asking* (Kintsch, 1998). The questions can have multiple-choice answers, or they can be open-ended. Questions measure the situation model representation when the information necessary to answer is not explicitly present in the text, e.g. problem solving questions -'linking information from separate sentences within the text and applying this information to a novel situation'- or bridging inference questions -'the information occurs in the text but requires linking two or more sentences to answer the question' (McNamara, Kintsch, Songer, & Kintsch, 1996). An example for the Eiffel Tower text would be:

8. Name two reasons why the planned Eiffel Tower elicited negative reactions.

Here, the reader has to link the two reactions mentioned in the text and see that they have something in common, namely the fact that they are negative reactions to the planned Eiffel Tower. Asking such inference questions is a simple, often used methodology to assess situation model representations.

However, problems often occur when this method is used. McNamara et al. (1996:14) discuss the use of specific types of inference questions, namely multiple choice, true-false and short answer ones, as follows: 'The attempt to measure the deeper understanding of the problem domain that these texts communicated to the readers by means of inference questions was unsuccessful, however, perhaps because there were too few of them (only four

or five questions in each set) or perhaps we simply asked the wrong questions'. The main problem in the McNamara et al. study is low internal reliability, indicating that the separate questions may not measure the same concept. This could be explained by the fact that comprehension is not considered to be a unitary construct, but a complicated mix of text and reader characteristics such as prior knowledge and purpose (Duke, 2005). This problem with the reliability of question asking raises serious concerns about its validity: do these questions actually measure what we intend to measure, namely the situation model representation? However, bridging inference questions and problem solving questions (McNamara et al., Experiment 2) seem appropriate questions to assess situation model representations.

We can conclude that, despite the fact that question asking is the traditional method for measuring comprehension, it may be a difficult method in terms of internal reliability. The separate questions often do not seem to measure the same construct. It seems difficult to ask exactly the relevant questions to measure the situation model representation that the reader has constructed in a reliable way. Question asking will be included in our experiment on the validity of situation model tasks in Chapter 4.

Scaling or rating methods

A second important technique is *scaling*: a set of key words in a domain is organized by experts, and the degree to which the organization of the participant resembles the organization of the expert, indicates the participant's level of understanding (Kintsch, 1998). An example for the Eiffel Tower text would be:

9. How strongly are the following persons related?

Gustave Eiffel – Emile Zola : very related 1 2 3 4 5 6 7 not related at all

Emile Zola – Guy de Maupassant: very related 1 2 3 4 5 6 7 not related at all

The correct answer would be that Zola and de Maupassant are more related in the text (because they share an opinion) than Eiffel and Zola (who have opposite opinions). The most basic way of obtaining an organization of these concepts is presenting the participants with pairs of keywords and asking them how related these concepts are. This method is considered to be very time-consuming for both the participant and the researcher (Kintsch, 1998; Royer,

Cisero, & Carlo, 1993). A recent example of an experiment in which this method was used, is Roebben (2004). He assessed comprehension of texts about financial products by using a scaling task (*jugements de proximité de mots-clés*). Participants had to make proximity judgments before and after reading the text, in order to measure a change in the situation model representation (Britton & Gülgöz, 1991). Finally, Roebben compared the results on the scaling task to the results that he obtained with bridging inference questions. Whereas the bridging inference questions were sensitive enough to show significant differences between novices and experts, the scaling task failed miserably: no effects were observed. Roebben concludes that the scaling method is not able to capture the complexity of situation model representations.

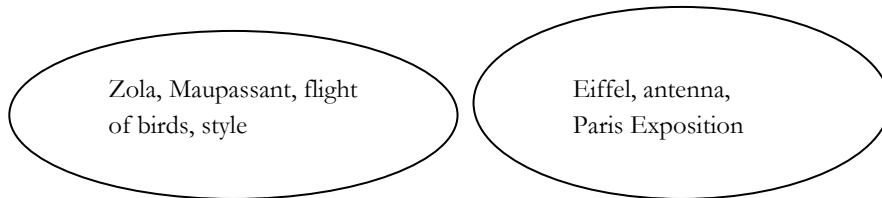
On the basis of these results, it can be concluded that the scaling method seems extremely time-consuming and does not seem to be sensitive enough to be a useful operationalization of the situation model representation. It will not be included in our investigation of the validity of situation model representations in Chapter 4.

Categorization tasks

Categorization methods, such as *card sorting*, have been used for a long time in psychology (Miller, 1969) and psycholinguistics. Only recently have they been applied to text comprehension as *sorting tasks*. When performing a sorting task, participants have to organize concepts into groups following the idea that similar concepts are placed on the same pile. In the sorting tasks, participants organize a set of key words from a text into groups. An example for the Eiffel Tower text would be the following task.

10. Make groups on the basis of the text. If you feel that two words belong together, put them in the same group. Draw a circle around each group.
Emile Zola, Gustave Eiffel, Guy de Maupassant, flight of birds, style, Paris Exposition, antenna.

The correct answer would be that Zola and de Maupassant could be placed together in a group (same opinions), and flight of birds and style could go together (arguments against Eiffel Tower). One could even place all these terms together: they are all related to a position against the Eiffel Tower. In the other group, people and arguments in favor of the Eiffel Tower could be placed.



Note that several categorizations are possible on the basis of common sense: one could make a case for putting all *writers* in one group, or all *people* in one group. Although this categorization makes sense, in the particular context of this text it is clear that the reader failed to recognize the opposition in the text if such a categorization is made. However, an important criterion for sorting tasks is that without reading the text, more than one categorization should be possible (Kintsch, 1998).

Sorting tasks were first applied to text comprehension in McNamara et al. (1996) and McNamara and Kintsch (1996). In both studies, expository texts from schoolbooks were used and participants had to organize cards into categories, before and after reading the text. This enabled the researchers to conclude what the effect of the text was on the organization of knowledge and on the situation model representation. They constructed an ideal sorting matrix, which was compared to the participant's matrix before and after reading the text. The researchers then performed a vector analysis to check whether the sorting after reading the text was closer to the ideal model than sorting before reading the text.

Sorting tasks seem to be a very appealing method to assess situation model representations, because they seem to be a relatively direct measure of whether or not a participant has linked concepts in his or her situation model representation. However, there are some points of concern. Kintsch's (1998:299) criticism on sorting tasks is that they produce 'results that are a bit too orderly and logical and do not necessarily reflect the memory structures that are operative in memory retrieval'. Still, other methodologies such as question asking or scaling also simplify the situation model representation considerably. The sorting task does not seem to simplify *more* than other methods.

In my view, a real problem with McNamara et al.'s sorting task is the comparison between pre-test and post-test. In the pre-test, participants were instructed to categorize these items how they thought the concepts should go

together. After categorizing, they labeled each pile they made. In post-tests, the researcher asked the participant to categorize according to the text. This means that participants were performing a different task (categorizing according to the text) compared to the one they executed before reading the text (categorizing according to their prior knowledge).

In conclusion, I consider the sorting task a very promising method to assess situation model representations, because it allows the researcher to measure many more links or relations in one sorting task than with several separate questions. Furthermore, the method focuses strongly on the relations between concepts in the representation rather than on the way they are literally expressed in the text, and participants can express their ideas on the conceptual representation without even formulating them. The act of sorting is sufficient. This is an advantage, because the influence of language proficiency is smaller than when participants have to formulate answers themselves. Sorting tasks focus on the representation of conceptual relations between crucial concepts from the text. This seems to be the closest we can get to situation model representations. However, I propose to use the sorting tasks in a post-text only design. As a consequence, the scoring technique changes, because there are no pre- and post-measurements to compare. The matrix multiplications and vector scores are no longer necessary. Although this means that the method of analysis will definitely be less sophisticated, there is an upside to this. The sorting task method now becomes easier to use for educational purposes, which is one of the most important applications of situation model representation tasks. In summary, the sorting task plays a prominent role in our search for the most valid assessment of situation model representations in Chapter 4.

Recognition methods

A fourth method to assess situation model representations are recognition methods (see the earlier section 2.1.1, Fletcher and Chrysler, 1990) In recognition methods, target words or sentences are presented to participants, who are then asked to judge whether or not the target appeared in the passage they just read. Target sentences can differ from the original sentence on several levels of representation: they can be originals (same surface code as the text), paraphrases (same text base as the text, but different surface code) changed-meaning (same situation model, but different text base), or changed-model (even the situation model has been changed). If more levels differ from the

original (e.g. surface, text base and situation model), people are expected to respond less often ‘yes, I’ve seen this sentence before’ as in a situation where only one level differs from the original (e.g. surface code). This methodology (see also paragraph 2.1.1) was first used by Schmalhofer and Glavanov (1986), and was followed by many other researchers, such as Fletcher and Chrysler (1990). An example for the Eiffel Tower text would be question 11. The answer to this question would have to be ‘no’, because the tower elicited strong negative reactions.

11. Did you read the following sentence?

However, the controversial tower elicited some positive reactions.

Yes, I’ve seen this sentence before - No, I have not seen this sentence before

This paradigm was recently used in an experimental study by Long, Wilson, Hurley and Prat (2006). They studied the effect of coherence marking on situation model representations with a recognition task. Participants read a text about Star Trek, and were then presented a sentence from the text. Their recognition task was to answer the following question: Is this a sentence from the text? If their judgement was that it was in fact a sentence that came from the text, they were then asked whether they had a ‘vivid, conscious, aware’ representation of that sentence or that they simply knew they had read it (in terms of Long et al., *remembering* versus *knowing*). The *remember* judgments are situation model measurements, according to the researchers, because they involve conceptual processing. Text and prior knowledge have to be integrated in order for a representation to be vivid and conscious. The *knowing* judgments are considered to be text base measurements⁵.

Although recognition methods have a clear principal relevance, such as in the studies by Chrysler and Fletcher (1990) and Schmalhofer and Glavanov (1986), they are less relevant for practical assessment of text comprehension. Recognition methods can show *that* a situation model representation has been constructed, but are less useful to assess what relations and concepts are present in the situation model representation. This method will therefore not

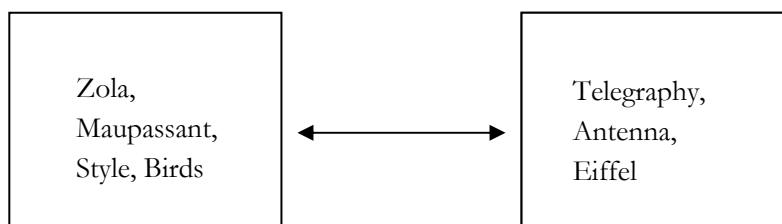
⁵ I am not convinced by the authors’ claim that the *remember* versus *knowing* distinction is apt for distinguishing between text base and situation model representations. Whether people claim to have a vivid representation of the sentence or not, does not seem to be an adequate measure of the situation model representation.

be taken into account in our search for the most valid method to assess situation model representations in Chapter 4.

Mental model tasks

Another type of situation model questions are mental model tasks. They have only recently been used in text comprehension. Mental model tasks are typically schematic pictures or diagrams. Key words from the text have to be placed in the right box or place in the schematic representation (Royer, Cisero, & Carlo, 1993). An example for the Eiffel Tower text would be:

12. Put the following words in the correct box: Style, Telegraphy, Antenna, Birds, Zola, Eiffel, Maupassant.



According to the text, a correct solution would be to put arguments and people in favor in one box, and arguments and people against in the other. McNamara (2001) applied this type of questions to test text comprehension, in combination with bridging inference questions. She suggests that bridging inference questions and mental model tasks measure the same construct, which could indicate that mental model questions are a suitable task for assessing situation models.

However, Boscolo and Mason (2003) used a comparable mental model task to measure the influence of coherence marking on situation models and conclude that these diagrams do not measure the same construct as bridging inference questions do. A possible explanation for the differences in results between these two methods is the fact that verbalization is needed in order to answer the inference questions, which is not the case in diagrams. This would be a reason to prefer mental model tasks over bridging inference questions.

In conclusion, mental model tasks seem to be a promising method to assess situation models, because surface code or text base representation cannot be sufficient for the reader to answer these questions. However, this task has only been applied to text comprehension fairly recently and so far, its use seems

44 – CHAPTER 2

to lead to different conclusions. It seems that this method is worth considering as an option to measure situation model representations, and that is why its validity is tested in Chapter 4.

Cloze Tests

Ever since Taylor (1953), cloze tests are known as a valid measure of readability and later also as a reliable measure of reading comprehension (Levenston, Nir, & Blum-Kulka, 1984). Cloze tasks are constructed by randomly omitting every *n*th word from a text. The reader has to fill in the blanks. An example for the Eiffel Tower text is:

13. The Eiffel Tower was erected for the _____ Exposition of 1889. Of the 700 proposals _____ in a design competition, one was unanimously _____, a radical creation from the French structural _____ Alexandre Gustave Eiffel. However, the controversial tower _____ some strong reactions, for different reasons. The _____ one was that some people— including Maupassant _____ Zola— found the Eiffel Tower useless and _____. (etc.)

Cloze tests are a traditional and accepted method to assess text comprehension. However, the cloze test has been criticized over the years, mostly because of the fact that it is not always clear what is being measured. For instance, Alderson (1983) pointed out that cohesion and coherence are not measured while these are important factors in text comprehension. Leventson, Nir and Blum-Kulka (1984) refined this position by stating that because words are omitted randomly in a cloze task, restoration of the text might involve text base representations (roughly equivalent to their term ‘micro-processing’) or situation model representations (‘macro-processing’), or even both.

Cloze tasks are very easy to use: their construction is relatively easy, there is no difficult scoring technique and the instruction for the participants is simple. In addition, cloze tasks are known to be internally very reliable: all items are assumed to measure the same construct. Therefore, cloze tasks could in principle form a very useful situation model task. However, I am not sure which level of representation is addressed; it might even differ between cloze items. Depending on the gap, it might concern a missing word in a fixed expression, which implies that a surface code representation is measured. Other gaps might concern information in the same sentence, without any integration of information between sentences or integration with extra-textual information

such as prior knowledge. Therefore, I would expect the cloze task to measure mostly surface code and text base representations.

Conclusion

From this overview of methods to assess situation model representations, it becomes clear that there is a large variety of methods to assess this crucial level of text representation. Our aim is to investigate the influence of coherence marking on the situation model representation. Therefore, the first step needs to be to find a reliable method to test this influence. This leads to the following Research Question:

RQ 2: What methodology is most suitable when assessing situation model representations?

I test the validity of the methods that have been selected here as promising situation model representation tasks in Chapter 4. So far, we have defined the situation model representation as ‘comprehension on the deepest level’. We have seen how coherence marking and prior knowledge can influence text comprehension at the level of the situation model representation.

If comprehension is assessed, does that provide a complete picture of the situation model representation? Or could there be other factors that are also part of the situation model representation? I started this chapter by saying that more aspects are included in this representation than comprehension alone. The next sections propose and explain these other aspects of the situation model representation.

2.2 Extending the view on mental representation

The main question of this thesis is: How does coherence marking influence what the mental representation that readers construct of a text? The previous section explained what the current view on ‘mental representations’ is: a strong focus on what readers understand and remember after reading a text. This is a narrow view of text representation. We have seen in the previous section how situation model representations are generally defined: they are extremely flexible and dynamic representations, constructed by the reader as interplay between the text itself and prior knowledge, previous experiences, opinions, feelings, reactions, and so forth. It is therefore more than logical to measure

more than just text comprehension when we investigate effects of text characteristics on the mental representations. If these text characteristics act on the situation model level, they have the potential of influencing people's feelings, opinions, and knowledge. All these aspects need to be included in our view of the mental representation that readers construct when reading a text.

If we take the Eiffel Tower Text in this Chapter (see Example 7) as an example, it is easy to see how such an informative text has the capacity to do much more than just convey information about the history of the Eiffel Tower. Even people who have never seen the Eiffel Tower in their life will have formed an opinion about the Tower after having read this text. Not only does the text influence our opinion about the Eiffel Tower, we also form opinions about the text and its writer. If we take other text types than informative texts, such as persuasive texts, these other aspects of the mental representation become even clearer. If a text aims at changing people's opinions, that text is even more likely to influence those opinions than in the case of the Eiffel Tower text.

In this dissertation, I include several other aspects aside from text comprehension in my view on the mental representation. I start by proposing that the reader's opinions about the text and the text's point of view should also be considered part of the representation that readers form of a text.

2.2.1 *Appraisal*

When we want to investigate the mental representation, a relevant concept is *appraisal*, a subjective judgment of situations or events, related to the needs and to the goal of the judge (Ortony, Clore, & Collins, 2001; Roseman, Antoniou, & Jose, 1996; Scherer, 1999). Appraisal can be applied to different things: events (*the party last Friday was great*), people (*the fifth grade math teacher was awful*), situations (*forgetting my wallet when I go to the supermarket makes me feel ashamed*) and objects (*I love my grandfather's car*).

A specific object of appraisal is text. *Text appraisal* is a subjective judgment of a text, related to the goal of the reader. In other, more text oriented studies (de Jong & Schellens, 1994; Lentz & de Jong, 2003; Oversteegen, van Keulen, & van Wijk, 2002), appraisal is also referred to as *text evaluation*. In both cases, the same subjective judgment of the text is the focus of attention.

Appraisal is considered to be a multidimensional concept. In other words, there are several components that together form the concept of appraisal. For instance, the appraising questions relate to issues such as ‘Does this text help me attain some goal?’ or ‘Is it pleasant to read?’ or ‘Did other people like this book?’. Appraisal theory distinguishes several so-called *Stimulus Evaluation Checks* (SECs), such as *appealingness* and *praiseworthiness*. Text evaluation theory distinguishes so-called *dimensions*, such as *relevance* and *clarity*. In Table 2.1, only the relevant dimensions of appraisal to the current context (that of text appraisal) are represented (Ortony, Clore, & Collins, 2001 based on p. 69).

Dimension	Appraisal
Appealingness	liking/ disliking
Desirability	relief/ disappointment/ satisfaction
Praiseworthiness	approving/ disapproving

Table 2.1 *Dimensions of appraisal, taken from Ortony, Clore and Collins, 2001*

The first type of appraisal is that of *appealingness*. Suppose that I come across a brochure about a short weekend in Paris. How would I form an opinion about the brochure in question? Relevant issues might be: Do I like this brochure? How does it look? Is it attractive? In more text oriented approaches the same term *appealingness* is used.

A second relevant type of appraisal is that of *desirability*. Desirability relates to how a text can help us realize our goals. We read texts to fulfill our need for information, or for entertainment. A question that relates specifically to the Paris brochure could be: Is it easy to find the right information? Am I able to make sense of it? More text oriented approaches refer to this type of appraisal as *accessibility*, a term that makes more sense in the context of text appraisal than desirability.

A third relevant type of appraisal is that of *praiseworthiness*. This type of appraisal concerns our beliefs and standards. For example, we believe that we should respect our parents, or that we should eat two pieces of fruit every day. When we form an opinion of other people, events, or in this case texts, we consider whether or not it is in accordance with our beliefs and standards. In the case of a brochure on Paris, it could state that Paris is the most romantic city in Europe. I would personally agree, but other people might think that that

is not true, because they think Venice is the most romantic city in Europe. In terms of Oversteegen et al., these questions concern *persuasion* and are considered separately from appraisal. This aspect could very well be considered an aspect of appraisal, but in this dissertation, I have chosen to address this persuasive aspect separately because of its complexity (see 2.3).

I propose, in the context of text appraisal, to use the following terms for the two relevant dimensions: appealingness and accessibility. Persuasion, the third relevant dimension of appraisal, is viewed as a separate factor. In the next section, I discuss the possible effects that coherence marking might have on the two appraisal types: *appealingness* and *accessibility*.

2.2.2 *Effects of coherence marking on appraisal*

In this dissertation, I am specifically interested in the effects of coherence marking on text appraisal. I expect coherence marking to influence judgments on text style, text point of view and text structure. This means that *appealingness* (style), *persuasion* (standpoint) and *accessibility* (structure) are the relevant dimensions of appraisal in the context of coherence marking. Persuasion is discussed separately in 2.4. For now, I focus on appealingness and accessibility.

How does coherence marking influence people's text appraisal? Are there specific hypotheses to test? Although not many, there are some studies that have included this subjective judgment variable in their study. Land, Sanders, Lentz and van den Bergh (2002) tested the effect of coherence marking on text appraisal for pupils in high school. These pupils judged the text without coherence markers to be easier to understand, but the version with markers was found to be clearer. There seems to be some inconsistency in these judgments, making it difficult to formulate a concrete hypothesis on the basis of this study. This might be due to the fact that the pupils involved in this study seemed to have trouble answering these appraisal questions.

Some studies have investigated the possible interaction of coherence marking and text interestingness, one of the aspects of appraisal. Here, appraisal is not a dependent variable, but an intervening one. The central question in these studies was: Is coherence marking beneficial to comprehension when people find a text to be boring? Or, on the contrary, is coherence marking beneficial to comprehension when a text is interesting? Hidi and Baird (1986) state that marking of coherence only influences text comprehension if the used texts are considered to be boring. However,

EFFECTS OF COHERENCE MARKING: PREVIOUS RESEARCH - 49

experimental work by Spooren, Mulder and Hoeken (1998) shows that the structure of a text *does* influence comprehension, regardless of its interestingness. Texts with a problem-solution structure were read faster than texts with a list structure. The effect of structure was evident, regardless of differences in interestingness. Although these studies concern the nature of coherence relations (Sanders & Noordman, 2000), they do not aim to answer the question how coherence *marking* influences appraisal, so again, it is difficult to formulate concrete hypotheses on the basis of these studies.

A particularly interesting study, that can help formulate a concrete hypothesis to be tested, was done by Garbarino and Edell (1997). They found that the more cognitive effort was needed in order to process a document (low accessibility), the more negatively it was evaluated (low appealingness). Since coherence marking is expected to make processing of a text easier, this consequently could have a positive effect on text appraisal, compared to implicit texts.

In sum, the effects of coherence marking on appraisal are rarely investigated and if it is investigated, results are not clear. However, Garbarino and Edell provide reasons to expect coherence marking to have an effect on appraisal. This aspect of the situation model representation deserves more attention in the empirical work on coherence marking. Hence, we can formulate Research Question 3 as follows:

RQ 3: What are the effects of coherence marking on text appraisal? More specifically, what are the effects of coherence marking on appraisal dimensions appealingness and accessibility?

The expectation is that the presence of coherence marking has a positive effect on both appraisal types. Coherence marking makes processing easier (more accessible) and the text more attractive (more appealing). Judgments on accessibility and appealingness are therefore expected to be more positive.

2.2.3 *Meta-cognition*

Aside from affective factors that should be included when studying the mental representation, there are more aspects to consider. An important aspect of text representation that should be included in research on text characteristics in general is meta-cognition.

Meta-cognition can be defined as ‘knowing about knowing’ and includes knowledge and awareness about the cognitive processes (Gordon & Braun, 1985), such as *I don’t understand this, I will need more time to solve this problem, or I can’t study with the TV on*. Evaluating the current state of one’s ongoing comprehension and potentially regulating reading processes are very important meta-cognitive skills in reading comprehension (Baker, 1985). Meta-cognition is strongly linked to appraisal, discussed in the previous paragraph. When readers formulate judgments about text accessibility (this text is difficult), they are probably basing this judgment on their meta-cognitive evaluation of their own reading process. If this process was problematic to them, they are more likely to formulate negative judgments on appraisal. This means that these concepts of appraisal and meta-cognition are related, but not identical.

2.2.4 *Effects of coherence marking on meta-cognition*

One meta-cognitive effect in particular is often expected to be influenced by coherence marking: Feeling of Knowing or Illusion of Knowing, relating to the sense of understanding readers may have after reading a text. Glenberg, Wilkinson and Epstein (1982: 597) define the Illusion of Knowing phenomenon as follows: ‘The belief that comprehension has been attained when, in fact, comprehension has failed’. There is a discrepancy between factual understanding, objectively fixable, and the subjective assessment the reader makes.

This illusion of knowing is often investigated through a *recognition-of-contradiction* paradigm. If readers fail to see the contradiction, then there is an Illusion of Knowing. Epstein, Glenberg and Bradley (1984) focused on particular text characteristics that may cause an Illusion of Knowing; they have shown that readers are more likely to misjudge their comprehension, when the contradictions in the text concerned inferences. In some cases, the contradiction that they had put in the text was directly visible, and in other cases, the contradiction involved comparison with an unwritten inference. An example of an explicit contradiction example is the following text (14):

14. Plants have the ability to manufacture all necessary vitamins. Plants have strong protection against extinction. Plants have no strong protection against extinction.

EFFECTS OF COHERENCE MARKING: PREVIOUS RESEARCH - 51

Clearly, the last two sentences in this example form a direct opposition. An example of an implicit contradiction is the following text (15):

15. Plants have the ability to manufacture all necessary vitamins. The ability to manufacture vitamins is a strong protection against extinction. Plants have no strong protection against extinction.

In the second example, the opposition is between the inference that results from sentence 1 and 2, namely *plants have strong protection against extinction*. This inference is contradicted by the last sentence. Epstein et al. have shown that in example 15, readers were more likely to detect the contradiction and correctly assess their comprehension than in example 14. Since coherence marking also prompts readers to make an inference (see Chapter 1), coherence marking could lead to a higher accuracy in predicting one's comprehension performance.

The most relevant study involving both Feeling of Knowing and coherence marking was done by Rawson and Dunlosky (2002). They investigated to what extent readers based their estimated performance on the difficulty they experienced in text processing. In these experiments, participants predicted a better score on comprehension questions when they had read a text with coherence markers than when they had read a text without coherence markers. This indicates that the easier a text is to process, the higher people assess their comprehension. These studies provide indications for the importance of coherence marking for Feeling of Knowing, which leads to the following research question:

RQ 4: What is the effect of coherence marking on the reader's Feeling of Knowing after having read the text?
--

Based on the studies that I discussed in this paragraph, I expect coherence markers to cause such a wrongful feeling of having understood the text. Coherence marking is expected to make processing of the text easier. The presence of the markers gives a clear instruction on how to relate the two text segments. When there are no markers present, readers have to decide for themselves how the two text segments are related. Coherence markers improve comprehension and reader's assessment of their own performance. The Feeling

of Knowing that readers may experience after an explicit text is therefore justified. Readers are correct in assuming that they have understood the text better.

2.3 Persuasion

I started out with the (narrow) view of mental representations as text comprehension. Considering how rich and elaborate situation model representations are, and how factors such as opinions, previous experiences and opinions interact with the text to form a mental representation, I proposed to add appraisal and meta-cognition. We have seen in 2.3 that appraisal is a broad concept, and that I have chosen to address one of the dimensions separately, namely acceptability or persuasion.

2.3.1 Persuasion

Persuasion occurs when people change their attitude about a certain topic. However, attitudes are not the only relevant determinant of behavior. In their Theory of Reasoned Action, Fishbein and Ajzen (1975) distinguish between three main determinants of human behavior: beliefs, attitudes and intentions⁶. These levels are represented in Figure 2.2.

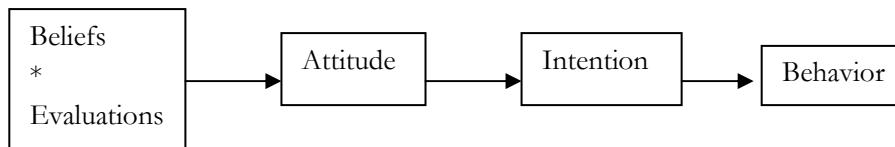


Figure 2.2: *Theory of Reasoned Action, Fishbein & Ajzen, 1975*

The levels in the diagram are related: the most likely predictor of human behavior, the last step in the diagram, is the intention to that behavior. For example, the most likely predictor of the behavior *recycling the garbage* is the intention to recycle the garbage, *I intend to recycle my garbage*. In a more indirect way, we can predict behavior when we know the attitude toward that specific behavior, in the case of our example the attitude towards recycling of the

⁶ In the context of this dissertation, I make use of a simplified version of the TRA model. Components such as *motivation to comply* and *subjective norm* are not included. They have been shown to be a less strong predictor of intention and behavior than attitude and belief (Eagly & Chaiken, 1993).

garbage, such as *I think that recycling the garbage is very important*. In most cases, this particular attitude towards recycling will lead to a positive intention and then to the desired behavior. But in some cases, it is possible to feel strongly about recycling but to still not show the desired behavior, because for instance *it is raining and the recycle-station is too far away*. The most indirect predictors of behavior in the diagram are beliefs. Beliefs concern facts, although people may hold beliefs that are not actually true. *The recycle station is three blocks away* could very well be true, but it could also be a mistake if the person in question is misinformed.

Throughout this dissertation, the persuasive texts always aim at changing behavior. They try to persuade the reader to act in a certain way, or to refrain from a certain behavior in the future. This does not mean, however, that texts that are simply aiming at changing an opinion are not to be considered persuasive texts. On the contrary, attitude change is very important and it is the key to persuasion. Behavior can only be changed if the attitude is consistent with this behavior. For instance, I can only convince someone to recycle their garbage if they first have a positive attitude about recycling. Still, the current focus on behavior is chosen for experimental reasons: all three determinants (beliefs, attitude and intention) can be included in the dependant variables.

2.3.2 *Coherence marking in persuasive contexts*

What do we know about the influence of coherence marking on persuasion? Does coherence marking increase or decrease the persuasive power? Not many studies have been done to investigate this aspect of coherence marking.

An interesting study on this subject has been conducted by Heller and Areni (2004). They used advertising claims to test the persuasiveness of several connectives. In the construction of the materials, they simply replaced one type of connectives with another type. For instance, they compared an advertising claim *The new brand X oven fan is very powerful, yet extremely quiet* to the constructed counterpart *The new brand X oven fan is very powerful and thus very quiet*. The design of experimental texts shows one of the pitfalls of connectives research: connectives cannot be simply 'swapped' around in order to compare supposed effects. It can lead to a mismatch of connective and coherence relation (see also Graesser, Millis, & Zwaan, 1997 for the same warning). In the current example from Heller and Areni (2004), the relation between being powerful and being quiet is simply not a causal, but a contrastive one. Marking this relation with the

connective *yet* makes sense, marking it with *thus* does not. Although the hypotheses and the idea of this study are very interesting, the unnaturalness of the materials makes it impossible to compare the persuasiveness of these different types of markers.

In a recent corpus study, Vivanco (2005) recently observed that in six technical advertisements, almost no connectives or coherence markers were present. Vivanco concludes that markers are probably omitted to keep texts as short as possible, for financial purposes, but also to maintain the consumer's attention. Although this corpus was very small and limited to the domain of academic texts only, the absence of connectives could mean that writers of the persuasive text type 'advertisement' expect them to have some negative effect on persuasion.

By contrast, we find several indications for a supposed persuasive effect of coherence marking in linguistic theory. Anscombe and Ducrot (1983) defined *mots du discours* as words that have no or little informative value, but mark the attitude of the speaker. Examples are *mais* (but), *même* (even), *décidément* (definitely). Coherence markers seem to fall in the category of 'linguistic argumentative operators': they clearly are words with an argumentative function in discourse. Cognitive linguists (Verhagen, 2005) as well as argumentation theoreticians (Snoeck Henkemans, 2001) have argued for a similar approach: discourse connectives are analyzed as contributing an argumentative value. Despite the analytical appeal of this idea, one crucial question has so far remained unanswered: what is the effect on the reader? This leads to the following research question:

RQ 5: What effect does coherence marking have on persuasion?

In this chapter, we saw how the Eiffel Tower text had the ability to change or form people's opinions, even though this was not an example of an explicitly persuasive text. This raises the issue of genre: are the effects of coherence marking different in one text genre than in another?

There are no clear expectations on this point for comprehension effects or appraisal effects. I expect coherence marking to act in the same way in both genres. However, I do see differences for persuasion effects: they are expected to be stronger in a persuasive than in an informative setting. In an informative context, readers can choose to change their minds. In a persuasive

context, they are forced to either maintain their original position or be influenced by the text. Either way, the text attempts to influence their opinions.

RQ 6: Are there differences concerning the effects of coherence marking in informative texts and in persuasive texts?

Of course, this raises yet another question: what exactly is the difference between these two genres? In the following section, I present both a theoretical distinction between these two genres and a distinction based on textual characteristics.

2.3.3 Distinction between informative and persuasive genre

The distinction between informative and persuasive texts is not an easy one to make. The most commonly used definition bases the distinction on the author's intention (for instance O'Keefe, 1990). However, this author intention is not always easy to identify. Some persuasive texts try to convince the reader by only providing information: does that mean the text is persuasive or informative? Obviously, the persuasive intent is not always easy to recognize, making it very complex to define the genre of a text by inferring the author's intent.

A solution would be to base our definition of persuasive text on a more objective criterion. Are there are textual characteristics that often or almost always occur in persuasive texts? A list of such characteristics can certainly help make the distinction between informative and persuasive texts clearer, especially in an empirical context. An overview of textual characteristics can form a more objective criterion for attributing a text either to the persuasive genre or the informative genre. In an empirical context, such a list can even help construct prototypical informative and prototypical persuasive texts in order to investigate genre and its influences. In the following sections, several studies that discuss (textual) features of persuasion are presented.

2.3.4 O'Keefe's common features of persuasion

O'Keefe (1990) gives in the first chapter of his book *Persuasion* a list of 'common features' of persuasion. These features do not necessarily apply to texts, but they can offer a starting point for a list of criteria for persuasive texts.

The first feature that O’Keefe presents is that of success. In O’Keefe’s view, we can only speak of persuasion if we succeed in our attempt to do so. O’Keefe’s second and third features are related: the presence of a goal or criterion, and the existence of the intent to reach that goal. For instance, the Amnesty campaign against the death penalty has as a goal that the death penalty has to be abolished, and the criterion could be that people sign the petition in order to prevent the death penalty. The fourth feature concerns the freedom that the receiver has. He or she has to have some degree of free will in the decision to take. This means, for instance, that I cannot persuade someone to stop sleeping; this is something we do not have free choice over. The fifth characteristic concerns the fact that persuasion has to involve communication. Persuasion has to be achieved through communication. Forcing someone to take a test or else fail the class does not mean this person has been persuaded to take the test. The sixth and last characteristic O’Keefe mentions the fact that there has to be a change in the mental state of the receiver. There can be *only* changes in mental states or attitudes, but they could possibly lead to behavioural changes.

2.3.5 Biber’s textual dimensions

Although the characteristics by O’Keefe help forming a theoretical framework for persuasion, they are not clear *textual* features that are prototypical for persuasive text.

Biber’s (1988) approach is helpful when defining prototypical features for persuasive texts. Biber described linguistic characteristics of different types of texts in English. He explained variation between these types by using the notion of textual dimensions. With a multivariate analysis, he identified which of 67 linguistic features typically co-occur, in order to define these dimensions. Each dimension has positive and negative features.

The first dimension is *Involved* versus *Informational*. The Involvement pole is characterized by private verbs (feel, think, believe), that deletion, wh-questions and first person pronouns. Text features that are essential for the informational pole are, among others, nouns, long words and attributive adjectives. The second dimension is *Narrative* (past tense, third person pronouns) versus *Non-Narrative* (present tense). The third dimension is *Explicit* versus *Situation-dependent reference*. The situation dependent reference makes use of time and place adverbials, whereas the explicit reference is independent from

a specific situation. The fourth dimension is *Overt expression of persuasion*. This dimension, contrary to the others, only has positive features. Biber states that this dimension marks the degree to which persuasion is marked overtly. Linguistic features for this dimension are modals (events *will, should, can* or *might* occur), suasive verbs (*command, demand, instruct*), conditional subordination (specifies the conditions that are necessary in order for certain events to occur), infinitives that mark the speaker's attitude towards the proposition in the infinitive clause (e.g. *happy to do it*). In other words, this dimension marks whether a speaker's own point of view is marked or not.

Biber specified that within a genre, there can be large variation between texts. This is specifically the case with Overt Expression of Persuasion. Biber analyzed two different types of persuasive texts: editorials and academic texts. Some of these persuasive texts argue for a particular point of view, whereas other texts persuade in a more covert way. Biber concludes that these texts can range from extremely persuasive and argumentative to markedly non-persuasive.

A study by Connor & Upton (2003) also provides more insight into the prototypical characteristics of persuasive texts. They applied Biber's dimension to yet another specific type of persuasive texts: direct mail letters. According to this study, we can expect informative and persuasive texts to be informational, non-narrative and refer to explicit situations. Thus, it is difficult to base the distinction between the genres on these three dimensions. This means that the important category for distinguishing between the two genres seems to be Overt Expression of Persuasion. In the study by Connor and Upton (2003), the score of direct mail on this dimension is not decisive: there are some linguistic tools present in the text from this dimension, but not all. Again, this seems to be the relevant finding here from both Biber (1988) and Connor and Upton (2003): they have provided a list with textual features, but these features are not expected to be *all* present in persuasive texts. These features add to the persuasive nature of a text: the more features are present, the more the text approaches a prototypical persuasive text.

2.3.6 Other prototypical characteristics of persuasive texts

In addition to Biber's characteristics, there are some other aspects that are certainly important in persuasive texts. These factors are expected to be present not in all persuasive texts, but in more *successful* persuasive texts. In other words,

these characteristics are not necessarily present, but when they are, they supposedly make the text more persuasive. For O’Keefe, a text is only persuasive if it succeeds in persuading the reader (see 2.4.4). These characteristics would then be part of the prototypical characteristics of persuasive text. However, following Hoeken (1998), I do not completely agree. In the context of this dissertation, a text is considered to be persuasive, even though it does not succeed in its attempt to change behavior or attitude. Still, the characteristic features of successful persuasive texts could provide more insight into the distinction between informative and persuasive texts, so I have chosen to include them in this overview.

One of these aspects concerns the explicitness of conclusions. O’Keefe (1997) conducted a meta-analysis in order to be able to determine whether standpoint or conclusion explicitness would be a sensible strategy in persuasive communication. Since persuasive texts have a goal, should this goal be made explicit or not? Messages with explicit standpoints are in general more convincing than messages without an explicit standpoint. Therefore, the feature of standpoint explicitness is also included in the list of prototypical persuasive features.

A difficult matter is that of emotion. Murphy (2001) makes a strong case for evoking the emotion and affect of the reader. Arguments alone, so she says, will not convince the reader. However, readers have all sorts of psychological reactions at their disposal, one of which is denial. An emotional appeal if it is too strong will not lead to persuasion but to rejection. However, it is a characteristic often found in persuasive texts.

Anscombe and Ducrot (1983) define *mots du discours*, words that have little informative value, but mark the speaker’s involvement. Examples are *mais* (but), *d’ailleurs* (by the way) and *au moins* (at least). These words are more likely to be present in the persuasive genre than in the informative genre, although they can be present in both genres.

A last, crucial remark that needs to be made concerns the *appearance* of a text: how is it presented? As a brochure? As a newspaper article? As an infomercial? Petty and Cacioppo (1981), Schellens and de Jong (2004), Koelen and Martijn (1994), Murphy (2001), and many more researchers in the field of persuasion have concluded that persuasive texts are often presented as informative text. This means that writers of persuasive texts try to make them look as informative texts. This makes the distinction that I am trying to make in

this paragraph problematic and clearly is a difficult aspect when we want to define crucial textual features for persuasive texts: the texts cannot be too clearly persuasive, for then they will not succeed in their attempt to influence attitudes and behavior. But we can say that when more features of persuasive text are present in the text, it resembles a prototypical text in its genre.

2.3.7 Pragmatic relations and pragmatic argumentation

An important concept in persuasion is pragmatic argumentation, or pragmatic relations. Sanders (1997) showed a clear link between text type and type of coherence relations. He concluded from a small corpus study that informative texts are dominated by semantic relations, meaning that discourse segments are mainly related because of their propositional content. By contrast, persuasive and expressive texts showed a higher quantitative and qualitative occurrence of pragmatic relations. Pragmatic relations exist because of the language user's goal oriented communicative act; the connection is at the illocutionary level.

In Chapter 1, I explained the distinction between pragmatic and semantic relations by using the terms objective and subjective relations. The following examples are taken from the corpus in Sanders (1997). The first example is taken from an encyclopedia, considered a prototypical informative text by Sanders, and the second example is taken from an advertisement in a tabloid magazine, a prototypical example of a persuasive text.

16. Immanent. Staying in. Indicating what is typical of a certain area and does not exceed it.

17. Constipation is unpleasant and inconvenient. So you'd better do something about it.

In 16, we see a relation of ELABORATION, a description of characteristics of a concept that was introduced earlier. This is an example of an objective or semantic relation, typical for informative texts. In 17, there is a clear ARGUMENT-CLAIM relation. It does not describe a state-of-affairs in the world, rather, the writer argues towards a conclusion and even urges the reader to do something. Here, the relation is between a negatively evaluated situation which leads to a speech act of the writer: 'do something about it!' This is a subjective or pragmatic relation, prototypical for persuasive texts.

The correlation that Sanders showed between text type and relation type indicates another prototypical characteristic of persuasive texts: they are

dominated by pragmatic relations, more than informative texts. In the remainder of this dissertation, this type of relations will be referred to as *subjective* relations, whereas semantic relations are referred to as *objective* relations.

Findings like Sanders' (1997) were corroborated in a corpus study by Schellens and de Jong (2004), who showed that one characteristic is always present in persuasive communication: pragmatic argumentation, also referred to as argumentation from consequences. This type of argumentation typically stresses advantages to the desired behavior or disadvantages to the non-desired behavior⁷. The following text passage (18) is illustrative for pragmatic argumentation.

18. If you take enough exercise, your muscles, lungs, heart and arteries are used to working. And that feels a lot better. You feel fitter and more energetic – energetic enough to do lots of enjoyable things, for example. (2004: 304)

This text passage focuses on the positive effects of the desired behavior. In my view, pragmatic relations, and more specifically pragmatic argumentation is central to the opposition between informative and persuasive texts.

In conclusion, I have presented several prototypical textual features of persuasive texts. In Table 2.2, these features are repeated. These features help base the distinction informative-persuasive on a more objective criterion than simply inferring the author's intent. If more of these prototypical features are present in a text, it resembles a prototypical persuasive text. If they are not present, then the text is more likely to be perceived by the reader as informative. The list in Table 2.2 (see the next page) forms the basis for the experimental distinction between informative and persuasive texts throughout this dissertation.

Finally, it is important to stress that these features are not *necessary* to make a text a persuasive one. Without these features, a text can still be persuasive and have as a goal to influence opinions and behaviors. But these features are more likely to occur in a persuasive text than in an informative text, and the more they occur in a text, the more a text resembles a prototypical

⁷ If we relate pragmatic argumentation to attitudes instead of behavior (cf. TRA model of Fishbein and Ajzen), it focuses on advantages of a certain object or goal, and disadvantages of a certain object or goal. For instance, if people want to argue that *La vita è bella* is the most beautiful movie they have ever seen, they can focus on advantages of seeing it or the disadvantages of not seeing it (*you would really miss out on something!*)

EFFECTS OF COHERENCE MARKING: PREVIOUS RESEARCH - 61

persuasive text. The more these features are absent, the more a text resembles a prototypical informative text. In Chapter 6 and 7, this issue will be addressed in more detail.

Study	Features for the persuasive genre	Features for the informative genre
1. O'Keefe	Intention is to change mental state or behaviour	Intention is to provide information
2. Biber : Overt expression of Persuasion	Modals (should, might, can) Suasive verbs (demand, ask) Infinitives in constructions like happy to do it, hoped to see i	- (no negative pole in this dimension)
3. Sanders: Pragmatic relations	Dominance of pragmatic relations	Dominance of semantic relations
4. Schellens & de Jong: Pragmatic argumentation	Pragmatic argumentation Argumentation cause-effect Argumentation from example	No argumentation
5. O'Keefe: Standpoint	Explicit standpoint	No explicit standpoint
6. Murphy: Emotion	More Emotion	Less emotion
7. Anscombe & Ducrot: Informative value	More Mots du discours	Less mots du discours

Table 2.2 *Prototypical textual features for the persuasive and the informative genre*

2.4 Innovative aspects of this dissertation

In sum, a lot of empirical work has been done to determine the effects that coherence markers can have on the mental representation that a reader forms of the information in the text. There seems to be an interaction between coherence marking and prior knowledge, but a final conclusion cannot be drawn yet, because of several methodological flaws, among them situation model representation tasks. In this dissertation, I set out to find the most suitable methodology in order to measure situation model representations (RQ2). Consequently, it will be possible to test the interaction hypothesis of prior knowledge and coherence marking on text comprehension (RQ1).

In addition, I want to broaden the scope from studying the mental representation by only including aspects of text comprehension, to a view where appraisal (RQ3), meta-cognition (RQ4) and persuasion (RQ5) are included. Situation model representations are formed at the interaction of the text and these very factors: opinions, knowledge, experiences. Together, they

form the ingredients for the mental representation that readers construct from a text. It makes sense to include these aspects when we investigate influences of any particular textual characteristic on the mental representation.

The final addition that I propose here is to include different text genres (RQ6). Although informative texts are certainly able to influence people's opinions, persuasive texts are even more likely to do so. Therefore, both genres are included in the research in this dissertation.

In short, the approach in this dissertation adds to existing knowledge about linguistic coherence marking and its effects in three ways: (1) by including an important reader characteristic, prior knowledge, (2) by including more than one aspect of text representation such as affective and meta-cognitive components and (3) by comparing these results between two different text genres.

This chapter presents the results from a pilot experiment on the effects of coherence marking on the mental representation, including text appraisal, Feeling of Knowing and persuasion, in addition to comprehension. We use texts in two different genres, which allows us to make a direct comparison. The results from this first experiment show 1) the importance of broadening the view of mental representation including factors such as appraisal, feeling of knowing and persuasion and 2) the usefulness of a genre manipulation that permits a direct comparison between both genres.

⁸ Previously published in Dutch with minor editorial changes in *Tijdschrift voor Taalbeheersing*, with co-authors Leo Lentz and Ted Sanders (Kamalski, Lentz, & Sanders, 2004).

The goal of this pilot experiment is twofold. The first goal is to investigate the effect of two independent variables, coherence marking and prior knowledge, on the situation model representation. We operationalize the situation model representation with four dependent variables: comprehension, appraisal, meta-cognition and persuasion (see Chapter 2). The hypotheses presented in this paragraph all result from the overview in Chapter 2, in which I argued that research concerning influences of text characteristics on the reader needs to broaden the view on representation.

The second goal of the pilot experiment is to test whether it is possible to manipulate the factor text genre in such a way that the informative and persuasive texts still contain largely the same information, but differ only in the context in which this information is presented. The advantage of keeping the text content constant is that the same tests and questions can be used for the experimental texts in both genres so we can compare effects on comprehension directly between genres. This is why in the pilot experiment all the effects of coherence marking and prior knowledge on the three dependent variables mentioned above are tested in two different text genres: the informative and the persuasive genre.

3.1 Variables in Pilot Experiment 1

In this section, the independent variables coherence marking and prior knowledge and the dependent variables comprehension, appraisal, Feeling of Knowing and persuasion are outlined.

3.1.1 Coherence marking and prior knowledge

Chapter 1 defined and clarified the notion of coherence and how it can be linguistically marked. This definition led to five categories: headings, organizers, lexical cue phrases, connectives, and referential coherence (see 1.3.1). All of these categories of linguistic coherence markers are included in the current pilot experiment. In this pilot experiment, the assumption is that we can generalize over coherence relations and coherence markers. Different types of coherence relations, such as causal and additive relations, or positive and negative relations (see 1.2) are marked. Also, different types of markers, such as connectives, lexical cue phrases and headings are used (see 1.3). For now, the working

hypothesis is that these markers have a general effect on comprehension, appraisal and persuasion. In Chapter 6 and 7, I will return to this issue.

Previous research has shown that coherence marking and prior knowledge interact: for low knowledge readers, the presence of coherence markers lead to a higher comprehension score; however, for high knowledge readers, the absence of coherence marking leads to a higher comprehension score (for instance McNamara, Kintsch, Songer, & Kintsch, 1996 and section 2.1.3 in this dissertation). It is therefore important to include the reader characteristic prior knowledge in studies investigating coherence marking. In this experiment, an expert-novice paradigm is used: two groups of participants are selected based on the expectation that their prior knowledge differs.

3.1.2 Comprehension

The first dependent variable in this study is text comprehension, or, in other words, the traditional definition on the mental representation of a text. Readers construct mental representations of texts at three different levels: the surface code, the text base and the situation model representation (see 2.2). At the level of the situation model representation readers integrate textual information with their prior knowledge. Since we are interested in the possible interaction of coherence marking with prior knowledge, the situation model representation is relevant here (for more discussion on this issue, see 2.2.3).

With regard to the dependent variable comprehension, I expect to replicate the interaction effect reported by McNamara and Kintsch (1996) and McNamara et al. (1996), with carefully constructed texts that vary systematically in coherence marking across conditions (as explained in 1.2.2). Hence Hypothesis 1 can be formulated as:

H 1: Readers with more prior knowledge perform better on comprehension questions after reading an implicit version. Readers with less prior knowledge perform better on comprehension questions after reading an explicit version.

3.1.3 Appraisal and Feeling of Knowing

I argued in Chapter 2 that when a study aims at investigating effects of a text characteristic on the mental representation more communicative effects need to

be included than comprehension alone. Two of these aspects that I include in this pilot experiment are appraisal and Feeling of Knowing.

Text appraisal is a subjective judgment of a text, related to the goal of the reader (for more information see 2.3). Garbarino and Edell (1997) found that participants who had to invest more energy into the processing of an advertisement also had a more negative judgment of this advertisement. Because coherence marking supposedly makes processing of a text easier, this consequently has a positive effect on text appraisal, compared to the text in which coherence marking is absent. This supposition leads to the following hypothesis:

H2: When coherence remains implicit, reading demands more cognitive energy from a reader. Readers find this to be annoying, and thus, in comparison to the explicit version of the text, the implicit version is judged more negatively.

Feeling of Knowing, a meta-cognitive judgment that is closely related to appraisal, is the impression that readers may have of understanding the text when, in fact, comprehension has failed (Glenberg, Wilkinson, & Epstein, 1982; Rawson & Dunlosky, 2002). Coherence markers are expected to cause a feeling of having understood the text, but this impression is correct for low knowledge readers. For high knowledge readers, the impression of having understood the explicit text better than the implicit text is wrong: they are actually expected to perform better on comprehension questions after having read the implicit version. Hence, the FOK hypothesis can be formulated as follows:

H3: The explicit versions cause a greater Feeling of Knowing than the implicit versions. This impression is correct for low knowledge readers, but not justified for high knowledge readers.

3.1.4 Persuasion

Relatively little is known about the effect of coherence marking on persuasion. This part of the pilot-study is therefore more explorative in nature than the previously presented hypotheses.

Brown and Stayman (1992) reported that many recent studies suggest that appraisal of an advertisement may be the best indicator of advertising effectiveness. This finding could very well apply not only to advertisements, but also to persuasive texts in general. In the case of coherence marking in persuasive texts, the expectation would be that the explicit versions are more persuasive than the implicit versions, because they are evaluated more positively (see hypothesis 2a) and therefore are more convincing.

However, the contrary prediction is also a plausible option. When readers become aware of an attempt to influence them, they build resistance to the text and become more difficult to persuade. This is the so-called forewarning effect (see Chen, Reardon, Rea, & Moore, 1992; Romero, Agnew, & Insko, 1996). In this situation, the implicit versions are expected to be more persuasive than the explicit ones, because the reader is less aware of the attempt to influence his or her opinions and therefore less resistant to the persuasion.

3.2 Pilot Experiment 1

In order to investigate the effects of coherence marking and prior knowledge on comprehension, appraisal, FOK and persuasion, an off-line reading experiment was conducted.

3.2.1 Materials

The persuasive texts had to meet one important criterion: both comprehension and persuasion had to be measured within one experiment. Public information brochures meet this demand, because they are used to change cognitions, attitudes and behavior by providing information (Kok & Damoiseaux, 1996). Both topics selected for the pilot experiment, genetic manipulation and organ donation, lead to persuasive texts that try to convince the reader by incorporating information and facts in the text. This characteristic makes them ideal texts to include in an experiment in which both comprehension and persuasion are measured.

For each topic, both an informative and a persuasive version were constructed, based on existing materials from brochures and websites. The textual characteristics that are prototypical for persuasive text, presented in 2.4.3 and further, formed the basis of the distinction between informative and persuasive texts. An example of such a characteristic is pragmatic argumentation, also referred to as argumentation from consequences. This type

of argumentation typically stresses advantages to the desired behavior or disadvantages to the non-desired behavior. An example of stressing advantages of the desired behavior is *by stopping to eat genetically manipulated foods, genetic manipulation will eventually disappear and the diversity in nature will stay intact*. An example of stressing the disadvantages to the non-desired behavior *By eating genetically manipulated foods, you will contribute to the loss of natural diversity*. These examples occur, although not literally, in the persuasive texts. The informative texts did not contain pragmatic argumentation. A text passage from both the informative and the persuasive organ donation text are provided in Appendix 1.

The four texts (informative and persuasive for both the organ donation and genetic manipulation topics) were subsequently manipulated to create an implicit and an explicit version of each. Marking of coherence is taken in a very broad sense in this pilot experiment. For now, I generalize over different coherence relations that are marked, as well as over different types of markers that are used to mark these relations (see Chapter 1), because I do not expect these factors to have different effects on the reader. The following aspects have been manipulated in the texts: global coherence (headings and organizers), relational local coherence (connectives and lexical cues), and referential local coherence (repeating the antecedent). In Table 3.1, examples of these manipulations for the organ donation text are given. Also, in Appendix 1, the coherence markers are underlined.

Category	Explicit version	Implicit version
Headings	Why register as an organ donor?	No heading
Connectives	People have to wait a long time, <i>because</i> there are not enough organs available.	People have to wait a long time. There are not enough organs available.
Lexical cue phrases	Whether or not you chose to donate your organs, is centrally registered. <i>This is the reason that</i> codicils are not necessary anymore.	Whether or not you chose to donate your organs, is centrally registered. Codicils are not necessary anymore.
Referential coherence	Establishing brain death is only possible when a person dies from brain damage. <i>Brain damage</i> can be caused by a traffic accident.	Establishing brain death is only possible when a person dies from brain damage. <i>This</i> can be caused by a traffic accident.

Table 3.1 *Examples of coherence markers in the genetic manipulation text*

3.2.2 Pretest

The texts and operationalizations were analyzed in a reader-focused pretest (see for the distinction reader-focused and expert-focused de Jong & Lentz, 1996; Lentz & De Jong, 1997)⁹. The first aim of this pretest was to test whether the experimental texts were natural enough. Texts have to be natural in order to eliminate the risk of including so-called *textoids* (Graesser, Millis, & Zwaan, 1997) or texts that have clearly been constructed for experimental purposes and do not resemble real-life texts. Second, the pretest tested the genre manipulation. Were the informative and persuasive texts constructed in a successful way? Is it possible to keep the content constant, but vary the context in order to create informative and persuasive texts on the same topic? This second goal is related to the first; again, the informative and persuasive texts have to be both similar and natural.

The text versions were pre-tested on a group of readers without prior knowledge of the text topic (n=8). They used the plus-minus method (de Jong, 1998) to comment extensively on the texts. Plus-minus methods are used to locate points in the texts where readers experience problems on the one hand and to test the naturalness of a text on the other. The plus minus method is conducted as follows: participants read the text and note plusses and minuses at places where they have positive or negative comments on the text. In the current pretest the exact location and nature of the plusses and minuses is not the focus¹⁰, but rather we focus on the overall report on the naturalness of the texts.

In the pretest, no remarks were made about the naturalness of the texts. When asked specifically to identify the writer, participants named newspapers for the informative texts and both the government and Greenpeace for the persuasive texts. None of the participants thought that the text had

⁹ An expert-focused pretest was also conducted, but this pretest mainly focused on the quality of the coherence marking manipulation. Several experts provided comments that helped improve manipulations. Their opinions were helpful in the sense that the texts could not be 'stuffed' with coherence markers, and the experts helped strike the right balance. Here, we will not go into the expert focused pretest, only the reader focused pretest is described in this chapter.

¹⁰ Although the texts were considered natural, participants did make some remarks concerning the manipulation of coherence marking, such as *I miss the headings* or *This jumps form one subject to another*. However, these remarks were only made by some of the participants and they still considered the texts to be natural.

been written for experimental purposes. This leads to the conclusion that the texts were natural enough to be part of the pilot experiment.

A second important question of the pretest was the difference between informative and persuasive texts. Was the manipulation of genre successful? In order to answer this question, the participants categorized the texts on a 4-point scale based on their understanding of the communicative intention of the writer: What do you think was the intention of the author? To inform or persuade? The results are presented in Table 3.2. The informative genre results in lower scores than the persuasive genre ($\chi^2(df=3)=92.9, p<.0001$). The results show that readers are able to make a clear distinction between the genres. Also, this distinction closely follows the distinction that was made by experts. This leads to the conclusion that the manipulation of genre was successful, in the sense that readers and experimenters classified the texts in the same way.

Text version	Mean score
1. Genetic Manipulation Informative	2.00
2. Genetic Manipulation Persuasive	3.33
3. Organ Donation Informative	1.33
4. Organ Donation Persuasive	3.50

Table 3.2 *The distinction informative – persuasive on a 4-point scale (1=to inform, 4= to persuade, n for each text=8)*

We can conclude from the pretest that the text versions were natural enough, and that the informative-persuasive manipulation seemed to be successful. The reader clearly perceives the texts as either informative or persuasive.

3.2.3 Dependent measures

In this paragraph, the measures that were used to assess the dependent variables comprehension, appraisal, Feeling of Knowing and persuasion are described.

Measuring the situation model representation can be operationalized in many ways (see 2.2.4 and Kamalski, 2004). An often used, rather traditional method is the open bridging inference question (among others McNamara, Kintsch, Songer, & Kintsch, 1996). In Chapter 2, we have seen an example of this type of question (example 8). Answering bridging inference questions

requires integrating at least two sentences from the text. This method is known to be sensitive enough to show an effect of coherence marking: Degand, Lefevre en Bestgen (1999) and Degand and Sanders (2002) used this method when they established an effect of coherence marking on text comprehension. McNamara et.al. (1996) used this method to establish an interaction between coherence marking and prior knowledge. In the current pilot experiment, four bridging inference questions were constructed for each text. The following passage is taken from the materials in this pilot-study. The text (1) is about genetic manipulation, and the question (2) is an example of one of the bridging inference questions. Both have been translated from the original Dutch materials.

1. Genetically manipulated organisms can cross with natural plants. The new organisms that arise because of this are resistant against insects and pesticides, whereas the natural organisms are not. The natural plants die out and as a consequence the bio-diversity of the natural flora is threatened, as well as biological agriculture that wants to remain free of genetic manipulation.
2. Question: The text states that genetic manipulation can possibly endanger the diversity of natural flora. Explain in what way.

In order to answer this bridging inference question correctly, a reader has to link information from the second sentence (*pesticides are only dangerous to natural organisms*) to information from the third sentence (*natural organisms die, diversity decreases*) and establish a causal relation between these events.

Measuring text appraisal is often done with a list of 7point Likert scales with semantic differentials, such as *difficult – easy*. This method has often been used and tested. Examples of Dutch studies that have used this methodology are van Wijk (1996), Oversteegen, Keulen and van Wijk (2002) , and Land, Sanders, Lentz and van den Bergh (2002). The items from their lists were taken together and assigned to one of the categories of text appraisal that I discussed in paragraph 2.3, namely *appealingness* and *accessibility*. Previous research has shown that not all scales within one dimension yield the same outcome. In the study by Land et al., participants found the explicit version clearer, but the implicit version easier. In order to be able to find these apparent contradictions, more than one scale was selected for each category. The selected items are represented in Table 3.3 (on the next page).

These semantic differentials are presented to the participants with a 7-point Likert scale in the following way:

3. Do you think this text is
 very difficult 1 2 3 4 5 6 7 very easy

Dimension	Operationalization	
Appealingness	Clear	Vague
	Readable	Not readable
	Credible	Not credible
	Professional	Not professional
	Reliable	Not reliable
Accessibility	Difficult	Easy
	Effortful	Effortless
	Simple	Complex
	Coherent	Not coherent
	Halting	Fluent

Table 3.3 *Operationalization of the evaluation scales*

FOK was operationalized as follows: participants had to indicate on a 4-point scale how well they thought they understood the text. This method is identical to the one used by Glenberg, Wilkinson and Epstein (1982), but different from Rawson and Dunlosky (2001). In the latter study, participants had to assess their own chances of having provided the correct answer after every comprehension question. This operationalization of FOK is more related to the question answering process than to the text processing itself. Because we want to gather information on the reading process, the first method was the better choice.

Persuasion is generally measured at different levels, based on the distinction in the Theory of Reasoned Action (Fishbein & Ajzen, 1975): beliefs, attitudes and intentions (see Figure 2.2 in this dissertation). In the current pilot experiment, all three levels of persuasion are measured: beliefs, attitude and intention. Although intention is the most likely predictor of behavior, the central measure of persuasion is *attitude change*. Therefore, attitude level is the central focus of persuasion in this experiment, but the surrounding levels are also measured to increase the likelihood of finding an effect.

In this study, the three levels of persuasion are operationalized by means of statements, which the participant could agree or disagree with on a 7-point Likert scale. An example of a belief-statement on genetic manipulation is:

genetic manipulation decreases natural variance in plants. A possible attitude-statement is genetic manipulation is bad. From the attitude follows the intention-statement I intend to eat only biological foods.

3.2.4 Participants

Seventy-nine students of Utrecht University participated in this experiment: 26 of them History students, 53 of them Biology students. The average age for the History students was 20.6 ($SD=2.0$), for the Biology students 20.3 ($SD=2.0$). Of all History students 42.3% were male, for all Biology students 33.3% were male. Prior knowledge was not manipulated in the experiment, but the Biology students were expected to know significantly more about genetic manipulation and organ donation. This assumption was controlled for by asking 4 prior knowledge control questions, before participants read the texts in the pilot experiment. These questions assessed basic knowledge about human anatomy and genetic manipulation. A t-test showed that the groups indeed differed significantly on prior knowledge ($t(78)=-10.59$, $p<.001$); Biologists scored higher than History students. Thus, the a priori groups represent two distinct levels of prior knowledge, and are considered as low knowledge and high knowledge readers for the following analyses.

3.2.5 Distinction informative-persuasive

The characteristics from Chapter 2 were used to construct an informative and a persuasive text version (see also 3.2.1). In order to enhance the genre distinction even further, the instructions were varied between the informative and persuasive condition. Prior to the persuasive text, participants read: 'Read this text as if you were to take part in a discussion afterwards'; prior to the informative text, they received the instruction to 'read as if you were to take an exam afterwards'. Furthermore, text construction clearly aimed at signaling the genre from the very beginning of each text. Clearly persuasive titles (*You too can save lives*) versus informative ones (*Organ donation in The Netherlands*) contributed to the distinction.

Both in the pretest and in the experiment, participants appeared to be extremely accurate in recognizing the appropriate genre. At the end of the experiment, they were asked to rate the writer's goal on a 4-point scale between informing and persuading. The participants classified the texts according to the experts' classification of informative and persuasive: ($t(78)=-9.95$, $p<.001$).

3.2.6 Design

In this study, the experimental factor *prior knowledge* is a between subjects factor with values *high* and *low*. The experimental factor *text version* is a between subjects factor with the values *implicit* and *explicit*¹¹. Furthermore, there are two genres: informative and persuasive, and two text topics: genetic manipulation and organ donation. In total, 8 different texts were used in this experiment: 2 topics * 2 versions * 2 genres. These factors were integrated in a Latin Square Design: every participant read 2 texts, one of which implicit and the other explicit, one of which persuasive and the other informative, one of which on genetic manipulation and the other on organ donation. The experiment took about 40 minutes. Participants were instructed not to turn back the pages they had already read. Each package contained: the prior knowledge questions, the instruction to the text, the text itself, FOK questions, appraisal items, comprehension questions, persuasion statements, and finally the genre-control question.

A pretest-posttest design in order to establish attitude change was deliberately *not* chosen in this experiment. Such a design would have made it possible to compare initial attitude with attitude after the text. However, when readers are asked to formulate their own original existing attitude, the chances diminish that there will be any persuasion at all. Consciously establishing one's own attitude automatically strengthens it (see for instance Hoeken, 1995). Therefore, a comparison was made between the average attitude after reading the implicit text and the average attitude after the explicit text, between subjects. These implicit and explicit conditions were randomly assigned, so we can assume that the average attitude for both groups was equal before reading the text. In both groups, there were participants who disagreed with the text's point of view, participants who agreed with the text's point of view, and participants without a clear opinion on the text topic. Any difference between groups *after* reading the text has to be attributed to the text version.

¹¹ Every participant read one implicit text and one explicit text. Why was the factor text version not analyzed as a within subjects factor? The reason for this is that I wanted to make a distinction between the two text genres. Every participant read one informative and one persuasive text. In order to separate the genres in the analysis, text version needed to be considered a between subjects variable.

3.3 Results

Each hypothesis is analyzed for the informative and persuasive genre separately. The factor text topic appeared not to influence the results, so the results are collapsed over text topic. Two-way ANOVAs were calculated to test the effects of coherence marking and prior knowledge on the dependent variables.

3.3.1 Comprehension

The answers on the open-ended comprehension questions were scored as follows: 0=incorrect, .5=doubtful, 1=correct. This did not cause any problems, since most of the answers clearly fell in category correct or incorrect. The doubtful answers were scored with the help of an expert in biology, and they were mostly assigned to category doubtful. Internal reliability between the 4 questions was measured with Cronbach’s alpha: for the genetic manipulation text, $\alpha=.33$, for organ donation, $\alpha=.45$. These results are disappointing (compare van Wijk, 2000). This issue of reliability will be addressed in the Discussion and in Chapter 4. In the meantime, the effects of coherence marking and prior knowledge on text comprehension will still be analyzed by combining the open-ended questions in the analyses.

Table 3.4 shows the effects of coherence marking and prior knowledge on comprehension questions for the informative and for the persuasive texts.

	Informative Implicit version	Informative Explicit version	Persuasive Implicit version	Persuasive Explicit version
Less prior knowledge	1.39 (.77)	1.93 (.76)	1.46 (.78)	1.92 (.76)
More prior knowledge	2.44 (.75)	2.29 (.76)	2.22 (.75)	2.39 (.85)

Table 3.4 *Mean scores and SD on comprehension for informative texts (0= poor, 4= perfect comprehension)*

In the informative genre, an interaction-effect of marking of coherence and prior knowledge occurred on comprehension ($F(1,77) = 3.72, p<.05, \eta^2=.05$). Low knowledge readers perform better after reading the explicitly marked text than after reading the implicit version ($t(24) = -2.58, p < .05$). High knowledge readers perform equally well after reading either version ($t(52) = .06, p > .9$). There is also a significant main effect of prior knowledge ($F(1,77) = 15.49, p <$

.01, $\eta^2 = .17$): high knowledge readers perform better than low knowledge readers. Figure 3.1 depicts this interaction in the informative genre.

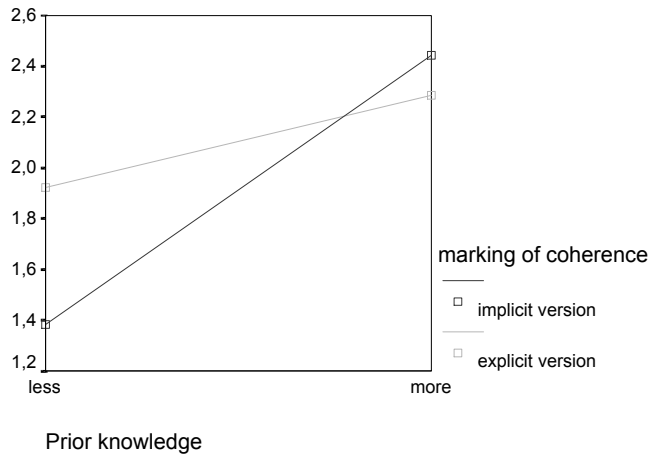


Figure 3.1: *Effect of marking of coherence on comprehension for readers with more and less prior knowledge (mean, 0 = poor, 4 = perfect comprehension)*

For persuasive texts, the situation differs, as is represented in the last two columns of Table 3.4. There is no interaction effect, but there is a main effect of marking of coherence ($F(1,75)=2.71$, $p=.05$, $\eta^2=.04$) and a main effect of prior knowledge ($F(1,75)=10.40$, $p<.01$, $\eta^2=.12$). On average, high knowledge readers perform better than low knowledge readers. Moreover, for both knowledge groups, comprehension scores are higher after having read the explicit version than after the implicit version.

3.3.2 Text appraisal

In Chapter 2, I argued that communicative effects such as appraisal and Feeling of Knowing need to be included if we want to get a complete picture of the functioning of coherence marking. In Table 3.5, the results are presented for the effect of coherence marking on text appraisal. Reliability was calculated with Cronbach's alpha (appealingness $\alpha=.88$, accessibility $\alpha=.82$). The reliability for these scales is good (van Wijk, 2000). The reliability for all dimensions together is also good ($\alpha=.85$).

The expectation was that the explicit version would cause a more positive appraisal than implicit versions. If we take all the items together and

calculate a mean appraisal score, then this expectation is confirmed ($F(1,155)=3.48$, $p<.05$, $\eta^2=.04$). The mean score for explicit texts is 2.0, the mean score for implicit texts 2.31, on a scale from 1 (very positive) to 7 (very negative). This effect holds for both the informative as the persuasive text.

Previous research has shown that different items can yield different results. Therefore, in Table 3.5, we differentiate between the different items and dimensions. Only significant results are included in the Table: in all cases, the difference between explicit and implicit texts is significant ($p<.05$). Not all 11 items are included, for reasons of clarity.

Dimension	Item	Informative e Implicit	Informative Explicit	Persuasive Implicit	Persuasive Explicit
Appealing	Clarity	2.37 (1.17)	1.77 (.66)	(n.s.)	(n.s.)
	Readability	2.23 (1.17)	1.80 (.76)	2.43 (1.15)	1.97 (.93)
	Reliability	(n.s.)	(n.s.)	3.53 (1.36)	2.95 (1.41)
Accessible	Simplicity	3.00 (1.20)	2.35 (1.08)	2.58 (0.98)	2.97 (1.20)
	Effort	2.48 (1.36)	1.90 (.96)	(n.s.)	(n.s.)
	Coherence	2.70 (1.14)	2.20 (.97)	3.02 (1.56)	2.49 (1.19)

Table 3.5 *Effects of marking of coherence on text appraisal: mean scores and SD on a 7-point scale (1=positive, 7 = negative)*

In the informative texts, appraisal of the implicit version is more negative than appraisal of the explicit version. This is in accordance with hypothesis 2a. In the persuasive texts, the effects do also occur, but they are less distinct. Still, the implicit version is judged more negatively on most items and dimensions, but not on all scales and all dimensions (see Table 3.5). Moreover, the evaluation by the participants is more complex: they consider the implicit version to be less clear, less coherent and less professional, but it is also considered to be the simplest version.

3.3.3 Feeling of Knowing

For Feeling of Knowing, the explicit versions were expected to cause a higher FOK than the implicit ones. Significant effects only occurred for the informative versions. There is a main effect of coherence marking ($F(1,76) = 2.89$, $p<.05$, $\eta^2=.04$): the explicit text causes a higher FOK than the implicit text, confirming hypothesis 2b. Apparently, the explicit text gives a reader the impression of having understood the text better than the implicit text. There is also a main effect of prior knowledge ($F(1,76)=6.77$, $p<.01$, $\eta^2=.09$):

participants with prior knowledge think they have understood the text better than participants without prior knowledge.

It is also important to check whether the readers' impressions of their own understanding is correct or not. Linear regression analysis was used to calculate whether FOK could be a predictor for the actual comprehension scores. FOK does appear to be a reasonable predictor of the actual comprehension scores for the informative texts ($R^2=.31$, $p=.06$), but not for the persuasive texts ($p>.1$). Readers appear to have more trouble assessing their level of understanding in persuasive contexts.

3.3.4 Persuasion

Aside from appraisal and FOK, I argued that persuasion needs to be investigated when we want to know how a certain text characteristic influences the reader. In this experiment, persuasion was measured at the beliefs-, attitude- and intention-level. Reliability for the five persuasion statements (beliefs, attitude and intention) taken together is acceptable: α genetic manipulation = .76, α organ donation = .75 (van Wijk, 2000).

In the persuasive condition, there are no effects of coherence marking on persuasion, whether we look at the belief-, attitude-, or intention level. This could be an indication that coherence marking does not influence persuasiveness of a text. However, in the informative condition – texts that by definition do not aim at changing attitudes – there is an interaction effect of prior knowledge and marking of coherence at the attitude level. In Table 3.6, the results on attitude are represented. The lower the score, the more participants agreed with standpoints from the text.

	Informative texts		Persuasive texts	
	Implicit	Explicit	Implicit	Explicit
Less knowledge	5.00 (1.00)	2.71 (1.11)	4.50 (1.38)	3.83 (1.72)
More knowledge	4.46 (1.45)	5.08 (1.71)	4.14 (1.88)	4.07 (.44)

Table 3.6 *Effects of marking of coherence on attitude in informative and persuasive texts (mean scores and SD on a 7point scale, 1= agreeing with text, 7= disagreeing)*

The interaction effect of coherence marking and prior knowledge on persuasion that occurs in the informative texts ($F(1, 77)=9.37, p<.05, \eta^2=.21$) means that readers without prior knowledge agree more with the text after reading the explicit version than after reading the implicit version. There also are two main effects. The main effect of prior knowledge on persuasion ($F(1, 77)=3.71, p<.05, \eta^2=.09$) shows that participants with prior knowledge disagree more than participants without this knowledge. The main effect of coherence marking on persuasion ($F(1,77)= 3.11, p<.05, \eta^2=.08$) means that the implicit version causes more disagreement than the explicit version.

Still, it is remarkable that these effects occur in the informative texts, which do not have persuasion as their goal. Research by Petty and Cacioppo (1979) can give an indication for the reason for these seemingly unexpected results. Petty and Cacioppo asked participants to read exactly the same text, but half of them thought it was an informative text; the other half thought it was a persuasive one. If they were under the impression that they were not being 'influenced', they accepted the standpoints in the text without too much criticism. However, when they thought they were being manipulated, they were more critical to the content. This could have happened in the Pilot Experiment: the informative text could have been considered to be neutral. This could have caused the readers to accept the information as true, without questioning it.

Clearly, more research is needed to make sense of these unexpected persuasive effects of coherence marking. In the persuasive genre, there is no clear effect of coherence marking on persuasion. In the informative genre, the influence of coherence marking is clearer. I therefore interpret these results as indications for a possible persuasive effect that markers may have. Persuasive effects of coherence marking will be the focus of Chapters 6 and 7.

3.4 Conclusion and Discussion

In this Pilot Experiment, the influence of marking of coherence and prior knowledge was examined on four different dependent variables: comprehension, appraisal, Feeling of Knowing and persuasion.

3.4.1 Conclusion

For comprehension, the results of McNamara en Kintsch (1996) and McNamara et al. (1996) have been replicated with carefully manipulated text versions that only differed in the linguistic marking of coherence. There was an

interaction effect of prior knowledge and marking of coherence: readers without prior knowledge obtained a higher score on comprehension questions after reading the explicit version, for readers with prior knowledge, both versions resulted in comparable results. This last result, the fact that for high knowledge readers there was no difference in comprehension between the implicit and the explicit text differs from the result in McNamara and Kintsch's (1996) and McNamara et al.'s (1996) studies. They found that for high knowledge readers, the implicit text results in better understanding of the text. Furthermore, the interaction between prior knowledge and coherence marking was only found for the informative texts; in the persuasive texts, the explicit version leads to better understanding for both groups.

The effects on text appraisal for both the informative and the persuasive texts were as predicted: the implicit versions were judged more negatively on the four dimensions involved than the explicit version. For Feeling of Knowing, the coherence markers in the informative texts caused a higher Feeling of Knowing. FOK was also a good predictor of actual comprehension. In the persuasive texts, there were no effects.

Effects on persuasion, the third dependent variable, were counter-intuitively only found in the informative genre. Readers without prior knowledge are more convinced after reading the explicit version; readers with this knowledge are more persuaded after reading the implicit version. In the persuasive genre, no effects on persuasion were found.

3.4.2 Discussion

The results from the Pilot Experiment provide some very interesting starting points for further research. The experiment tested whether genre could be manipulated in such a way that a direct empirical comparison would be possible. This is definitely the case: we have seen that a specific text factor (namely coherence marking) and a specific reader characteristic (namely prior knowledge) seem to interact on comprehension in the informative genre, but not in the persuasive genre. The content was kept as constant as possible between both genres, only the context was varied. This direct comparison gives rise to further empirical investigation: why do these effects differ from one genre to another?

However, there are several caveats. First, there is a clear need for a more valid and reliable operationalization of text comprehension and situation

model representations. The comprehension questions in the pilot experiment turned out to be unreliable. Their internal reliability, calculated with Cronbach's alpha, was disappointingly low (around 0.4). In Chapters 4 and 5, this problem is tackled. The experiment on coherence marking effects on comprehension has to be repeated (Chapter 5) in order to draw more reliable conclusions.

Second, the results on persuasion were not as expected, because no effects at all occurred in the persuasive genre. Although this part of the study was explorative, we did not expect that the effects on persuasion would only occur in the informative genre. This result shows that the theory about persuasive effects of coherence marking is in urgent need of more theoretical and empirical work. In Chapters 6 and 7, I will return to this issue.

In sum, we can conclude from this Pilot Experiment that coherence marking indeed seems to influence text comprehension, text appraisal, FOK and persuasion. Therefore, it seems worthwhile to enlarge the view on mental representation, or more specifically the situation model representation, by including these factors. Clearly, the pilot experiment underlines the importance of including all three types of effects in order to advance the theory on coherence marking and reach a better understanding of their exact role and functioning.

4 How to measure situation model representations¹²

Although situation model representations are crucial when we want to investigate effects of coherence marking on text comprehension, there is absolutely no consensus on how to measure these representations. Therefore, this chapter focuses on the question: which situation model methodology is the most valid? In an experiment on the validity of situation model tasks among four hundred high school pupils, four different methods were compared (question asking, cloze tests, sorting tasks and mental model tasks). Correlations and reliability analyses show that the sorting task is the most valid method to use when assessing situation model representations.

¹² Parts of this chapter were also reported in earlier publications (Kamalski, 2004; Kamalski, Sanders, Lentz, & van den Bergh, 2005) and in Kamalski, Sanders, Lentz & van den Bergh (submitted).

In this dissertation, I investigate the effects of coherence marking on the situation model representation. This is the level of representation where prior knowledge and text information interact. Coherence marking should also interact with prior knowledge (see among many others McNamara, Kintsch, Songer, & Kintsch, 1996). Therefore, the situation model representation is the crucial level supposedly affected by coherence marking.

However, it is not clear what method should be used when assessing the situation model representation. There is absolutely no consensus in previous research on what method to use. For instance, in the particular context of research on coherence marking and the effect on the situation model representation, many different methodologies have been used in previous studies. In Chapter 2, these different methodologies were explained in more detail; for example, Boscolo and Mason (2003) used mental model tasks; Britton, Glynn, Meyer and Penland (1982) used a recall test; Degand, Lefevre and Bestgen (1999) used open inference questions; McNamara, Kintsch, Songer and Kintsch (1996) used, among other methods, sorting tasks. There are many more examples of studies that all aim at studying the influence of coherence marking on the situation model representation, many of which use different methodologies to study the same effect. It is hardly surprising that different conclusions are drawn in these studies. On the basis of the sorting task and the bridging inference questions, McNamara et al. conclude that coherence marking and prior knowledge interact in the situation model representation. On the basis of bridging inference questions and mental model tasks, Gilabert et al. conclude that there is no interaction between coherence marking and prior knowledge. Of course, these studies differ in more respects than the methodology. But a first step that needs to be taken in order to draw reliable conclusions concerning the effect of coherence marking on the situation model representation is to validate these methodologies. What exactly is being measured with a sorting task, or a mental model task, or inference questions? Do these methodologies measure the same construct? Do they measure what we intend to measure, namely the situation model representation? These questions need to be answered before we can analyze the effects of coherence marking on the situation model representation. Therefore, the focus in this chapter is on validity of situation model methodology. The central question in this chapter is (see also Chapter 2):

RQ 2: What methodology is most suitable when assessing situation model representations?

Researchers are not the only ones searching for a valid method to assess the situation model representation. Teachers and test developers also struggle with problems of validity, reliability and generalizability (Paris & Stahl, 2005; Pearson & Hamm, 2005). Hence, the central question is not only relevant in experimental contexts, but also in educational contexts.

Finding the most valid method to assess situation model representations can be a true challenge. Situation model representations are known to depend strongly on a specific text and a specific reader (see also Chapter 2). If these representations are formed with the use of prior knowledge, prior experiences, and so on, they will differ considerably from one reader to another. One might conclude that it is simply impossible to find a method to assess these dynamic representations in a stable way. In other words: how does one decide whether a situation model representation is 'correct' or 'wrong'? However, in educational contexts, there is a practical need for methods to also have some normative aspect: they must enable us to decide whether or not a reader has understood the information in the text. The aim of this chapter is to use situation model methodology in such a normative context. In sum, the challenge is: if a situation model representation really differs from one situation to another, what method is most suitable to measure this flexible representation?

4.1 The Validation Experiment 2

As promising as the new operationalizations of situation model representations may seem, the question remains as to whether they have been really validated empirically. Although we have seen in Chapter 2 how the foundations of situation model representations and their assessment have been laid by Schmalhofer and Glavanov (1986), Zwaan and Radvansky (1998) and others, it appears that the actual testing implications have never been validated. It will be beneficial to determine in a systematic way which method is the most valid one. Correlation experiments on validity provide a tool to answer such questions, because they investigate whether or not an observed score reflects the underlying theoretical construct that the investigator intended to measure (see for instance Cronbach & Meehl, 1955; Holleman, 2000).

4.1.1 The selection of four situation model tasks

Four different methods are selected to be tested in this validation experiment (see also 4.1). The sorting task and the mental model task are chosen because they seem to be promising methods to assess situation model representations, even though they were only developed recently. These methods were compared to more traditional ones: question asking and cloze tests. This choice was made because the traditional methods offer a ‘base-line’ to which the results of the newer methods can be compared. In Chapter 2, these methods (and others) are described in a very detailed way. Here, only a brief description of the four methods under investigation is given.

In educational contexts, reading comprehension is most commonly assessed with question asking (see for a recent overview of comprehension assessment methods, Pearson & Hamm, 2005). Many theoretical arguments exist in favor of and against their use, but in practice, they seem to be considered an effective method of assessing students’ learning from text (among many others Fellenz, 2004; Scouller & Prosser, 1994). The so-called cloze task (Taylor, 1953) is another common reading comprehension assessment: words are omitted from a passage and the reader is asked to fill in the blanks with the appropriate words. Although there is a strong criticism of the depth of the comprehension that is assessed with this method (see for instance Shanahan & Kamil, 1984), it still seems to be considered a practical method in education (Bachman, 2000). Sorting tasks and mental model tasks are categorization methods that rely on the reader’s ability to group words that were taken from the text. These methods were only recently applied to text comprehension and situation model representations.

4.1.2 Hypotheses

The goal of this experiment is to investigate whether the recently developed methods such as sorting tasks and mental model tasks are valid methods for measuring situation models. This is done by comparing these methods to more traditional ones.

The hypotheses are related to the different aspects of validity. The first step is to establish whether or not a method is internally reliable. A test is internally reliable if the separate questions or items measure the same construct. Internal reliability is a necessary, but not sufficient, condition for validity. Secondly, if different operationalizations are all considered to be situation model tasks, they

must correlate highly (convergent validity). This means that we expect question asking, sorting tasks and mental model tasks to correlate highly. A third aspect of validity is divergent validity: the operationalization does not measure other, related constructs. Three other (related) cognitive constructs are measured in the experiment: IQ, prior knowledge and attitude/opinion¹³. These factors are expected to play a role in situation model representations, but they should not be identical to them. The fourth and last aspect of this question is known-group validity: does a certain method have discriminative abilities between groups? The methods are expected to be able to discriminate between 10th grade pupils and 12th grade pupils: 12th graders perform better on average than 10th graders. As for the variance in both groups, I expect less influence of a specific text on comprehension results in grade 12 than in grade 10. In education, pupils are constantly trained in text comprehension and over the years, they also develop their reading skills. The general reading skills have become more stable and less dependent on a specific text with a specific text topic, structure and style. Therefore the influence of one specific text on text comprehension is expected to diminish. In other words, from grade 10 to 12 the variance between texts should decrease, but the variance between subjects should increase.

Hence, we can now present the following four hypotheses (QA= question asking, CL= cloze, SO= sorting task, MM= mental model, IQ, PK = prior knowledge and ATT = attitude). The 25% limit of cumulative explained variance is arbitrary, meaning that a considerable part, but not all, of the variance can be explained by these related constructs. We can formulate the hypotheses as follows:

H 0: Each method has to be internally reliable.

H 1: SO & MM correlate higher with QA (situation model) than with CL (text base)

H 2: a) 12th graders perform better than 10th graders, b) variance due to text decreases from grade 10 to 12,

H 3: The influence of IQ, PK and ATT respectively has to be equal for all methods. Their cumulative explained variance has to stay under 25%.

¹³ We have seen in Chapter 2 how readers use prior experiences, knowledge, feelings, opinions, etc, to build situation model representations. Therefore, it is expected that these factors are all related, but not the same construct.

4.1.3 Materials

The materials for this experiment are taken from Dutch high school exams constructed by Cito, the Dutch national examination centre. The Cito is a reputable institution with experience in testing skills and reading comprehension since 1987. It is in the top three of international organizations in test development and measuring learning performances. All texts in the validity experiment had been actually used as national exams in 2001 and in 2003. These exams presented students with several two-page texts, followed by multiple choice and open questions in order to determine their level of text understanding. These texts all expressed the writer's opinion about a certain phenomenon in society (the death penalty, the Dutch political system, robotics and environmental planning in The Netherlands).

IQ was assessed with a non-verbal test taken from Horn (1969). Verbal IQ-testing is difficult in a classroom context, where 25 or more students are being assessed simultaneously. Therefore, a non-verbal test was used (Rymenans, 2004). Prior knowledge was measured indirectly, by asking the participants to assess their own level of knowledge on a 5-point Likert-scale (Do you know a lot about topic X? Do you read about this topic?). A third question asked participants to name a few key words for each topic, which was a more direct way of assessing their prior knowledge. Whether or not the participants' opinion differed from the one expressed by the author of each text was measured on a 5-point Likert-scale with questions such as 'Do you agree with the author?' and by formulating standpoints based on the text: 'The Netherlands are too crowded' and asking participants whether they agreed or not.

4.1.4 Situation model task construction

The original questions from the Cito exams were mostly multiple choice questions, although some were open-ended¹⁴. In Appendix 2, a translated example is given of a text passage of the texts in this experiment on the lack of confidence of the Dutch people in Dutch politics and the corresponding question. The other tasks were constructed: a cloze test, a sorting task and a

¹⁴ Van den Bergh (1990) showed that there is no significant difference in reliability between multiple choice items and open-ended questions. Therefore, in the current study, these questions are analyzed together.

HOW TO MEASURE SITUATION MODEL REPRESENTATIONS - 89

mental model task. Only post-tests were performed for all methods, including sorting tasks.

In the construction of the cloze task, every 8th word was omitted from the original Dutch text. The first paragraph was kept intact, for the participant to establish what the topic of the text was, before having to fill in the blanks. Since the passage was translated, not exactly each 8th word appears as omitted. In the original Dutch text, the word in the exact 8th position was omitted.

For both SO and MM, approximately 20 key words and concepts were selected for each text. These concepts played an important role in the argumentational structure of the text. The choice of concepts was determined by the principle that they should enable the participant to reconstruct the rationale behind the text. With these 20 concepts, a model was constructed in which these concepts were linked in the same way as the text did (for example, causes were linked to their consequences, or examples of the same phenomenon were linked). The following example illustrates the sorting task construction.

In Chapter 1, a text passage about the Eiffel Tower was introduced and in Chapter 2 a corresponding sorting task was given. Here, this example is repeated in order to explain sorting task construction.

Text: The Eiffel Tower was erected for the **Paris Exposition** of 1889. Of the 700 proposals submitted in a design competition, one was unanimously chosen, a radical creation from the French structural engineer **Alexandre Gustave Eiffel**. However, the controversial tower elicited some strong reactions, for different reasons. The first one was that some people— including **Maupassant** and **Zola**— found the Eiffel Tower **useless and monstrous**. Second, nature lovers thought that it would interfere with the **flight of birds** over Paris. It was almost torn down in 1909, but was saved because of its **antenna** - used for telegraphy at that time. Still, the Tour Eiffel is now completely accepted by French citizens, and is internationally recognized as one of the symbols of Paris itself.

In order to construct the corresponding sorting task, the important rationale in the text needs to be identified. In this case, the opposition between criticism and positive aspects to the Eiffel Tower in 1889 is very important in the text. If readers miss this opposition, they have not understood the crucial information in the text. If this opposition is considered to be important, this implies that the terms that are in opposition need to be included in the sorting task. Also,

central terms to the text topic need to be included, such as Eiffel and Paris Exposition. For reasons of exposition, they are printed in bold in the example text. One important remark that needs to be made is that terms with a strong negative or positive direction, such as the adjective *monstrous* are neutralized when mentioned in the sorting task. In this case, monstrous is replaced by *style*, a neutral term. Otherwise, the nature of the word already indicates in which group the concept needs to be placed: negative words with other negative words and positive words with positive words. Then, the text would no longer be needed to fill in the sorting task.

In Appendix 2, a second example is given: part of a sorting task from the validity experiment. For the mental model task, the same concepts and key words were used, but they had been integrated in a schematic representation of the information in the text. This means that for each text, 3 methods were constructed and added to the existing QA questions. This resulted in 16 different tasks: four tasks for four texts. All materials and tasks were pre-tested.

4.1.5 Participants and procedure

Four hundred high school pupils from 4 different Dutch secondary schools participated in this experiment; half of them were in the 10th grade, the other half in the 12th grade. The average age of the pupils in the 10th grade was 15.2, for students in 12th grade 17.1. In total, 220 female pupils and 180 male students participated in this experiment.

Every pupil received a package with personal questions on the first page (name, age, school, level, and so forth). After the IQ test, the ‘comprehension section’ started: prior knowledge questions about the text topic, the text, the comprehension task and finally the attitude questions. This section was repeated 3 more times for different texts. Pupils completed the tests in their own classroom. The experiment took about two hours.

All tasks were performed in a post-text version. Participants were allowed to turn back the pages to check their answers in the text, because the answers to situation model questions cannot be found literally in the text. However, they were instructed to turn back as little as possible, because the available time for the experiment was limited. They were told that these results would give them a fair estimate of the result they would obtain if they were taking the exam right now, which would give them an indication of the work

that still needed to be done before the exam. With this instruction, we hoped to influence their motivation for participating in this experiment. Also, their schools and teachers made it clear that this was an obligatory test for them. The participants had 20 minutes to complete each text in combination with the questions. They were warned after 10 minutes that half of their time was up. Time pressure was added to the procedure in order to resemble the natural exam situation as much as possible. Even though participants may have differed in the amount of time spent on the text and the task, we controlled systematically that they could not spend longer than 20 minutes with the added time-pressure.

4.1.6 Design

Of course, it is not possible to have students perform all four different methods of reading comprehension for just one text. In that case, the answers for one method would influence the scores on another. Therefore, we opted for a design in which all four types of methods were asked for all texts. This results in a design in which all students take only one method for one text, and another method for another text. For instance, the first student answered the traditional comprehension questions of text 1, a sorting task for text 2 and a mental model task with text 3. The second student started with the mental the model task of text 1, the sorting task of text 2 and the question asking test of text 3. In addition to these three tests, one of the methods was repeated within students. So, the first student would take another mental model test for text 4, and the second student would take another question asking test for the 4th text. In essence, this is an experimental design for correlational data. In this design, texts are crossed with methods. In Table 4.1 the allocation of students to methods (type of reading comprehension questions) and texts is presented.

Method	Text1	Text 2	Text 3	Text 4	Total
Question Asking	74	73	72	73	292
Cloze	71	73	70	73	287
Sorting task	124	121	117	122	484
Mental model	119	118	125	123	485
Total	388	385	384	391	

Table 4.1 *Number of students per text and per type of question*

From Table 4.1 it appears that each text is read as often as the other texts (see total; horizontal entry). It is also easy to see that the two situation model methods (sorting task and mental model) are somewhat overrepresented; for both traditional methods the number of observations is about 60% of the number of observations for all experimental methods.

4.1.7 Scoring

Scoring of the QA questions was done exactly in accordance with the scoring-table provided by the Cito: wrong answers were always scored as 0, correct answers sometimes 1, 2, or 3, depending on the weight attributed to the question.

In the CL task, only exactly the same answers were considered to be correct, no synonyms or misspelled words were accepted. The correct answers were attributed 1, the wrong answers 0. Afterwards, the total score was calculated, the maximum being 55.

In the SO, the participants' scoring was compared to the analysis that several experts made of the texts and the way the key concepts were linked. Several categorizations were considered to be correct: for instance causes with causes, consequences with consequences, or each cause with its consequence. When a participant placed a key concept in a wrong category, his or her maximum score was lowered by 1 point. For instance, a participant who misplaced 4 concepts out of the total 20 would receive a score of 16 points.

In the MM, the participants saw a schematic diagram in which every empty text box had to be filled with a key concept. With every box that did not contain the right word, the score lowered by 1 point. Again, a participant who placed 4 concepts in the wrong boxes would receive a score of 16 points.

4.2 Results

A multi-level analysis (Quené & van den Bergh, 2004) was performed on the data, in order to investigate effects between subjects and between texts. First, the method of analysis is outlined. Then, the characteristics of the variables and the internal reliability of the operationalizations are presented. Next, I will discuss and calculate the different aspects of validity: convergent validity, known group validity and finally divergent validity.

4.2.1 Method of analysis

In order to analyze these data a multi-level model is specified. Let $Y_{(ij)}$ be the score on text i ($i = 1, 2, \dots, I$) of participant j ($j = 1, 2, \dots, J$). Each score is therefore nested within the combination of text and participant. That is, data are collected according a cross-classified design (Goldstein, 2003). In order to analyze these data a multilevel model is specified. If I define four dummy-variables, one for each method ($QA_{(ij)}$, $CL_{(ij)}$, $ST_{(ij)}$ and $MM_{(ij)}$), which are turned on -equals 1- if the test score represents the respective type of measurement and turned off -equals 0- otherwise), the model can be written as:
 $Y_{(ij)} = \beta_1 * QA_{(ij)} + \beta_2 * CL_{(ij)} + \beta_3 * ST_{(ij)} + \beta_4 * MM_{(ij)}$.

For each method a mean score ($\beta_1 - \beta_4$) is estimated. However, in this model, random components are not yet specified. Random components concern: the text, the individual and the residual score. The four (mean) regression weights can be written as:

$$2. \beta_1 = \beta_{1(ij)} + e_{1(ij)} + u_{1i0} + v_{0j},$$

...

$$B_4 = \beta_{4(ij)} + e_{4(ij)} + u_{4i0} + v_{0j}.$$

In the equation above $\beta_{1(ij)} - \beta_{4(ij)}$ denote the mean score of individual i on all the texts with method j . For each method the residual score of individual j is estimated ($u_{1i0} - u_{4i0}$), and one residual score for texts is estimated (v_{0j}). Hence, it is assumed that the differences between texts are equal across types of method. The third residual score ($e_{1(ij)} - e_{4(ij)}$) can be defined as a combination of the interaction between individual and text, as well as random noise. It is assumed that all residuals are normally distributed with an expected score of 0 and a variance of $S^2_{e1(ij)} - S^2_{e4(ij)}$, $S^2_{u1i} - S^2_{u4i}$, S^2_{v0j} respectively.

If equation (2) is substituted in (1) one gets the complete model with fixed (Equation 1) and random effects (Equation 2). This model can be easily extended with either fixed effects (for instance prior knowledge) or random effects.

4.2.2 Descriptives and reliability of the operationalizations

In Table 4.2, the mean scores on the four methods are presented in terms of percentage of correctly answered questions. The variance per method is divided into three categories: variance between texts, variance between subjects, and residual variance.

Over all methods, the variance between texts is 24.61, the standard deviation 4.85. This specific type of variance provides us with a criterion for the interpretation of the mean scores in Table 4.2: if the mean scores per *method* differ more than one standard deviation of variance caused by the *text*, this is a strong indication that the difference between methods is an important one. Table 4.2 shows that all methods differ more from the other methods than 4.85, except the sorting task and the mental model. This means that the performance of a pupil is influenced more by the method than by the text that is tested. The scores on MM and on SO do not differ more than 4.85 and this difference could therefore be due to text-factors instead of method-related factors. We also see in Table 4.2 that for question asking, all the variance in the data is residual variance and that no variance can be attributed to the interaction between text and individual (see also Table 4.4).

Based on these estimates we can approximate the internal reliability for the situation model tasks. Table 4.3 shows the scores on Cronbach's alpha per text per method. The internal reliability for the questions was low. The reliability of the cloze test was good, the reliability of the sorting task was acceptable, and the reliability of the mental model task was low, except for the second text.

Method	Mean (% correct answers)	S ² residual	S ² between subjects
Question asking	31.99 (2.63)	316.5 (26.7)	0.00 (0.00)
Cloze	44.48 (2.57)	156.8 (21.8)	41.8 (2.3)
Sorting task	67.78 (2.58)	178.8 (24.2)	144.9 (28.5)
Mental model	70.60 (2.61)	280.8 (36.3)	119.6 (37.0)

Table 4.2 *Descriptives (mean and S²) in percentages of correct answers (n=400), standard error between brackets*

However, Cronbach's alpha is known to be an underestimation (Lord & Novik, 1968). A possibly more adequate measure of reliability is split-half reliability, where we randomly divide all items that purport to measure the same construct into two sets. This is not possible for the sorting task and the mental model task, because these tasks consisted of three items. The split half reliability scores for QA and CL are added to Table 4.3. Although they are slightly higher than the Cronbach alpha's, the conclusions still remain the same.

HOW TO MEASURE SITUATION MODEL REPRESENTATIONS - 95

Method	Text 1		Text 2		Text 3		Text 4	
	α	SH	A	SH	A	SH	α	SH
QA	.23	.26	.42	.50	.30	.55	.53	.60
CL	.83	.88	.82	.89	.80	.88	.79	.91
SO*	.61	-	.65	-	.67	-	.79	-
MM*	.50	-	.94	-	.50	-	.45	-

* performed on a subset n=60

Table 4.3 *Internal reliability in Cronbach's alpha and Split Half (SH)*

The split-half method still underestimates the reliability, but these scores come closer to the real reliability. Split half and Cronbach's alpha have in common that they both calculate reliability *per* text. Another way of analyzing reliability is looking at the proportion of the variance that can be attributed to the participant and not to other factors (see Table 4.4), calculating reliability *over* texts. This reliability was calculated as follows: $\rho = S^2_{\text{subject}} / (S^2_{\text{subject}} + S^2_{\text{residual}})$. Ideally, this proportion has to be large: a method is then optimally capable of distinguishing between subjects.

Method	Proportion of variance attributed to the subject	Number of tasks needed for .80 reliability ¹⁵
QA	0.00	-
CL	0.21	15
SO	0.44	5
MM	0.30	9

Table 4.4 *Proportion of variance attributed to the subject and number of tasks needed for .80 reliability*

These results show that in order to obtain a reliable result for a specific pupil, an endless number of tasks are needed. For the cloze task, 15 tasks have to be completed by a pupil in order to attain a reliability of .80. For the sorting task, only 5 tasks are needed for a .80-reliability, and for mental model tasks 9 tasks are needed. In Table 4.4, the sorting task has the highest variance attributed to

¹⁵ Calculated with the Spearman Brown prediction formula, also known as Spearman Brown prophecy formula (for instance in Lord & Novik, 1968).

the subject, therefore the least variance attributed to the text. If we only look at the proportion of variance attributed to the subject, the sorting task is the best option to assess situation model representations.

4.2.3 Convergent validity

In order to investigate whether these methods measure the same construct, the correlations between all four methods were calculated (see Table 4.5).

Method	QA	CL	SO	MM
QA	-	-	-	-
CL	-	1		
SO	-	0.46	1	
MM	-	1	0.50	1

Table 4.5 *Correlation coefficients among 4 text comprehension methods*

Correlations between question asking tasks and other methods cannot be calculated, because the variation attributed to the subject equals 0. Correlations between all other methods are positive, as is expected. The correlation between the mental model and the cloze task is very high; the correlation between the sorting task and the cloze task is also considerably high. This is contrary to the expectations: the cloze task was expected to measure text base representations, and the mental model and the sorting task were expected to measure situation models (see also the Discussion). Table 4.5 indicates that the cloze task, the sorting task and the mental model task all measure the same mental construct. This could be an indication that they are possible methods to assess situation model representations. In paragraph 4.3.6, this surprising result is discussed and reanalyzed.

4.2.4 Known group validity

One of the criteria for validity is discriminative ability. Two groups are chosen in an experiment, because they are known to differ. In this particular experiment, the two groups were 10th grade pupils and 12th grade pupils. The 12th grade pupils were expected to perform better on all tasks. In the following Table, we compare regression weights by means of Chi-square with 1 degree of freedom.

HOW TO MEASURE SITUATION MODEL REPRESENTATIONS - 97

Method	Mean 10 th grade	Mean 12 th grade	X ² (df=1)
QA	29.55 (2.76)	34.11 (2.83)	5.06*
CL	41.13 (2.61)	48.42 (2.70)	20.12*
SO	64.43 (2.63)	71.55 (2.74)	17.10*
MM	65.13 (2.75)	76.81 (2.66)	42.76*

* p<0.05

Table 4.6 *Comparison of mean scores in percentage of correct answers for both school levels (10 and 12), standard error between brackets*

Table 4.6 shows that this expectation was confirmed: all tasks were able to discriminate effectively between the two school levels. On all tasks, 12th graders performed on average better than 10th graders. On the basis of Table 4.6, there is no preference for one single method.

Another aspect of known group validity is the variance. In Table 4.7, both the variance between subjects and the residual variance is calculated separately for grades 10 and 12.

Method	S ² residu		S ² subject		Proportion of subject variance on total variance	
	Grade 10	Grade 12	Grade 10	Grade 12	Grade 10	Grade 12
QA	294.6	334.9	0	0	0	0
CL	113.7	212.2	40.46	7.8	0.26	0.03
SO	159.5	203.1	115.6	148	0.42	0.42
MM	282.1	266	144.4	28.4	0.34	0.10

Table 4.7 *Comparison of variances for both school levels (grade 10 and 12)*

On the basis of the variance attributed to the subject, question asking is completely unreliable for both grades. No variance can be attributed to the subject. The same problem occurs with the cloze task, but only for grade 12, where only 3% of the variance can be attributed to the subject.

Results on the sorting task and the mental model task depend on both the subject and on other factors. However, when we look at the proportion of

variance attributed to the subject, we see that for the sorting task this variance is constant from grade 10 to grade 12. For the mental model task, variance due to the subject decreases from grade 10 to 12, from 34% to 10%. Ideally, a pupil should perform more constantly in the 12th grade than in the 10th grade. This is not the case for either of the methods, but the fact that for mental model tasks, variance attributed to the subject decreases is an indication that the sorting task would be the more successful method.

We have seen in Table 4.6 and 4.7 that the sorting task is not only able to discriminate between levels in school, but that the variance due to the subject is also constant over time. The sorting task is the only method that combines these two characteristics.

4.2.5 *Divergent validity*

Related constructs were also measured (IQ, Prior Knowledge and Attitude toward the text topic). They are expected to influence text comprehension, but are not the constructs that I aim to measure: situation model representations. Internal reliability for these operationalizations was good: the IQ test proved to be reliable ($\alpha=0.84$). The reliability of the prior knowledge questions was acceptable (for text 1: $\alpha=0.71$, 2: $\alpha=0.61$, 3: $\alpha=0.64$, 4: $\alpha=0.64$). For the attitude questions, reliability was good except for the fourth text (1: $\alpha=0.83$, 2: $\alpha=0.82$, 3: $\alpha=0.79$, 4: $\alpha=0.50$).

Two models were fit onto the data for each related construct: one in which the influence was estimated freely, and a second one in which the influence was constrained to be equal over methods. If there was no significant difference in fitting the data, the second model was preferred. Indeed, the second model did not cause a significant change in fitting the data for the variable IQ ($\chi^2(6)= 5.1611$, $p= 0.52$). The influences of IQ on test scores are therefore constrained to be equal. The model fitting did cause a significant difference for the other two variables (for PK: $\chi^2(6) =13.903$, $p= 0.03$, for attitude: $\chi^2(6) =15.929$, $p= 0.01$). In Table 4.8, the regression weights of these three constructs for each situation model task are represented. Regression weights show per unit in these three constructs the amount of change that is caused in text comprehension. For instance, per unit change in IQ, the score in comprehension in the 12th grade will be 0.56 points higher (second column). For one unit more in prior knowledge, the score on comprehension in the 12th grade will go up with 0.48 points.

HOW TO MEASURE SITUATION MODEL REPRESENTATIONS - 99

Method	IQ		PK		Attitude	
	10	12	10	12	10	12
QA	0.13	0.56	0.49	0.48	-0.57	-1.9
CL	0.13	0.56	0.97	0.09	-0.80	0.29
SO	0.13	0.56	2.21	0.75	0.74	-0.70
MM	0.13	0.56	1.39	0.45	0.01	0.07

Table 4.8 *Regression weights for related constructs IQ, Prior Knowledge, Attitude and the situation model tasks*

The first conclusion that can be drawn from the results presented in Table 4.8 is that the influence of IQ is equal for all methods. Therefore, this does not provide a preference for a specific method. There is a larger influence of IQ in the 12th grade than in the 10th grade.

The variable prior knowledge also influences the scores on all methods, but we see in Table 4.8 that the influences on the sorting task and the mental model task are more important than the influences on question asking and cloze. At the situation model level, prior knowledge should have an influence on text comprehension. Therefore, this could be a reason to prefer SO and MM. The influences of attitude on text comprehension methods vary, but they are all small influences and do not provide a preference for one method over another.

Together, these variables do not account for *all* the variance in the comprehension scores. The total model, including the three related variables, accounts for 45% of the variance in the test scores. The model without these related factors accounts for 39 % of the variance in the test scores. Only 6 % of the variance in test scores can be explained by IQ, prior knowledge and attitude, and therefore I am clearly measuring another, separate construct. The 6 % explained variance by the related variables does not exceed the 25%-limit from the original hypothesis.

4.2.6 *Post-hoc analysis of the cloze tasks*

The results on convergent validity showed that mental model, sorting and cloze tasks correlated highly. This finding could indicate that all these three tasks actually measure comprehension at the text base or surface code level, not at the situation model level. After all, it is most likely that the cloze tasks measures

text base or surface code representations. However, an alternative explanation could be that the cloze task measures both text base and situation model representations, and that this high correlation indicates that we are measuring at both the text base and the situation model level. More insight into the exact functioning of the cloze tasks is needed in order to test this explanation. With this aim, a post-hoc analysis was conducted, following the method of Levenston, Nir and Blum-Kulka (1984). With this method, we try to find out what level of representation is used to fill in the gaps in a cloze task. The random gaps in a cloze task can be categorized into three categories of components of knowledge that a reader uses in order to find the right word.

In the first category of gaps in a cloze task, readers have to use linguistic knowledge. The context within the sentence provides enough information for the reader to fill in the gap. An example for this category, taken from the validation experiment, has been translated from the Dutch materials. The text topic is *Problems in The Netherlands due to the high population density*.

3. There is no other country that has _____ kilometers of highway and as many cars as ours. [Dutch word to be filled in: 'zoveel', English: 'as many']

In this example, the reader could already guess what word to fill in on the basis of the grammatical structure of the second part of the sentence. Moreover, 'as many cars' provides a clear indication of the word to be filled in.

The second category of gaps in a cloze task is based on pragmatic components. Readers need extra-textual information to fill in these gaps. Knowledge about the world in general provides the basis for these answers.

4. The Dutch Interdisciplinary _____ Institute (DIDI) states on its website '...'. [Demographic]

The topic of the paragraph is immigration, growth of population, and so forth. We also know that the word has to start with the letter d. When readers combine these textual clues with their knowledge of the world, they should come up with 'demographic'.

The third category of gaps in a cloze task concerns a textual component: readers use their ability to link sentences by following the structure of a text.

5. The growth of the population will diminish in the 21st century. Problem _____, one might think. [solved]

This gap requires the integration of information we find in several passages in the text. The problems are explained, then the prognosis for the future is given, and then, the conclusion can only be: problem solved.

In terms of text base and situation model representations, the first category (linguistic) is likely to pertain to text base and surface code representations, whereas the second and the third (textual and pragmatic) would include situation model representations. In the following Table 4.9, the number of gaps in each text of the experiment is calculated for each category.

In each cloze task, about 70% of the gaps could be filled in by only using surface code and text base representations. About 30% of the gaps could only be filled with the help of situation model representations. In conclusion, cloze tasks *can* measure situation models, but they also measure text base and surface code representations, and the ones we used predominantly did the latter.

	Text 1	Text 2	Text 3	Text 4
Linguistic	37 (67%)	37 (67%)	38 (69%)	42 (76%)
Pragmatic	7 (13%)	6 (11%)	9 (16%)	2 (4%)
Textual	11 (20%)	12 (22%)	8 (15%)	11 (20%)
Total:	55 (100%)	55 (100%)	55 (100%)	55 (100%)

Table 4.9 *Exact nature of each gap in the cloze task: number of gaps in each category (linguistic, pragmatic or textual)*

Whether or not this categorization of cloze items, following Levenston, Nir and Blum-Kulka (1984), provides an explanation for the data from the validation experiment was checked by means of factor analysis. In the factor analysis, for both the first and the second text 7 factors were extracted, for the third text 6 factors and for the fourth text 8 factors were found. In the following Table, the underlying statistical factors in the data are compared with the categorization from Table 4.9. In Table 4.10 we see what percentage of linguistic, pragmatic or textual items rely on each of these factors. These percentages do not necessarily add up to a 100 %, because one item or gap can rely on more than one factor. For instance, for text 1, 65% of all linguistic items

from that text fall into factor 1, 14% of all pragmatic items rely on factor 1, and so forth. The percentages in bold show what category relies most on this factor.

Factor:	1	2	3	4
	Lg- pr- t	Lg- pr -t	Lg- pr- t	Lg- pr- t
Text 1	65- 14- 20	12- 43- 58	15- 71- 16	15 -0- 8
Text 2	55- 60- 46	42-100- 77	27- 20- 46	16- 20- 15
Text 3	29- 89- 89	44- 11- 0	32- 55- 11	18- 0- 0
Text 4	61- 0- 50	31-100-50	19- 50- 17	20- 0- 17

Table 4.10 *The percentage of items from a specific category (linguistic – lg-, pragmatic –pr- or textual –t) per factor (1 to 4) per text (1 to 4)*

Table 4.10 shows indications for the existence of all three categories of items. The clearest indication for this is to be found in text 1; we clearly see that linguistic items rely on factor 1, textual items on 2, and pragmatic items on 3. In other texts, the distinction is not always as clear, but it is still present. However, even for text 1, we see that factors are rarely based on one type of item only: in almost all the factors, all three types are present. It is not surprising that I cannot separate these three types of items completely in the factor-analysis: surface code and text base representations are used in order to construct the situation model representation, so all three levels interact. However, the factor analysis still shows that the explanation offered by Levenston et. al., that all cloze items do not measure the same level of text comprehension and text representation, seems a valid explanation. This could very well be the reason why the correlation between cloze, mental model and sorting task is so surprisingly high.

4.3 Conclusion and Discussion

I will now evaluate each method from the validation experiment separately, on all the criteria that I have formulated earlier in this chapter.

4.3.1 Question asking

The internal reliability scores of question asking seem to indicate that more than one concept is being measured. There is no convergent validity with the other comprehension methods that I tested in this validation experiment. Divergent validity does not cause any problems. Concerning the known-group

HOW TO MEASURE SITUATION MODEL REPRESENTATIONS - 103

validity, I can say that, although the questions can discriminate between the two school levels, all variance for 12th grade pupils has to be attributed to the text, not to the individual. This means that an endless number of tests would be necessary before one can reliably establish a pupil's comprehension level, which is an impossible task. In conclusion, question asking seems unreliable and cannot be considered a valid way to assess either situation model representations or text comprehension in general.

4.3.2 Cloze task

For the cloze tasks, the internal reliability is good. A negative result is that the variance attributed to the subject is lower than for the sorting task and the mental model. Convergent validity is established between the cloze task and both sorting tasks and mental model tasks. Known-group validity is acceptable, but the variance between subjects is almost completely absent in grade 12, making results on this method almost completely dependent on other factors. The conclusion for cloze tasks is that there are indications that they partly measure situation model representations, but that this is not the dominant representation that is measured. The post-hoc analysis confirms this conclusion. The randomly constructed cloze tasks relied for 70% on text base and surface code representations, making cloze tasks less valid tasks for situation model assessment.

4.3.3 Sorting task

The internal reliability is acceptable and the variance between subjects is good (see table 4, where the score for sorting tasks is actually the best). Convergent validity is established between sorting tasks and both cloze tasks and mental model tasks. Known group validity is good, and contrary to other tasks, the variance between subjects stays constant from grade 10 to grade 12. Divergent validity poses no problems. In conclusion, sorting tasks seem to be a reliable and valid method to measure situation model representations.

4.3.4 Mental Model task

The internal reliability is poor, but still better than for question asking. The variance between subjects is not as good as for the sorting tasks, but is acceptable. The correlations between methods seem to indicate that mental model tasks measure the same construct as cloze tasks and sorting tasks.

Known group validity is acceptable, except that the variance between subjects decreases from grade 10 to 12. In my view, a reliable method to assess situation model representations should yield more constant results. Divergent validity causes no problems. In conclusion, there are some objections to the mental model task as a method for measuring situation model representations, but it can be an acceptable second choice.

4.4 Discussion

In this dissertation, I have argued that in order to develop valid methods for text comprehension assessment, it is important to address the level of situation model representation, since this is a crucial level of text representation. In this chapter, I have investigated to what extent four different text comprehension methods indeed tapped this level of representation. In this experiment, the sorting task proved to be the most valid method to measure the situation model representation that readers construct on the basis of the information in the text in combination with their own prior knowledge. The mental model task was the runner-up.

What do these results tell us about the situation model theory? The theory on situation models predicts that situation model representations possibly vary per situation, depending on the text and on the available prior knowledge. This could lead to the conclusion that there is no possible instrument or methodology to measure these complex representations in a constant way. Text comprehension certainly depends on a specific text and the specific topic, but it also depends on a skill that is constant within the reader in a repeated measures design. Over time, the skill of text comprehension can of course develop, but in such a short experimental period, the results should remain stable. In the validation experiment reported here, I have shown that situation model representations *can* be measured in a valid and reliable way, even with different methods. Two methods are successful in measuring situation model representations and the results from these methods highly correlate with one-another. Even if situation models can vary per text, there is a constant factor in play.

The sorting task turned out to be the most valid method in this experiment. The sorting task was first used to assess situation model representations by McNamara and Kintsch (1996) and McNamara, Kintsch, Songer and Kintsch (1996). However, there are many possible practical

realizations of a sorting task; in the current experiment, I only investigated one of them. Therefore, it is impossible to conclude that *any* sorting task would be the best method to use. One of the decisions to make in constructing a sorting task is whether to provide the participant with categories in which to categorize the keywords, or whether to leave it completely up to the subject. I followed McNamara and Kintsch's and McNamara et al.'s example and did not provide ready-made categories. I do not know whether doing this would affect the validity of the method. A second difference has to do with the scoring technique used. McNamara and Kintsch's scoring matrix-technique was not applied in this validation experiment because we wanted to investigate whether simpler solutions could be used. Therefore, our scoring method was simple: for each wrongly categorized term, the total score was lowered by one point. This scoring technique makes the sorting task more usable in different contexts. A third difference with the McNamara and Kintsch studies is that the experimental texts differed greatly. The texts in the current experiment were argumentative whereas the texts used by McNamara and Kintsch were informative schoolbook texts. Still, in both cases, the crucial criterion for constructing the sorting tasks was that the keywords had to be open to multiple categorizations on the basis of common sense. In other words, the answer could not be so obvious that reading the text would become unnecessary. Although many aspects of the practical realization of the sorting task may differ, this one aspect is constant and crucial for sorting tasks.

Apart from expected results, there were also unexpected outcomes of this experiment, such as the results on validity from the cloze task. I expected the cloze task to tap into the text base or maybe even the surface code representations, but I found some evidence in the data for influences of situation model representations. This validation experiment and the post-hoc analysis show that cloze tasks can certainly be useful and valid. A non-random cloze task could measure the situation model representation even more precisely, if only items with a pragmatic or textual component are included in the task (Boogaard, Sanders, & Admiraal, 1996).

In conclusion, this experiment on the validity of situation model tasks has shown that the traditional method of question asking seems to be the least valid method. On the contrary, the task that has only recently been used in text comprehension research, the sorting task, appears to be the most valid and reliable method for assessing situation model representations. The results also

show that cloze tasks and especially mental model tasks should be taken into account when looking for adequate measures of comprehension.

The validation experiment was conducted with four different argumentative texts. Further research is needed in order to investigate generalizability over texts and text types. However, with this validation experiment, it is shown that it is imperative to use this sorting method if we are looking for a valid and reliable way to assess situation model representations.

In Chapter 3, the effects of coherence marking on text comprehension were investigated with open questions. In this chapter, these effects are re-tested, this time making use of a sorting task. The central questions remain: Does the effect of coherence marking on comprehension depend on readers' prior knowledge of the text content? What is the effect of coherence marking on appraisal and Feeling of Knowing? Do these effects differ between the persuasive and the informative genres? Effects on persuasion are not considered in this chapter, but will be the sole focus in Chapters 6 and 7.

¹⁶ The research in this chapter has also been reported in two papers, co-authored by Leo Lentz and Ted Sanders (Kamalski, Lentz, & Sanders, 2006) and (Kamalski, Lentz, & Sanders, submitted).

In 2.1.3, I presented an overview of studies on the effect of coherence marking on the situation model representation. It is difficult to distill one clear picture from all these different, sometimes even conflicting, results on comprehension. There are three possible explanations for these differences between studies and their resulting conclusions.

The first possible reason for differences between the studies is the methodology that is used to assess situation model representations. In studies on coherence marking, the methodology varies from traditional methods such as multiple choice questions, open ended questions and recall tasks to more recently developed tasks such as sorting tasks and mental model tasks. This inconsistency makes it difficult to compare the experimental results. This problem was addressed in Chapter 4 and the results indicate that the sorting task is the most reliable method to use.

A second source of variation concerns the way in which coherence markers are varied across several experiments (Degand & Sanders, 2002). Even in the seminal work of McNamara and Kintsch, many different textual aspects were varied at the same time: adding elaborative information, identification of anaphoric references, and even supplying background information (for instance in the implicit version, the text mentioned *ARVN* and in the explicit version, this was explained as *South Vietnamese Army*). I believe it is of great importance to investigate the interaction of coherence marking and prior knowledge found by McNamara and Kintsch, but to do so with experimental texts that differ only in linguistic markers of coherence: in the explicit versions coherence is marked by connectives and cue phrases, whereas in the implicit version it is not, but the coherence relations in both text versions are identical.

A third reason why the studies on coherence marking differ concerns the operationalization of prior knowledge. This explanation for different results is discussed in the next section.

5.1 Operationalization of prior knowledge

There are four different operationalizations of prior knowledge. In some studies, prior knowledge is not manipulated but *assessed* (Boscolo & Mason, 2003; Langer, 1984; McNamara, 2001; McNamara, Kintsch, Songer, & Kintsch, 1996; Voss & Ney Silfies, 1996). On the basis of the results on the prior

knowledge assessment, participants are then divided into groups (for instance low prior knowledge and high prior knowledge).

Other studies made use of a so-called *expert-novice paradigm*. Two groups of participants are selected that are expected to differ with respect to their level of prior knowledge. An example is the experiment that Birkmire (1985) conducted. She selected physics and music students and these students read a text on laser technique. The physics students were expected to have more prior knowledge about the text topic laser technique than the music students.

Under the *instruction paradigm*, researchers provide half of the participants with the necessary prior knowledge; the other half do not receive such training. Thus, the experimenters create within the experiment a group with and a group without prior knowledge. The specific ways of instructing the participants vary from an instruction booklet (McNamara & Kintsch, 1996) to highly interactive lessons taking 30 to 60 minutes (Gilbert, Martínez, & Vidal-Abarca, 2005; McKeown, Beck, Sinatra, & Loxterman, 1992). McKeown et al. (1992: 38) explain the content of their lesson as follows: '(the lesson) did not present the sequence of events described in the text, but introduced major agents and concepts needed to construct a representation of that sequence from the text materials'.

Finally, a last option is to *vary the materials*. Prior knowledge is in this case manipulated in an indirect way: the experimental materials are varied (Blanc & Tapiero, 2001). When experimenters present a text with familiar content, all participants are part of the high prior knowledge group. By choosing a text that is highly unfamiliar, readers now do no longer possess the necessary prior knowledge and become the low prior knowledge group for this text.

In sum, there are several ways to manipulate prior knowledge. In this chapter, I will explore several experimental methods to distinguish novices from experts. Although the main goal is to clarify the relationship between prior knowledge and coherence marking and the effects they may have on several dependent variables, a second goal is to compare the validity and reliability of operationalizations of prior knowledge.

5.2 Experiment 3: comprehension, appraisal, FOK

This experiment is a replication of the Pilot Experiment (Chapter 3), with a few adaptations. First, the method to assess comprehension has been changed to a

sorting task, due to low internal reliability of the bridging inference questions (see chapter 4). This is a problem that often occurs when using inference questions to assess text comprehension (see for instance McNamara, Kintsch, Songer, & Kintsch, 1996, Experiment 1). In Experiment 4 in Chapter 4, I found that the sorting task was the most valid operationalization of text comprehension (Kamalski, Sanders, Lentz, & van den Bergh, submitted). This is why a sorting task was used to replace the open-ended questions.

A second adaptation is that in the current experiment, only marking of relational coherence was manipulated, and not marking of referential coherence, as was the case in the Pilot study. In doing so, marking of coherence was manipulated in a more homogeneous way than in the Pilot Experiment. We still manipulated marking of global coherence relations (headings and organizers) as well as marking of local coherence relations (connectives and lexical cue phrases) in the same way as we did in the Pilot Experiment. Table 5.1 presents a reminder of all types of coherence markers that were manipulated in the current study, and examples of each type.

Coherence marker	Example from explicit text	Example from implicit text
Headings	What are the consequences of genetic manipulation?	No heading
Organizers	There are two reasons for using genetic manipulation in food.	No organizer
Connectives	Environmental organizations say that genetic manipulation is dangerous, <u>because</u> it entails risks that the consumer is not aware of.	Environmental organizations say that genetic manipulation is dangerous. It entails risks that the consumer is not aware of.
Lexical cue phrases	<u>The goal of this manipulation is that</u> the plant creates a poison which will kill a certain plague of insects.	The plant creates a poison which will kill a certain plague of insects

Table 5.1 *Examples of coherence markers, taken from genetic manipulation text*

5.2.1 Hypotheses

The hypotheses in the current experiment are the same as in the Pilot Experiment (see Chapter 3). Here, I simply repeat them.

H 1: Readers with prior knowledge perform better on comprehension questions after reading an implicit version. Readers without prior knowledge perform better on comprehension questions after having read an explicit version.

H2: In comparison to the explicit version, the implicit version is judged more negatively. When coherence is left implicit, reading demands more cognitive energy from a reader. Readers find this to be annoying.

H3: The explicit versions cause a greater Feeling of Knowing than the implicit versions.

5.2.2 Materials

The same texts were used as in the Pilot Experiment, but with some slight adaptations. No effects of coherence marking were observed in the persuasive genre, but they were present in the informative genre. An explanation for this absence of an effect could be that the persuasive texts were ‘too’ persuasive, that is, so many persuasive features were used that the coherence markers did not have an effect on persuasion any more. Readers seemed to develop a resistance against these features. In other words, the coherence markers seemed to be ‘buried’ under a large pile of other persuasive features. Although this risk has been explicitly discussed in 2.4.7, the right balance was apparently not found. Therefore, the experimental texts were adapted in order to make them less overtly persuasive, which means that some persuasive clues as for the intent of the writer were removed. More specifically, the titles were adapted (*You too can save lives* in the Pilot Experiment vs. *Organ donation, a matter of choice* in the current Experiment), passages with emotional appeals were neutralized, etc. This was done carefully in order to preserve the persuasive nature of the texts. The checklist with prototypical features in persuasive text from Table 2.2 functioned as a criterion. In Appendix 3, two exemplary passages are represented.

On the basis of the results from Chapter 4, a different measure of text comprehension was used: a sorting task (see also Chapter 4). Participants were asked to categorize key concepts from a text (McNamara & Kintsch, 1996; McNamara et al., 1996). The sorting tasks resemble McNamara and Kintsch’s, but they are scored in a different way. The sorting tasks were constructed as follows: key words and concepts were selected by three text linguistic experts

for each text (approximately 20 words per text). These concepts played an important role in the argumentation structure of the text. The choice of concepts was determined by the principle that they should enable the participant to reconstruct the rationale behind the text. The concepts can be considered to reflect the macro-proposition of the text (van Dijk & Kintsch, 1983). With these 20 concepts, the experts constructed a schematic model in which the concepts were linked in the same way as the text did (for example, causes were linked to their consequences, or examples of the same phenomenon were linked). Finally, to score the answers, this model was used in a normative way: the subject's responses were compared to the model to decide whether their answer was correct or not. The three experts agreed on both the choice of concepts and the categorization of concepts that formed the normative model. In Appendix 3, the genetic manipulation text is reproduced with its sorting task. I highlight part of that task here in order to explain its functioning (1).

1. Here's a list of key words taken from the text. Make groups of words that you think should go together on the basis of the text. You can make as many groups as you want, and they can be of any size. Draw a circle for each group you want to make and put the right numbers in the circle.

1. solving the world hunger problem
2. crossing existing crops
3. moratoria
4. making new proteins

etc.

According to the normative model, key concept 1 and 4 had to be categorized together. Key concept 2 (an original application of genetic manipulation) and 3 (a solution to the dangers of genetic manipulation) did not belong in that same group. The participant's score was calculated by counting the number of items categorized in accordance with the normative model. Suppose a participant would categorize the above example as follows: 1, 4 and 3 together, and 2 apart. This means that out of a possible 4 items to be categorized correctly, this participant would receive a score of 3, since item 1, 4 and 2 were categorized according to the model. Item 3 was misplaced in a group where it did not belong and therefore, this item did not result in any points. If participants combined two or more groups into one big group, the score was lowered by 0.5

points per combination. In the case of our example, one single group of 1, 2, 3 and 4 would result in a score of 3: the category was combined with two others, lowering the score by 1 point. Finally, I calculated the proportion of correctly categorized items, and multiplied the proportion by 10, thereby creating a scale from 1 to 10. This means that final sorting scores range from 0 (no correctly sorted items) to 10 (all items correctly sorted).

Appraisal was measured in the same way as in the pilot experiment, since no problems with this operationalization occurred. In the following Table, taken from Chapter 3 and repeated here, the scales are repeated.

Dimension	Operationalization	
Appealingness	Clear	Vague
	Readable	Not readable
	Credible	Not credible
	Professional	Not professional
	Reliable	Not reliable
Accessibility	Difficult	Easy
	Effortful	Effortless
	Simple	Complex
	Coherent	Not coherent
	Halting	Fluent

Table 5.2 *Operationalization of the evaluation scales*

Feeling of Knowing was operationalized by two different questions (Glenberg, Wilkinson, & Epstein, 1982; Rawson & Dunlosky, 2002): on a 4-point scale, participants had to indicate whether they did not understand the text at all, or whether they understood the text perfectly. The second question asked participants to rate their performance in a hypothetical test. In the Dutch educational system, pupils receive grades on a 10-point scale, which was also used for this question.

5.2.3 Participants

As in the pilot experiment, an expert-novice paradigm was used to operationalize prior knowledge. Sixty-seven medical students participated in the high knowledge group, 64 law or history students participated in the low knowledge group. The medical students were expected to have more prior knowledge about organ donation and genetic manipulation than law students. This expectation was confirmed with a t-test: $t(129) = -20.3, p < .0001$. Students without prior knowledge received an average score of 0.29 on the

comprehension score, students with prior knowledge received on average 3.55 (on a scale from 0 to 5). Therefore, in the remainder of the analyses, these two groups can be considered a high-knowledge and a low knowledge group.

5.2.4 Design

The experimental factor prior knowledge was a between subjects factor with values high and low. The experimental factor text version was a between subjects factor with the values implicit and explicit. Furthermore, there were two genres: informative and persuasive, and two text topics: genetic manipulation and organ donation. In total, 8 different texts were used in this experiment: 2 topics * 2 versions * 2 genres. These factors were integrated in a Latin Square Design: every participant read 2 texts, one of which implicit and the other explicit, one of which persuasive and the other informative, one of which on genetic manipulation and the other on organ donation. The experiment took about 40 minutes. Participants were instructed not to turn back the pages they had already read, the same as in the Pilot Experiment.

5.2.5 Results on comprehension

The reliability for the sorting tasks in terms of Cronbach's alpha was considerably higher than in the pilot experiment (α genetic manipulation = .56, α organ donation = .64), but still not very good (van Wijk, 2000). However, considering the fact that this is not uncommon in text comprehension measures (see Chapter 4), these scores are acceptable. Text topic did not influence the results and therefore, the results presented in the following section are collapsed over text topic.

Comprehension was measured with a sorting task. The following formula was used in order to calculate the sorting score:

$$\text{SO score} = ((\text{SO1} + \text{SO2} + \text{SO3} + \text{SO4}) / \text{SO maximum score}) * 10$$

In other words, the proportion of correctly categorized items is calculated and then multiplied by 10 to make the SO score resemble normal test scores on comprehension, usually on a scale from 1 to 10.

Two-way ANOVAs were calculated to test the effect of coherence marking and prior knowledge on text comprehension. The results in Table 5.3 show that there is a main effect of marking of coherence for the persuasive genre: ($F(1,125) = 2.83, p = .05, \eta^2 = .14$) to the extent that the explicit version

is understood better. There is no effect for the informative genre (n.s., $p > .1$). In both cases, there is no interaction with prior knowledge and no main effect of prior knowledge.

	Informative	Persuasive
Without marking	6.04 (2.15)	6.03 (2.33)
With marking	6.49 (2.21)	6.70 (2.27)

Table 5.3 *Comprehension scores from Experiment 1, ranging from 1 (very low) to 10 (very high) on the sorting task for informative and persuasive texts (means, SD)*

5.2.6 Results on appraisal

A two way ANOVA with marking of coherence and prior knowledge showed a main effect of marking of coherence ($F(1,244) = 7.85, p = .05, \eta^2 = .03$) and a main effect of prior knowledge ($F(1,244) = 4.76, p < .05, \eta^2 = .02$), see Table 5.4. The implicit versions were, as expected, evaluated more negatively than the explicit versions. The readers without prior knowledge were more lenient in their judgments.

	Evaluation score for all texts	Score for informative texts	Score for persuasive texts
Implicit versions	3.23 (0.81)	3.01 (0.72)	3.50 (0.83)
Explicit version	2.94 (0.75)	3.07 (0.79)	2.89 (0.70)

Table 5.4 *Effects of marking of coherence on evaluation (mean scores on a 7-point scale (1 = positive, 7 = negative) and SD)*

Although I did not differentiate between genres in my hypotheses, I still checked whether this assumption was correct. These same effects were tested for the different genres. Although the effect was not significant for the informative texts, the data showed a strong trend in the expected direction. For the persuasive texts, the effect was significant: $F(1,127) = 19.03, p < .001, \eta^2 = .14$.

5.2.7 Results on FOK

Participants were asked to rate their own level of comprehension by means of two questions: one on a 4-point scale, and one on a 10-point scale. On both

scales, the effect is visible. We will report on both scales combined. There is a main effect of coherence marking in the persuasive texts, ($F(1,127) = 3.30$, $p = .03$, $\eta^2 = .03$). In the informative texts, the same effect is visible, but less strong: $F(1,122) = 2.40$, $p = .06$, $\eta^2 = .02$). In all cases, the texts with coherence marking cause a higher feeling of knowing than the texts without marking. It is very important to check whether FOK is accurate: are the participants able to *correctly* assess their understanding? A correlation analysis revealed a predictive relation in the informative genre ($r = .86$, $p < .05$), and in the persuasive genre ($r = .49$, $p < .001$).

Aside from the main effect of coherence marking, there is also a main effect of prior knowledge on FOK both in the informative ($F(1,125) = 14.15$, $p < .001$, $\eta^2 = .10$) and in the persuasive texts ($F(1,125) = 14.61$, $p < .001$, $\eta^2 = .10$). This effect is most visible at the 10-point scale, but is also present at the 4-point scale. On average, the low knowledge participants rated their understanding of the text lower (6.77) than the high knowledge participants (7.55). However, both groups rated their understanding as more than adequate (in the Dutch educational system, an average of 5.5 is needed to pass an exam).

5.2.8 Conclusion for comprehension

Experiment 1 showed that coherence marking leads to better comprehension for the persuasive genre, as in the Pilot Experiment. For the informative genre, there is the same tendency, but the results do not reach significance. This result in the informative genre differs from that presented in Chapter 3, where we saw a clear interaction between coherence marking and prior knowledge. Is this difference caused by the more reliable text comprehension method used in this chapter? This does not seem very likely, because the results for the persuasive genre *were* replicated. It is far more likely that the cause of these differences lies in operationalization of prior knowledge. The low knowledge readers behaved exactly the same in the Pilot Experiment and in Experiment 1: they performed better on comprehension questions after having read the explicit text in the informative genre and in the persuasive genre. For the high knowledge group, we see differences between both experiments. In the Pilot Experiment, genre interacts with the effect of prior knowledge on text comprehension ($F(1,104) = 8.62$, $p < .05$, $\eta^2 = .08$), in Experiment 1 we do not see this effect ($F(1,135) = 1.02$, $p > .3$). Therefore, the difference between these two experiments is likely to be caused by the high knowledge group. There can be several reasons for

this different behavior of the high knowledge group. For instance, they come from different backgrounds (Medical experts versus Biology experts). Therefore, their prior knowledge representation might have differed. Also, their behavior may have been different due to external factors, such as a hypothetical exam the next day or a big party the night before. There is no way of knowing whether one of these factors has played a role in the results. However, it seems advisable to keep as much control over the variable prior knowledge as possible, to eliminate as many differences between the high knowledge groups as we can. Therefore, I decided to replicate this part of the experiment, using an instruction paradigm in order to keep this variable under optimal experimental control.

5.2.9 Conclusion for appraisal and FOK

The effect of coherence marking on appraisal is clear: the positive effect coherence markers have on the reader's opinion of the persuasive text was replicated. The effect seems to be stronger in a persuasive context than in an informative context, and occurs for both knowledge groups.

Coherence marking also causes a higher Feeling of Knowing. Readers estimate their comprehension higher when the text they have read was explicitly marked. Participants were quite capable of assessing their own level of understanding in both genres and both knowledge groups. Coherence marking does not give them the *illusion* of having understood the text; their impression is accurate.

These results on appraisal and FOK all replicate the results from the Pilot Experiment. These effects seem robust and stable: coherence marking has a positive influence on people's opinion about the text.

5.3 Experiment 4: Comprehension reconsidered

In Experiment 3, the high knowledge group behaved differently than the high knowledge group in the Pilot Experiment. Although there are many possible reasons for this, one such possibility can be verified: the operationalization of prior knowledge. It seems to be a sensible idea to keep as much experimental control over the knowledge variable as possible. Therefore, in the current Experiment, the prior knowledge paradigm has been changed. We no longer make use of an expert-novice paradigm, but decided to manipulate prior knowledge directly in an instruction paradigm.

An instruction can be given in many different ways. Mostly, instructions are given in a booklet or through an interactive lesson (see the Introduction to this Chapter). We prefer a lesson over a booklet in the context of a reading experiment, because of the modality effect (among others Moreno & Mayer, 1999): students are expected to learn better when verbal input is presented auditorily as speech rather than visually as text. Still, a downside to the use of interactive lessons in the context of reading experiments is that for every session the same interactive lesson needs to be taught. Otherwise, there could be differences in prior knowledge between sessions. Since interactive lessons depend on the input of the participants, it is impossible to keep the content absolutely constant. Therefore, we have chosen another way to provide participants with the necessary prior knowledge: through short movie clips. These clips explain in a simple way how complex processes such as genetic manipulation work. Every time they are played in an experimental context, they are exactly the same. Again, the modality effect is still an excellent reason to prefer a movie clip over a text booklet.

5.3.1 Materials

Experiment 4 is an adaptation of Experiment 3. Materials and assessments are kept constant and the only difference lies in the fact that prior knowledge is manipulated by means of short movie clips, taken from Dutch current affairs shows. These clips were two short movie clips of approximately 15 minutes each. The clips were chosen on the basis of the two criteria. First, they had to have as little overlap with the content of the actual experimental texts as possible. Second, they had to activate relevant concepts and explain basic ideas that were necessary for comprehension of the experimental materials. We chose clips that applied the necessary knowledge to a slightly different but related situation. For instance, the genetic manipulation movie clip explained the process of manipulation of viruses and bacteria. We expected that this explanation would make it easier to understand the consequences of genetic manipulation in plants, which was put forward in the text. The organ donation movie clip introduced the technique of transplanting animal organs in human bodies. By explaining this whole process, we expected that the actual process of human organ transplantation, as well as the risks involved, would become clearer to the participants.

5.3.2 *Participants*

Two hundred and forty four participants, all bachelor students from Utrecht University, took part in the experiment. They were selected for being novices in the domains of organ donation and genetic manipulation, and the necessary prior knowledge was given to them for one of these two topics. No prior knowledge was provided for the other topic, so that they remained novices with respect to this topic. The success of this paradigm was tested by means of 3 control questions for each topic. A t-test was performed to check the prior knowledge manipulation. The analyses show that this technique of providing the necessary knowledge did have the desired effect: after having seen the movie clip on genetic manipulation, participants had more knowledge about genetic modification ($t(243) = 9.32, p < .001, \eta^2 = .13$) than those who had not seen the movie clip. The same effect occurred for organ donation ($t(243) = -7.97, p < .0001, \eta^2 = .10$).

5.3.3 *Design and procedure*

Prior knowledge was a between-subjects factor. Half of the participants saw the movie clip about genetic manipulation, the other half saw the movie clip for organ donation. These clips were shown to the participants in groups of approximately 25. After the clip, they read two texts: one on each topic, one implicit and one explicit text, and one informative and one persuasive text. The Latin square design was the same as in Experiment 3.

5.3.4 *Results*

The sorting tasks were analyzed in the same way as in Experiment 3. In addition, they were scored by a second rater to test whether two independent raters would attribute the same scores to each participant. The inter-rater reliability for these 102 observations was calculated with a correlation coefficient: Pearson = .93. Also, the internal reliability was acceptable: Cronbach's $\alpha = .60$. Again, as in Experiment 3, two-way ANOVA's were performed to investigate the effects of prior knowledge and coherence marking on comprehension.

The results for the informative genre show an interaction effect of coherence marking and prior knowledge ($F(1,243) = 8.03, p < .01, \eta^2 = .03$). When analyzing both text topics separately, there are two similar interaction effects (GM: $F(1,119) = 3.81, p < .05, \eta^2 = .05$ and OD: $F(1,116) = 5.70, p < .05, \eta^2 = .05$).

Therefore, these results are collapsed and presented together in Table 5.5 and Figure 5.1.

	Informative Implicit	Informative Explicit	Persuasive Implicit	Persuasive Explicit
Less prior knowledge	5.97 (1.72)	6.61 (1.80)	6.20 (1.92)	6.40 (1.87)
More prior knowledge	6.94 (1.55)	6.30 (1.94)	5.79 (2.31)	6.46 (2.07)

Table 5.5 *Comprehension scores from Experiment 4, ranging from 1 (very low)-10 (very high) on the sorting task for the informative and the persuasive genre (means and standard deviations)*

In Table 5.5, we see that the low knowledge group performs better after having read the explicit version than after having read the implicit version ($t(116) = 1.97, p < .05$). For the high knowledge group, the implicit version causes better comprehension scores than the explicit version ($t(122) = 2.01, p < .05$).

For the persuasive genre, the results show neither an interaction with prior knowledge nor a main effect of prior knowledge, but rather a main effect of coherence marking. The explicit text yields higher results on the sorting task than the implicit text. This effect is significant for the organ donation text ($F(1,120) = 58.78, p < .05, \eta^2 = .98$), for the genetic modification text there is only a trend toward significance ($p < .1$).

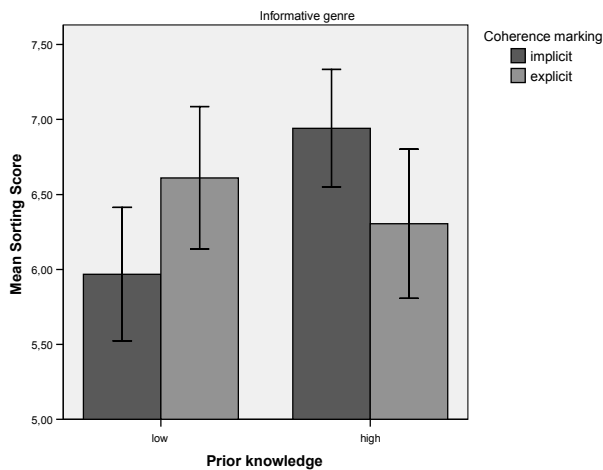


Figure 5.1 *Comprehension scores on a scale from 1(low) to 10 (high) for the informative genre for both knowledge groups in Experiment 4*

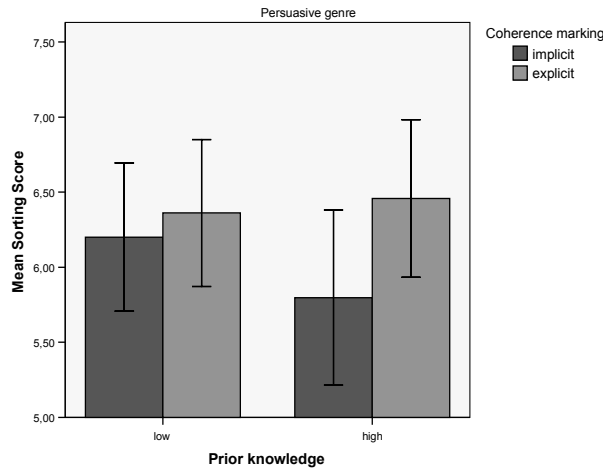


Figure 5.2 *Comprehension scores on a scale from 1 (low) to 10 (high) for the persuasive genre for both knowledge groups in Experiment 4*

5.4 Conclusion

In summary, Experiment 4 showed an interaction between coherence marking and prior knowledge on comprehension. In Table 5.6, the results from the Pilot Experiment, Experiment 3 and Experiment 4 are compared, and a distinction is made between both text genres.

	Pilot Experiment	Experiment 3	Experiment 4
Informative genre	Interaction of coherence marking and prior knowledge	n.s.	Interaction of coherence marking and prior knowledge
Persuasive genre	Main effect coherence marking	Main effect coherence marking	Main effect of coherence marking

Table 5.6 *Results on Comprehension from the Pilot Experiment, Experiment 3 and Experiment 4*

The most trustworthy results were found in Experiment 4. These results were clear, reliable and based on an improved text and methodology, compared to the Pilot Experiment and Experiment 3. The factor prior knowledge was under experimental control. We can conclude from Experiment 4 that for informative texts, prior knowledge and coherence marking interact. Low knowledge readers

benefit from the explicit texts. However, high knowledge readers perform better after having read implicit texts, probably because these texts cause them to process deeper and more thoroughly than when they read the explicit text. For persuasive texts, there is no such interaction: for both knowledge groups, the explicit text is the most beneficial in terms of text comprehension.

5.5 Additional analyses

The remaining question is: Why do we see no interaction in the persuasive genre, as opposed to the informative genre? A detailed look at the data shows that for low knowledge readers there is no difference in the effect of coherence marking between both genres ($F(1,234)=0.01$, $p>.9$). In both cases, low knowledge readers benefit from an explicitly marked text, whether the text is informative or persuasive. This effect can be explained by the integration and inference function of coherence markers: the markers probably help low knowledge readers integrate text segments and infer causal relations (Noordman & Vonk, 1997).

Yet, for high knowledge readers, the effect of coherence marking differs per genre ($F(1,246)=3.79$, $p<.05$, $\eta^2=.05$). When reading informative texts, high knowledge readers benefit from an implicit text that apparently causes deeper processing and comprehension. In other words, because the text does not help them they are ‘forced’ to integrate and infer themselves. This deeper processing results in better comprehension (McNamara & Kintsch, 1996). By contrast, in the persuasive genre the beneficiary effect of the absence of markers on comprehension is not present. The sorting tasks for the informative and the persuasive texts were identical, which enables us to compare these scores on comprehension between genres. When we look at Table 5.3, we see that in the informative genre, high knowledge readers on average performed 6.94 on the sorting task in the implicit condition, but only 5.79 on average in the persuasive genre. This is a significant difference ($t(124) = 3.26$, $p<.01$). For the explicit version, there is no such difference between the informative genre and the persuasive genre ($p>.05$). These analyses help refine our question: why do results on comprehension for the high knowledge readers on the implicit text change so much from one genre to another? I will elaborate further on this question and discuss a possible explanation in the general discussion in the following paragraph.

5.6 Comprehension: general discussion

The most important result from the Experiments in this chapter is the replication of McNamara and Kintsch's (1996) and McNamara, Kintsch, Songer and Kintsch's (1996) interaction effect of coherence marking and prior knowledge on text comprehension, but this time with texts that were varied systematically in the linguistic marking of coherence relations and text structure. In the current experiments, only purely linguistic markers of coherence were manipulated. It was with these experimental manipulations that I replicated the interaction effect of prior knowledge and coherence marking on text comprehension in the informative genre.

Additionally, I wanted to investigate which operationalizations were optimal for experimental research on these interactions. The operationalization of prior knowledge played an important role in the outcome of the Experiments and having prior knowledge under optimal experimental control was crucial. Although it seems attractive to operationalize prior knowledge with an expert-novice paradigm, thereby making use of elaborate stable knowledge structures, this was not the most stable operationalization in our Experiments. In both the Pilot Experiment and in Experiment 3, this expert-novice paradigm was used. In contrast, I used an instruction paradigm in Experiment 4. The results showed that the experts in the Pilot Experiment behaved differently than the experts in Experiment 3, but the novices behaved in exactly the same way in both Experiments. The problem did not lie in the selection of the novices: it did not matter whether they were History or Law students. Apparently, it did matter whether we chose Biology or Medical students. Therefore, a more precise manipulation of expert knowledge was needed. In Experiment 3, the prior knowledge that was needed to activate necessary concepts was provided to the participants. Apparently, the most successful empirical distinction between experts and novices was achieved through precise activation of relevant concepts by means of instruction. The conclusion that the instruction paradigm is the most successful operationalization of prior knowledge is contrary to the intuitive expectation: one would expect the expert-novice paradigm to be more stable, since it relies on extensive knowledge structures that have been established over time. The conclusion is the more remarkable because it is contrary to earlier findings by McNamara and Kintsch (1996), where instruction did not have the expected effects³. How can operationalizations that simply provide instruction and knowledge during 30

minutes be more reliable? The most plausible answer lies in the activation of relevant concepts (Anderson, 1984). The movie clips did not *literally* provide the same information as the text. They presented analogies and parallel situations to the ones in the text. For instance, when the text explained genetic manipulation in plants and food, the movie clip explained the process of human genetic manipulation. By providing equivalent but not identical information, it is plausible that the relevant concepts were activated in the reader's knowledge structure. Activation may play a lesser role when an expert-novice paradigm is used: then, the researcher relies on the expert's elaborated knowledge structures that novices do not have. This may be the reason why the operationalization of prior knowledge was most reliable when an instruction with a movie clip was given: the relevant concepts were activated in the reader's knowledge base. Clearly, further research is needed to confirm this tentative explanation. In Chapter 8, this surprising result is discussed in more detail.

A third goal of Experiment 3 and 4 was to compare effects of coherence marking and prior knowledge on text comprehension in two different text genres. The effects of coherence marking and prior knowledge in persuasive contexts seem to differ from those in informative contexts: I did not find the interaction effect that was present in the informative genre, but rather a main effect of coherence marking. In the additional analyses for Experiment 4, I have refined the crucial question to: Why do results for high knowledge readers on the implicit text change so much from one genre to another? Although I do not yet have any final answers to this question, I would like to speculate that in an informative context high knowledge readers use extra processing effort to process the implicit text deeper. In the persuasive context, however, they have the same extra processing effort at their disposal, but the context makes them use it differently: A persuasive context causes readers to judge and evaluate, to form, change or maintain an opinion, to engage in counter argumentation, etcetera. Since cognitive processing capacity is limited (among many others Gibson, 1998; Johnston & Heinz, 1978; Paas, Tuovinen, Tabbers, & van Gerven, 2003), there is not enough cognitive energy left for deeper processing that is needed to make up for the absence of coherence marking. This might be the reason why, in the persuasive context, high knowledge readers do not benefit from the implicit text. The resources that could be used to process deeply in order to benefit from such a text are

invested in other, persuasion-related cognitive processes and cannot be used for text comprehension.

Even though I can only speculate about the exact explanation for the differences between both genres, I have shown how useful it is to compare text genres by using texts that largely share the same content, but present this information in a different context. Furthermore, studying more than one genre provides more insight into the functions of coherence marking. The context in which the markers are presented influences the effects they have on comprehension. Secondary task methodology (Britton, Glynn, Meyer, & Penland, 1982; Paas, Renkl, & Sweller, 2003) could be a helpful method to test whether my cognitive effort explanation for the difference between the informative and the persuasive genre is indeed a valid one. Still, the results so far seem to indicate that coherence markers have more complex effects than simply making a text easier to understand. Coherence markers may make the process of understanding easier, and subsequently, in some contexts the extra cognitive resources available may then be attributed to other cognitive activities, such as weighing information in the text carefully and forming an opinion. This means that coherence markers do not only affect comprehension, but also persuasion. Markers function differently in informative texts, in which they play an important role in improving the cognitive representation of readers who do not have much content knowledge. The role of coherence markers in persuasive texts will be the focus of Chapters 6 and 7.

It's a miracle **but** we've made it (Audi Quattro). **And** you're done. (Amazon.com). **Because** so much is riding on your tires (Michelin). Peugeot. **Pour que** l'automobile soit toujours un plaisir. Danone. C'est bien **parce que** c'est bon ! L'Oréal. **Parce que** je le vaux bien. **Want** het leven is meer dan afwassen alleen (Dreft). Op een dag drink je geen bier meer, **maar** drink je Grolsch. Minder zout, **dus** je proeft meer kaas (Maaslander).

In this chapter, I propose four processing steps for coherence marking in the persuasive text genre. Together, these steps cause a forewarning mechanism: when readers identify the attempt to persuade them, they are no longer open to persuasion. This mechanism is triggered by subjective coherence marking and not by objective coherence marking.

In previous chapters in this dissertation, I have argued that linguistic marking of text coherence not only plays a crucial role in text comprehension, but that it also affects readers' opinions about the text (appraisal), their opinions about their own cognitive processes (meta-cognition) and their opinions about the text topic (persuasion). This chapter focuses on the latter: Can coherence markers influence persuasiveness of a text? And if so, how does this process take place? The advertising slogans at the beginning of this chapter are a few examples of famous and less famous advertising slogans with coherence markers. They set the tone of this chapter: Intuitively, the markers do more than simply mark a relation between two (persuasive) text segments, or in this case, they do more than mark the relation between the brand and the slogan or tagline. But what do they do exactly? If they can influence persuasion, how does this work?

Coherence markers are known to provide 'processing instructions' (see also Chapter 1): they give the reader information on how to integrate the incoming information with previously read information and they elicit inferences (Britton, 1994; Noordman & de Blijzer, 2000; Noordman & Vonk, 1997; Sanders & Noordman, 2000). Most of the research on coherence markers has focused on informative and narrative text. Could it be that coherence markers give the reader additional processing instructions, aside from integrating and inferencing, particularly in an argumentative context? Perhaps coherence markers not only provide information about how to integrate information and which inference to make, but also information about the text genre, the author and the author's intent. This information can then be taken as a processing instruction: readers process text differently according to the genre they have attributed to the text (Zwaan, 1994).

Not all coherence markers are expected to have this effect. Specifically, I expect the subjectivity of a marker to cause readers to recognize the persuasive nature of a text, and thus process it differently than they would process an informative text. Marking of subjective and marking of objective relations are known to have different effects on text processing (Traxler, Bybee, & Pickering, 1997; Traxler, Sanford, Aked, & Moxey, 1997). These differences have been attributed to the complexity of subjective relations compared to objective relations. My approach is slightly different: if we consider coherence markers as processing instructions (Noordman & de Blijzer, 2000), then subjective marking could provide *other* processing instructions. If these

subjective markers are prototypical to a genre and therefore may signal a genre, then genre-specific processing could be triggered. If the subjective markers signal the persuasive intent of the author, as I propose in this chapter, they may trigger specific persuasive processing and specific reactions to this intent.

In this chapter, I argue that subjective coherence markers can have a *forewarning* effect. In short, forewarning means that in verbal communication readers and listeners feel a strong influence of the sender, and identify this as an attempt to influence them. As a consequence, they build up resistance and persuasion does not occur. How does this process take place? Which steps must be taken in order for forewarning to occur? I propose four steps that are important in the process of forewarning. The first step is *detection of subjectivity* and concerns the text itself: which textual features can cause forewarning? More specifically, in the context of this dissertation, which coherence markers are expected to cause forewarning and which ones are not (6.3)? The second essential step is for the readers to *identify the author's intent* (6.4). The third step concerns their *reaction to the intent*: what do readers do after this intent has been identified as persuasive (6.5)? And the fourth and last step is the final *influence on the persuasive power* of a text (6.6). Before getting into these four steps, a short summary of the results in the Pilot Experiment is given. The Pilot Experiment already addressed the issue of coherence marking and persuasion, giving rise to some specific questions.

6.1 Markers of coherence in persuasive texts: summary of the pilot

Before describing the forewarning effect in more detail, it is important to keep in mind what has been discussed and empirically found so far concerning persuasion and coherence marking. In section 2.3.2, I described several reasons to suspect that coherence marking might influence persuasion. Several theoretical accounts of coherence marking would predict that coherence markers give argumentative direction to a text (Anscombe & Ducrot, 1983; Snoeck Henkemans, 2001; Verhagen, 2005). This means that coherence marking might influence the persuasiveness of a text. Although there are some recent studies to investigate these effects (Heller & Areni, 2004; Vivanco, 2005), there is no clear prediction on how coherence marking can influence persuasion. Does the presence of coherence marking make a text more or less persuasive? And how can we explain this?

In the Pilot Experiment (Chapter 3), these possible persuasive effects of coherence marking were initially tested. Surprisingly, coherence marking had no effect on persuasion in the persuasive genre. However, for informative texts, there was an interaction effect of coherence marking and prior knowledge on persuasion: low knowledge readers agreed more with the explicit informative text than with the implicit informative text. For high knowledge readers, the effect is quite the opposite: they agreed more with the implicit informative text than with the explicit informative text.

Unexpected as these results were at first sight, they are in need of a theoretical explanation. That the effects on persuasion occurred in the informative genre indicates that even informative texts have the possibility to change the reader's mind. This is not so surprising in the light of findings by Petty and Cacioppo (1979). They found that when participants believed that they were simply being informed, they were more likely to accept statements as true than when they thought they were being persuaded. Murphy (2001:676) adds the following: 'whether a text is written with persuasion in mind or for some alternative purpose like informing, it has the possibility of elaborating, transforming or altering an individual's conceptions'.

These results offer an interesting insight in the functioning of persuasive text. Apparently, texts that are clearly argumentative or persuasive, can overshoot themselves: people seem to be more persuaded by neutral information than by colored information. This conclusion is hardly new and has often been drawn before, for example, in the following quote by Anscombe and Ducrot: *On fait croire d'autant mieux qu'on n a pas l'air de chercher à faire croire*. [It is easier to persuade people when it looks like one is not persuading]. This phenomenon can be partly explained by the forewarning principle, introduced in the next section.

6.2 Forewarning

Forewarning has been studied in psychology for over 40 years (McGuire & Papageorgis, 1962; Petty & Cacioppo, 1979) and is defined as a social-psychological factor that causes resistance to persuasion. Forewarned participants are aware of the fact that someone is trying to persuade them. They will intuitively produce more counter-argumentation and will strengthen their own position. Therefore, a warned participant will be more difficult to persuade (see also 6.6 for more details on the effects of forewarning on persuasion).

Petty and Cacioppo (1979) conducted a famous experiment in order to provide evidence for forewarning and its inhibiting effects. In this experiment, half of the participants were 'warned' before listening to a message by the following instruction: 'the tape was designed specifically to persuade you'. The other half of the participants received the instruction: 'the tape was prepared as part of a journalism class project'. The message that they heard provided arguments in favor of a comprehensive exam that students would have to take prior to graduation. The results were very clear: Participants, who had been warned, agreed less with the comprehensive exam proposal than the un-warned participants. Forewarning is taken literally in this experiment; it consists of a verbal explicit warning of persuasion before message processing.

A broader approach to the forewarning principle is that of Friestad and Wright (1994). Their point of view is that likely targets for persuasive communication have topic knowledge, agent knowledge (about the sender), and persuasion knowledge. All these types of knowledge influence people's reactions to persuasive attempts. Because of all this knowledge, they may recognize attempts to influence them and experience forewarning. Forewarning is one possible reaction that people may experience, but there might be many others as well.

Although the concept of forewarning and influences of reader characteristics on the process of forewarning (Chen, Reardon, Rea, & Moore, 1992; Zuwerink & Devine, 1996) are widely recognized and accepted, little is known about the *text* characteristics that may provoke such a reaction. In fact, it remains to be seen whether or not text characteristics can cause the same forewarning reaction as an explicit verbal warning beforehand.

6.3 Step 1: Detection of subjectivity

If we apply forewarning theory to coherence marking, the prediction is that the explicit versions make the persuasive appeal clearer to the reader and are therefore less persuasive. However, this prediction is too simple. Not all markers of coherence can be expected to have a forewarning effect. We need to consider the well-known distinction between subjective and objective causality (see also Chapter 1): a causal relation either exists between two events in the world on which the author simply reports (objective), or because the author is constructing the causality (Sanders, 2005), namely that he is arguing towards a conclusion by presenting arguments in favor of a subjective claim. This

distinction resembles Sanders et al's (1992) source of coherence (semantic versus pragmatic relations) and Sweetser's (1990) distinction between content versus epistemic relations (but see discussions by Pander Maat & Degand, 2001; Pander Maat & Sanders, 2001 for the differences).

The following examples illustrate this difference between subjective and objective markers.

1. In the T.V.-series 'Lost', 48 people are stranded on an unknown island, because their plane crashed.
2. 'Lost' is the best T.V.-series of the last decade, because all my friends watch it.

In the first example, *because* expresses a causal relation that exists in reality. The speaker simply reports this relation, where there is a causal connection between a plane crash and being stranded on an island (at least in movies). A paraphrase (Sanders, 1997) would be something along the lines of 'The fact that they crashed leads to the fact that they are stranded on an island'. In the second example, *because* is used as a subjective marker. Apparently, the speaker considers the behavior of friends as an indication for the quality of a T.V. show. This relation is not verifiable in external reality, but the speaker takes responsibility for this causal claim-argument relation. This relation could be paraphrased as: 'The fact that my friends watch Lost leads to my claim that it is the best TV-show'. The alternative objective paraphrase, 'the fact that my friends watch Lost leads to the fact that it is the best TV-show', clearly does not capture the intended meaning.

In these examples, it is possible to use the same marker, namely the English connective *because* to express both an objective relation and a subjective relation. *Because* is one of the connectives that can mark both types of relations. This is not always possible as some lexical markers can only mark objective relations, such as *as a result* or *consequently*, and some connectives can only express subjective relations, such as *therefore*, which has specialized in conclusion relations (Knott & Sanders, 1998). In fact, the objective relation in 1) can be expressed by the Dutch *doordat*, but the relation in 2) cannot be expressed by *doordat*. In this case, Dutch speakers will prototypically use *want*, while *omdat* is possible. *Doordat* is an exclusive objective marker (Pander Maat & Degand, 2001; Pit, 2003; Sanders, 2005).

Previous research indicates that marking of subjective and marking of objective relations have different effects on text processing (Traxler, Bybee, & Pickering, 1997; Traxler, Sanford, Aked, & Moxey, 1997). These differences have been attributed to the complexity of subjective relations compared to objective relations. I propose a different explanation, as indicated above, namely that in a subjective context these markers give additional processing instructions to the reader concerning the text genre, the author and the author's intent. Readers process text differently according to the genre they have attributed to the text (Zwaan, 1994). I predict that the subjectivity of a marker can cause readers to recognize the persuasive nature of a text, and thus process it differently than they would process an informative text.

In sum, step 1 in the forewarning process consists of detecting subjectivity, a characteristic that provides additional processing instructions to the reader. In prototypical persuasive texts, there may be more characteristics that lead to detection of subjectivity and thus to forewarning aside from the marking of subjective relations. Subjectivity can be detected because of the appearance of a text ('it looks like an advertisement so it is likely to be persuasive') or because of the subjective content. In any case, when subjectivity is detected, it can subsequently provide processing instructions to the reader. The nature of these processing instructions is explained in the following steps.

6.4 Step 2: Recognizing and identifying the author's intent

Aside from subjectivity, another crucial notion for the forewarning hypothesis of coherence marking is that of author's intent. We know from previous research that readers are inclined to identify the author's intent (Zwaan, 1994). In Zwaan's (1994) study, readers process a text differently depending on their expectations about its genre. In other words, identification of text genre leads to different types and strategies of processing.

In the case of subjective and objective coherence marking, I predict the following. When people read a text with subjective coherence markers (see above), they infer the author's intent. When they read a text with objective marking, I expect that such an inference will *not* be made. In other words, I consider the objective situation as a default: when there are no cues present in a text to conclude otherwise, readers will assume that the intent of the author is to provide information. But when textual clues occur that seem to indicate otherwise, an inference is necessary to recognize and identify the author's

intent. Subjective marking leads to an author's intent inference, whereas objective marking is not strong enough to elicit such an inference. In the latter case, readers stick to their default expectation.

Gibbs (1984) is a strong believer in the importance of author intent for reading and understanding. He stated that readers are strongly disposed to attribute intentionality to human action. Many aspects of how we understand written text are influenced by the search for communicative intentions. 'Readers who are aware of who an author is, and what that person's status might be, are much better able to understand and evaluate expository and literacy texts' (1999: 177). A central idea to Gibbs' work is that even though people never know for sure what is meant by a text, a work of art or a movie, they will definitely try to figure it out.

Contrary to Gibbs' view is that of Barthes' 'Death of an Author', a 1968 essay in which Barthes explains how texts have meaning and existence independent of the author, but dependent on the context and reader. This point of view has been adopted by Dixon and Bortolussi (2001), who claimed that it is not relevant for a reader to know what the author may or may not have intended, and that readers do not ask themselves these types of questions. Although this is in fact a very theoretical debate, some empirical evidence has been gathered for Gibbs' point of view. These studies show that readers form an image of the author when they read text (for example Nolen, 1995; Tierney, LaZansky, Raphael, & Cohen, 1987).

A good example of a study showing how readers perceive an author and how text characteristics may influence this image is given by Nolen (1995). She showed that the reader's image of the author influences cognitive and affective reading experience. She investigated women's reactions to a statistics text with a visible author and a text with no clear presence of the author. The texts differed in the extent to which the author revealed attitudes and personality. In the visible author text, the author was clearly present: *I was a bratty little kid like that, or I know something about the biology of the situation*. In the other text, the authors revealed nothing about themselves and remained anonymous. They did not make use of the first person pronoun but wrote exclusively in the third person. The visible author text led to very detailed descriptions of the author in the think aloud protocols (for example *This author seemed to do a lot, whether they did it successfully or not for me, they did a lot of visualization*). However, the image that participants formed of this author varied

considerably, depending on self-efficacy, cognitive involvement and intrinsic motivation. Nolen concluded that a visible author of a text may be welcomed by readers when perceived as helpful and supportive. However, this effect was unpredictable: some students found the visible author to be manipulative. It is difficult to conclude which text version led to better comprehension, because Nolen focused primarily on the reader's estimate of their own understanding (meta-cognition) instead of using direct measures of comprehension. In any case, there seems to be a quite strong influence of what readers *perceive* the author's intent to be, not necessarily of the true author's intent.

Recognizing and identifying the author's intent is crucial for the forewarning mechanism. If an explicit textual characteristic is present to cause readers to deviate from their default expectation of informative text and recognize the author's intent, such as subjective marking, an inference is made to identify the intent as persuasive.

6.5 Step 3: Reacting to the identified author's intent

Once readers have recognized the attempt to influence their opinions and identified the author's intent, how do they react? We know from forewarning theory that recognizing the author's intent as persuasive eventually leads to diminished persuasion. But what happens between the moment that people notice the attempt to influence them and the end result of diminished persuasion?

The first possible reaction that people may experience is called *warning induced resistance*. The classical study by Petty and Cacioppo (1979) showed how this works. In their study, forewarning elicited both anticipatory counterarguments and counterarguments during processing. Also, the number of favorable thoughts towards the text topic, which were reported after having heard the message, reduced when participants were forewarned. In fact, forewarned participants tried harder to find reasons to disagree with the message arguments in order to reassert their attitudinal freedom. This reaction is referred to as *warning induced resistance*, because counterarguments and less favorable thoughts towards the text topic lead to an increase in resistance caused by the message. Wood and Quinn (2003) mention the same reaction, that they refer to as *biased processing*. People generate thoughts that support their existing view and counter the position to be advocated. Typically, this reaction

occurs when people are involved and motivated to process the textual information deeply.

A second, completely opposite reaction that people may experience is *anticipatory change* (Wood & Quinn, 2003). Forewarning makes people feel vulnerable. They interpret the warning as a threat to their identity and as a consequence they already shift to the attitude that is advocated in the message before even having heard or read it. By doing so, they avoid the threat of the actual message: they already made up their mind beforehand. Logically, this occurs more often if people are not very involved in the actual topic of the message. This reaction can also be explained in terms of Brown and Levinson's Politeness Theory (1987)¹⁷. Politeness consists of two specific kinds of desires: positive face (the desire to be liked, admired, etc) and negative face (the desire not to be imposed upon, not to be used or manipulated). In general, people tend to cooperate in maintaining face in interaction, based on the mutual vulnerability of face. The important question here is whether coherence markers can act either as face threatening or as face preserving. Barton (1995) described a corpus-analysis on connective expressions that she conducted with 50 academic essays, which she characterized as argumentative. She presented an analysis of the use of connectives and their *interpersonal* functions in academic discourse, using Brown and Levinson's Politeness Theory to explain the use of some connective expressions. Face-threatening acts (FTA's; disagreement, criticism) are often preceded by face-saving moves (establishing shared ground). Connectives that were used to mark the face-threatening act were mostly contrastive connectives (*but, however*), and markers that were used before these face threatening acts were *to be sure, of course, true enough*. An example of this finding in academic work translated to the texts used in the Pilot Experiment in Chapter 3 would be:

3. Of course, helping other people is important. But there is no other country in Europe where as few people are organ donors as in the Netherlands.

¹⁷ Politeness theory is originally interactional and conversational, the speaker and the hearer both being present at the same place at the same time. However, for instance in the view of Anscombe and Ducrot (1983) and Moeschler and Reboul (1994), argumentative texts are necessarily an (implicit) dialogue.

The strategy that the writer adopts is first creating common ground, then making the counterclaim: the actual FTA. The marker *of course* emphasizes the common ground, which is face preserving. The marker *but* emphasizes the face threatening act. Barton suggests that the markers are added to make the utterance less face threatening. Although Barton's analysis seems plausible to me, she offers no empirical evidence for the interpretation that connectives help softening face threatening acts. Moreover, a more general question remains: Do readers perceive attempts to persuade them through text as a FTA? In the case of example 3), the utterance can only be face threatening if the reader is Dutch and involved enough to feel threatened. I argue that face threatening by means of a persuasive text can only occur when readers are involved and when the situation applies to them personally. Politeness theory could be an explanation for why readers experience resistance when they feel they are being influenced.

In sum, there are two different reactions that people may have when forewarned: warning induced resistance and anticipatory change. Coherence marking is a possible forewarner that is present *in* the text, not presented *before* the text. Therefore, it is impossible that readers change their opinion before reading the text, as is the case in anticipatory change. I expect subjective coherence markers to increase the likelihood that readers recognize the author's intent, compared to the objective coherence markers. This might then lead to more internal counter-argumentation, less favorable thoughts toward the text topic, and more resistance to the content of the message. These processes occur internally and not consciously. They can be expressed externally, but they do not necessarily have to be expressed or consciously experienced by the reader.

6.6 Step 4: Maintaining the original attitude and making it stronger

So far, I have proposed that the forewarning mechanism for coherence marking consists of three different steps. The first step is the detection of subjectivity, the actual textual characteristic that causes the effect. The second step consists of recognizing and identifying the author's intent. Consequently, this leads to a possible internal reaction of counter-argumentation or face threatening. What is the eventual outcome of this process? How does this process influence the persuasiveness of a text?

There are three different reactions that readers may have to forewarning. The first reaction is that of *warning induced resistance* (see above): If readers experience resistance, engage in counter argumentation and experience less favorable thoughts, this is generally assumed to have no effect whatsoever on persuasion. Basically, people stick to their own original opinions and nothing changes. In this case, there is no persuasive effect of the text. This is the general effect that forewarning is expected to have: it increases resistance and therefore it inhibits persuasion.

A second possible reaction to forewarning can be *anticipatory change*, as we have seen above. In the case of anticipatory shift, some form of persuasion occurs: attitudes are changed in the desired direction, to avoid the attempt to influence them. As soon as readers notice the attempt to persuade, they shift their opinions. Still, this attitude change, if one can call it that, is considered to be of short duration and to be context dependent. That is, when the potential threat disappears, people tend to revert back to their initial attitude. Again, this is not the most plausible reaction in the case of coherence marking. Coherence marking is a less explicit forewarner than the literal message: ‘I will now try to persuade you’.

There is a third possible effect of forewarning on persuasion. Tormala and Petty (2002) formulated this effect as follows: ‘What doesn’t kill you makes you stronger’. They argue that once an attempt to influence has been resisted, beliefs and attitudes are strengthened. In this aspect, meta-cognition is very important; this position assumes meta-cognitive knowledge about the (persuasion) process (see also Chapter 2 and 3). In other words, resisting an attempt to influence might strengthen attitude certainty. This idea is at the basis of McQuire and Papageorgis’ (1962) *inoculation theory*: just as human bodies can be inoculated against diseases, beliefs and attitudes can be inoculated against persuasive attacks¹⁸. In their experiments, Tormala and Petty (2002) found that resisting a persuasive message does indeed increase the certainty with which people hold their initial attitudes. However, this effect did not occur when the attack was perceived to be weak; apparently, resisting a weak attack is considered less relevant. In essence, this means that when people feel they have

¹⁸ However, intensive counter argumentation such as predicted by the first possible effect that I listed, can also cause strengthening of original attitudes, but in a direct cognitive way. In this case, attitudes are strengthened through meta-cognition.

done a good job in resisting a strong message, they can infer that their attitude must be a valid one.

In sum, forewarning can affect persuasion in three ways: by strengthening one's own view through cognitive effort (counter argumentation), through meta-cognitive effort (resisting the appeal makes me stronger) and by causing anticipatory attitude change. Which of these three effects do I expect in the case of coherence marking? Certainly, the latter is not particularly relevant to the context of coherence marking, as previously discussed. The coherence markers do not occur at the very beginning of a text, such as the message 'I will now try to persuade you' in the classical forewarning experiments. The subjective markers occur at some point in the text, when readers are already actively processing the textual information. Attitudes and beliefs have already been activated, and it seems extremely unlikely for anticipatory change to occur at this point. However, readers could use an anticipatory change strategy from the moment they have encountered an explicit cue of forewarning in the text. This would only occur for readers who are not involved and not motivated. Readers who are more involved with the text topic are expected to react differently.

Both other reactions are possible in the case of subjective coherence marking. I expect this type of markers to cause readers to maintain their original position and strengthen it. I believe that internal counter argumentation, reduced favorable thoughts and trying to maintain attitudinal freedom are the most likely reactions to the forewarning character of coherence marking.

6.7 Conclusion

In this chapter, I have argued that the forewarning account is a plausible explanation for persuasive effects of coherence marking. This leads to the expectation that subjective coherence markers cause forewarning, because they lead to recognition of the author's intent to persuade. They consequently lead to more internal counter argumentation, they cause more resistance, and finally they strengthen the reader's original attitude. No persuasion occurs. Objective coherence markers are not expected to cause a forewarning reaction. They do not lead to recognition of the author's intent as persuasive, and thus they do not cause more counter argumentation and less favorable thoughts.

Aside from the forewarning mechanism, there may be other accounts from various other disciplines that also explain effects of coherence marking on persuasion, such as the social psychological theory on experiential processing (Meyers-Levy & Malaviya, 1999), Cognitive Load Theory (Paas, Renkl, & Sweller, 2003) and the cognitive linguistic Mental Space Theory (Fauconnier, 1984). These theories can provide additional insights into the functioning of coherence markers in a persuasive context, as we will see in the Discussion to Chapter 7. They all address aspects of coherence marking and its possible effects on persuasion, but the most complete and persuasive prediction can be formulated with forewarning theory.

This chapter started with a short summary of the results in the Pilot Experiment. Forewarning theory explains why we did not find any effects of coherence marking in the persuasive texts. In the Pilot Experiment, the persuasive texts were overtly persuasive. Titles already signaled the intent to persuade from the beginning of the text. These texts signaled the author's intent so clearly to the reader that the markers did not contribute any more. In other words, forewarning was already caused by other factors such as the titles, so that the specific coherence marking forewarning was simply invisible.

In the next chapter, I examine the forewarning hypothesis. If forewarning occurs, then the author's intent and the diminished persuasiveness of the text must occur when subjective marking is used. In other words, the forewarning hypothesis that is the focus of the next chapter can be formulated as follows:

H₁: Implicit and objective text versions are more persuasive than subjective ones.

H₂: Subjective marking makes the author's intent to persuade more visible than objective or no marking.

H₃: The Author's intent and the Persuasiveness are linked: when a reader recognizes the author's intent to persuade, this directly influences the (absence of) persuasiveness.

In the previous chapter, I proposed the forewarning hypothesis to account for persuasive effects of coherence marking. The current chapter reports two experiments testing this hypothesis: the first experiment (7.2 and 7.3) examines the effects of coherence markers on persuasion, whereas the second experiment (7.5 and 7.6) examines the effect of coherence markers on both persuasion and processing.

¹⁹ This chapter has been submitted as an article, co-authored by Ted Sanders, Leo Lentz and Rolf Zwaan (Kamalski, Lentz, Sanders, & Zwaan, submitted). Part of the research in this chapter has been realized at Florida State University in Tallahassee, FL, USA with the financial support of the Netherlands Organization for Scientific Research.

The aim of this chapter is to test whether the forewarning phenomenon (explained in detail in the previous chapter, short summary provided in 7.2) is a correct explanation for the persuasive effects of coherence marking. In this chapter, two experiments on the effects of coherence marking in a persuasive context are reported. The first experiment (7.3 and 7.4) focuses on off-line effects, whereas the second experiment (7.5 and 7.6) combines on- and off-line evidence. The crucial issue is that subjective markers are expected to have a different effect on persuasion than objective ones, because the author is so prominently ‘on stage’ (Langacker, 1990) that the author can hardly be neglected and therefore causes resistance (see Chapter 6).

7.1 The forewarning hypothesis for coherence marking

Forewarning (McGuire & Papageorgis, 1962; Petty & Cacioppo, 1979) causes resistance to persuasion (for more information see Chapter 6), because forewarned readers are aware of the fact that someone is trying to persuade them. The following definition of forewarning is used: ‘a text causes forewarning when it signals the persuasive intent of the author to the reader, thereby decreasing the chances of persuading the reader.’ Crucial in this definition is the link between diminished persuasion and recognizing the author’s intent as persuasive. If forewarning occurs, both these effects occur: not only diminished persuasion, but also the recognition of the author’s intent.

It is unlikely that objective markers of coherence cause forewarning. Also, texts without marking are not expected to cause forewarning. On the other hand, subjective markers may very well cause forewarning because the author is more visible, thereby rendering the attempt to influence the reader more visible. Hence, we can formulate the following Forewarning Hypotheses (H_f), taken from Chapter 6:

H_{f1} : Implicit and objective text versions are expected to be more persuasive than subjective ones.

H_{f2} : Subjective marking makes the author’s intent to persuade more visible than objective or no marking.

H_{f3} : The Author’s intent and the Persuasiveness are linked: when a reader recognizes the author’s intent to persuade, this directly influences the (absence of) persuasiveness.

7.2 Experiment 5: off-line effects of coherence marking on persuasion

7.2.1 Materials

The materials consisted of two Dutch texts with different topics: genetic manipulation and organ donation. These materials were previously used in Experiments 1, 3 and 4. The texts were on average two pages long, with approximately 25 manipulations of coherence marking per text. Four different experimental conditions were constructed: implicit, objective, subjective and both types of marking combined. This last condition, where all markers were used at the same time, serves as a comparison: it is not included in the hypothesis, because we do not know for certain whether the effects of objective marking and subjective marking cancel each other out. However, we have a strong intuition that this might be the case and to test this, the ‘all marking’ condition was included.

A very important remark that needs to be made is that the experimental comparison between objective and subjective marking is a complex one. There are restrictions on the use of connectives and lexical markers of relational coherence. Therefore, markers can only be used when the connected segments and the relation expressed fit in with these restrictions. For instance, the Dutch connective *daardoor* can only be used to mark a non volitional CAUSE-CONSEQUENCE relation (Pit, 2003; Stukker, 2005). It is very important to take these restrictions into account when constructing experimental materials. Otherwise, the texts would not seem natural to the reader and they would resemble what Graesser, Millis and Zwaan (1997) called ‘textoids’: experimenter-generated texts that have no ecological validity.

Marking of coherence was taken in a broad sense in this experiment. The categories of coherence markers that were manipulated have in common that they all explicitly mark a relation that otherwise would have to be inferred by the reader. The first category is that of global coherence. Both headings and organizers were manipulated. The following examples illustrate the types of coherence marking that were manipulated.

1. Why should you be an organ donor?
2. There are two types of donation, organ donation and tissue donation
3. What are the consequences of genetic manipulation?
4. What are the dangers of genetic manipulation?
5. There are two types of donation.

6. There are three important risks to genetic manipulation
7. Because there are not enough organs available, there are long waiting lists and people might die
8. Doctors can determine objectively whether a patient is brain-dead. Therefore, organ donation is a safe and careful process

The first category of coherence marking concerns global coherence (headings, organizers). Example 1 illustrates a heading. This heading indicates that the following paragraph will provide reasons. In the implicit version, there was no heading. 2) is an example of an organizer, indicating that in the following section, these two different types of donation will be explained. Again, in the implicit version, there was no organizer present.

How were the objective and subjective versions constructed? Every heading was constructed in two versions: a subjective one and an objective one. Headings were the only type of markers which I could manipulate at exactly the same place in the text in order to make a direct comparison between objective and subjective conditions. For instance, the objective heading 3) had a subjective counterpart 4). The objective heading does not give away the argumentative point of view. The subjective version expresses the writer's opinion on the subject, whereas the objective version does not.

The organizers were attributed to the subjective or the objective category. They could not be replaced by a marker of the other category and were simply omitted in the implicit version. This means that the subjective and objective organizers did not occur at the same point in the text. For instance, 5) was considered objective, since there is no clear argumentative direction in that statement. 6) was categorized as being subjective, because it has an argumentative direction to it. It expresses an evaluation by the author: these are not just consequences; these are risks, thereby expressing a subjective evaluation.

The second category of coherence markers concerned local coherence relations, manipulated by means of connectives (7) and lexical cue phrases. All connectives and lexical cue phrases were attributed to either the objective or the subjective version. Again, this means that objective and subjective connectives did not occur at the same location in the text. 7) is an example of objective marking. 8) is an example of subjective marking with a connective.

The following passage (9) shows an example of a text passage from this experiment, where the objective markers are underlined and the subjective

markers are italicized. The experimental text was much longer, about two pages. The implicit version contained very few markers; as few as possible were included without sacrificing the natural quality of the text. In both texts, no more than 5 markers were present in the implicit text. The objective marking version contained only the objective markers and the objectified headings. The subjective version contained only the subjective markers and subjectified headings. The all-markers condition contained all markers and the subjective headings. In Table 7.1, the experimental conditions in the experiment are represented.

9. Many genetically manipulated crops have been made herbicide-resistant, other crops have even been made resistant to insects with a gene from a bacterium. The goal of this manipulation is that the plant will develop a poison against a certain plague of insects. *Therefore*, genetic manipulation is designed to solve problems which are caused by agriculture itself.

Condition	Objective markers	Subjective markers
Implicit	-	-
Objective marking	+	-
Subjective marking	-	+
All marking	+	+

Table 7.1 *Conditions in Experiment 5*

The most important dependent measure is that of persuasion. Persuasion can be measured at different levels, namely beliefs, attitudes and intentions (Fishbein & Ajzen, 1975, see also 3.1.2). I measured persuasion at all three levels, but attitude change was the central focus of attention. The three levels of persuasion were operationalized by presenting the participants with statements for which they provided agreement ratings on a 7-point Likert scale (1= strongly agree, 7= strongly disagree). Five different statements were included. An example of a belief on genetic manipulation is: *genetic manipulation decreases natural variance in plants*. A possible attitude is *genetic manipulation is bad*. From the attitude follows the intention *I intend to eat only biological foods*. Only the attitude after reading the text was measured, not the initial attitude. Previous research has shown that asking people their opinion on a certain topic makes it very difficult to subsequently change their opinions with a text (Hoeken, 1994),

precisely because of the same forewarning mechanism that we are investigating here. Therefore, I used a post-text design only, in which the average scores on attitude per text version were compared. Because participants were assigned randomly to one of the conditions, the average initial attitude score should be equal across conditions. If so, post-text differences can be ascribed to the text manipulations.

The forewarning hypothesis was further tested by asking participants to rate the writer's intention. If forewarning occurs, they will be more likely to detect the writer's intention to change their attitude. This means that, although objective and subjective coherence marking is *not* manipulated at the same place in the text, I am still comparing off-line attitude measurements for all these conditions. Here we are not (yet) interested in the influence of *one* single manipulation of coherence marking, but in the effect that a text with many subjective markers may have on the reader. Therefore, I combine all objective markers in the objective condition and all subjective markers in the subjective condition. Although these markers do not occur at the same point in the text, we can still compare the off-line measurements on attitude between these conditions. This design does not enable us to see which exact marker is causing the effect, but it only gives insight in the collective functioning of *all* the markers in the same category (objective versus subjective).

7.2.2 Participants

Previous research with informative text has shown that the effects of coherence marking may vary as a function of the reader's prior knowledge (see Experiment 3 and 4 in Chapter 5). Therefore, prior knowledge about the text topics was controlled. An expert-novice paradigm was used, although Chapter 5 has provided evidence that an instruction paradigm should be preferred. At the time, I was unaware of these results: the experiment in Chapter 5 was conducted later than the experiments reported here. Two groups of participants were selected who were expected to differ with respect to their prior knowledge. One-hundred Dutch medical and 100 law students from Utrecht University participated in the experiment. Their average age was 20.24 ($SD = 3.58$). They were all in their first two years of their BA-program. 30% were male, 70% female. Given that the text topics were organ donation and genetic manipulation, I expected the medical students to have more prior knowledge at their disposal than law students. The level of prior knowledge was controlled

for by asking five questions about basic information on DNA and human anatomy. The expectation was confirmed ($t(190) = -47.677$, $p < .0001$, $\eta^2 = .75$): students without prior knowledge (law students) achieved an average score of .29 on the prior knowledge score, whereas students with prior knowledge (medical students) achieved an average score of 3.55 (on a scale from 0 to 5).

7.2.3 Procedure

The experiment took about 40 minutes. Participants were instructed not to turn back to the pages they had already read. After each text, the questions and statements concerning that particular text were given. This means that each package contained the following sections: the prior knowledge questions, the instruction to the first text, the first text itself, the persuasion statements for text 1, and finally rating of the author's intent. Then, this whole procedure was repeated for the second text.

7.2.4 Design

Prior knowledge was a between subjects factor with the levels high and low. Text version was a between subjects factor with the levels no marking, objective marking, subjective marking, and all markers combined. Each participant read two experimental texts in two of the four experimental conditions. Furthermore, there were two text topics: genetic manipulation and organ donation. The topics were varied within subjects. This resulted in the following design: 2 topics (organ donation vs. genetic manipulation) * 4 coherence conditions (no marking, objective marking, subjective marking, and finally a version with both objective and subjective marking in one text). These factors were integrated into a Latin Square Design.

7.3 Results

Since the results turned out to be the same for both text topics and both knowledge groups, the analyses were collapsed over text topic and knowledge level (see also the Discussion in section 7.3.6).

7.3.1 Effects of coherence marking on persuasion

The first question was whether or not the subjective and the objective version differed with respect to their persuasive effects. These effects were measured at belief, attitude and intention level (all items combined, for genetic manipulation

questions Cronbach's $\alpha = .72$, for organ donation $\alpha = .62$, with 1 measure at the belief level omitted). We can compare the effects on attitude for all the conditions in the experiment, because the text's content was exactly the same for all the conditions. They all contained the same facts. The only difference concerned the marking of coherence relations, but the relations themselves were present in all versions.

Table 7.2 shows the mean persuasion scores per condition. In accordance with the hypothesis, the version with the objective markers was more persuasive than the version with the subjective markers. An independent-samples t-test with persuasion as a dependent variable showed this difference to be significant ($t(122)=-1.651$, $p<.05$, $\eta^2=.02$, one-sided). Clearly, objective marking has a different effect on persuasion than subjective marking.

	Score on persuasion	Score on perceived intent
Implicit	4.10 (.67)	6.22 (1.02)
Objective markers	3.99 (.83)	5.49 (1.64)
Subjective markers	4.24 (.78)	6.30 (1.22)
Obj and subj. markers	4.07 (.75)	6.21 (1.20)

Table 7.2 *Effects of marking of coherence on persuasion and perceived intent. Persuasion scores indicate a mean (SD) on a 7-point scale (1 = agree completely with text standpoint, 7 not agree at all with text standpoint, intent scores are mean scores on a 7-point scale (1 = inform, 7 = persuade).*

However, there is no significant difference between the subjective condition and the implicit condition. This difference has to be significant in order for the forewarning hypothesis to apply: the subjective markers signal the intent to persuade, the implicit condition does not. These two conditions are thus expected to differ, but they clearly do not. We have to conclude that forewarning hypothesis 1 is partly confirmed.

Why is the difference between the implicit version and the objective condition so large? In other words: the objective version is more persuasive than the implicit version. This can be explained with the known effect that coherence markers have on text processing: they make processing easier. This may have resulted in stronger persuasive effect. A one way ANOVA indeed shows that according to the readers' own judgments, the no marking version (mean 3.49 on a 7-point scale) costs more energy than the objective version

(mean 2.86, $(F(3,190)=2.457, p=.06)$). This could be an explanation for the positive effect of objective coherence marking on persuasion.

7.3.2 *Effects of coherence marking on detection of persuasive intent*

We have established so far that the objective version was indeed more persuasive than the subjective one. However, it remains to be determined whether or not this is because a subjective version actually makes the attempt to persuade more visible to the reader. Therefore, a question regarding the perceived intent of the writer was included to measure this perceived intent. Participants had to rate the author's intent on a 7 point Likert scale from 'informing the reader' to 'persuading the reader'.

A one-way ANOVA with coherence version as grouping variable shows a significant effect on these answers ($(F(3,250)= 5.30, p=.001, \eta^2=.06)$). Tukey's post-hoc analysis shows that it was clearer for participants that the intent was to persuade when they read the version with all markers, the version with subjective markers or the implicit version, than when they read the objective version (see Table 7.2). These results confirm forewarning hypothesis 2.

To test the final forewarning hypothesis, I determined how the perceived intent of the writer influences the persuasive effect. The correlation ($r(251)= .13, p<.05$) shows that the perceived intent indeed influences the persuasiveness of a text. A second way of testing this influence is to include perceived intent as a covariate in the ANOVA. The effects should then diminish or even disappear. The latter is indeed the case: with perceived intent as a covariate, the effect of coherence marking on persuasion disappears ($(F(3,248)=.80, p>.5)$). In other words, the perception of the intent of the writer is crucial for the persuasive power of a text. If readers notice the attempt to influence them, the persuasive effect of a text diminishes. The subjective markers have this effect on the reader. This confirms forewarning hypothesis 3.

7.4 Discussion

In Experiment 5, we have seen that the text version with subjective markers was less persuasive than the version with the objective markers. The same effect was observed with the question on the writer's intention: the intention of the writer of the objective version was perceived to be less persuasive than that of the writer of the subjective version. The perceived intent was a successful

predictor of the persuasive effects of a text. This supports the hypothesis that when readers notice the attempt to influence or even manipulate them, the persuasive power of a text decreases.

Although these findings indicate that it is very important to distinguish between the two types of markers, i.e. subjective and objective ones, the forewarning effect has not been shown conclusively. I did find strong forewarning indicators: subjective markers cause more resistance than objective markers, and this effect is linked to the perceived intention of the author. There was no difference between the subjective and the implicit version. Hence it cannot be concluded that we are definitely dealing with a forewarning effect. This issue will be explicitly addressed in Experiment 6.

A second surprising result is that both knowledge groups yielded exactly the same results. One would expect individual differences such as prior knowledge to influence the effect of coherence markers. We have seen in chapter 5 that prior knowledge interacts with coherence marking on comprehension, and prior knowledge is expected to affect persuasion as well (Wiley, 2005). The absence of an effect of prior knowledge in the current experiment is an intriguing issue. It might be caused by the expert-novice paradigm that was used to operationalize prior knowledge. Chapter 5 has provided indications that a manipulation of prior knowledge is the preferred option.

In the current experiment, different text linguistic characters of markers were manipulated collectively (for instance connectives with lexical cue phrases with organizers). Even if we have found strong indications of a forewarning mechanism, it is not clear whether these markers all have the ability to contribute to the forewarning effect. It could also be the case that only the headings (*risks* versus *consequences*) were strong enough forewarners, so that the more subtle connectives did not affect persuasion and forewarning at all. The following study manipulates only connectives and lexical markers, the more subtle coherence markers compared to headings and organizers.

In summary, Experiment 5 showed an off-line effect of coherence marking on persuasion. Is this effect really forewarning, even though we only found a difference between the objective and the subjective version and not between the subjective and the implicit version? And if the effect is indeed one of forewarning, how does it affect on-line processing? This question was examined in Experiment 6.

7.5 Experiment 6

In Experiment 6, I tried to replicate the findings from Experiment 5 and, moreover, to investigate whether coherence markers can indeed be taken as forewarning markers by examining their effect on on-line text processing. In Chapter 6, I argued that forewarning occurs during processing. This has been shown in other studies as well (see 6.5).

In short, I expect subjective marking to cause readers to recognize the author's intent. This will then influence processing in such a way that it results in diminished persuasion. This process of recognizing the author's intent might take more time than simply understanding what is being said. In fact, an extra inference concerning the author's intent needs to be made. Inferencing generally takes time and therefore it might be reflected in reading times data (see also Chapter 1).

7.5.1 Hypotheses

The hypotheses follow from the theoretical overview in Chapter 1 and 6. We have seen that coherence marking can have both an integration effect and an inference effect (Cozijn, 2000; Noordman & Vonk, 1998). The integration effect means that immediately after a coherence marker readers speed up compared to the implicit version, because they can easily integrate information from the incoming segment to the previous one. The inference effect means that at the end of the sentence readers slow down compared to the implicit version, because they spend time inferring and validating the relation between the two text segments. The inference effect is often 'checked' with verification statements after the text: readers then have to validate statements from the text. If readers made the inference on-line, they are faster when answering the verification statements off-line. In other words, the inference effects seen in the reading times and the verification latencies are correlated.

7.5.2 On-line hypotheses

Specifically, on-line effects of coherence marking have mostly been investigated with objective marking. Therefore, the expectation for the objective markers is to replicate the integration and the inference effects.

H1: immediately after an objective coherence marker, readers speed up compared to the implicit unmarked version (integration effect) and at the end of the sentence, they slow down compared to the implicit version (inference effect).

It is more difficult to predict on-line effects of subjective marking. Among Graesser, Singer, and Trabasso's (1994) list of 13 types of inferences that readers possibly make when they read a text is the *author's intent*. This is a relevant inference in the present context of forewarning. Graesser et al. predict this inference only to be made on-line, if there is a strong pragmatic context for the inference. Subjective coherence marking might provide such a context. If forewarning occurs during reading, then the inference with respect to the author's intent is also made on-line. Subjective coherence markers are expected to generate these author's-intent-inferences. Objective markers are not expected to generate such an inference. Therefore, the hypothesis for subjective marking is:

H2: The subjective marker causes an additional inference to be made compared to the objective marker, concerning the author's intent. Therefore, there is an additional increase in reading times. The inference effect that we see when we compare the subjective marking version to the implicit version has to be larger than the inference effect in H1.

Obviously, the subjective and objective statement pairs differ completely in content. A direct comparison is impossible, since it is very likely that the results would be influenced by other factors, such as activation of beliefs about the truth of the statement (Voss, Fincher-Kiefer, Wiley, & Ney Silfies, 1993). Therefore, I only compare reading of the text with subjective marking and reading of the exact same sentence in the implicit version, and similarly the text with objective marking and the exact same sentence in the implicit version (see also 7.3.1). In the previous experiment, only off-line measurements were compared, making a direct comparison between the subjective and the objective text (as a whole) possible. Both texts contained the same information, only the markers differed. In the current experiment, where on-line measurements need to be compared, we can no longer make a direct comparison. We compare the objectively marked statement to the same

statement in the implicit text version. Also, we compare the subjectively marked statement to the same statement in the implicit version. This is an indirect method to compare the effects of the objective marker on on-line processing and the effects of the subjective marker on on-line processing.

Based on previous research, it is not easy to predict at what point in the text the resistance due to the forewarning effect will be experienced, and neither do we know exactly where the reader makes the inference about the writer's intent. The most likely point would be the end of the sentence, because that is the point in time where readers usually make causal inferences (see Noordman & Vonk, 1997 and many others). In the current experiment, the factor prior knowledge is not taken into account, because the first experiment showed that the effects on persuasion occurred for both knowledge groups.

7.5.3 Off-line hypotheses

Experiment 6 tested both on- and off-line effects of coherence marking, in order to replicate the earlier results and link off-line and on-line evidence directly. Verification statements were included to check whether a causal inference has been made during reading. Also, we asked the same question to rate the author's intent as we did in the previous experiment. Finally, we also measured persuasion with 5 attitude questions per text. This leads to the following off-line hypotheses.

H3: Participants answer the verification statements more quickly after a marked version (objective/subjective) than after an implicit version.

H4: Participants answer the question regarding the author's intent more quickly after the subjective version than after the implicit version or the objective version. They have made this inference on-line and are therefore faster when rating the author's intent.

H5: The subjective version is less persuasive than the objective version and the implicit version. Participants report a more positive attitude towards the text standpoint after the objective text.

7.5.4 Materials

Twenty-four short texts were used. They were all approximately one paragraph long, with an average of 75 words. Each text contained only one manipulation (compared to texts that were two pages long and contained 25 manipulations in

Experiment 5). The texts were of a persuasive nature, meaning that the authors all took a clear position in a debate and advocated a clear standpoint. Topics ranged from complex global or regional ones, such as the political situation in Israel or gun law policies in Florida, to more local and personal ones, such as parking around campus, sororities and fraternities, and college exit exams. Of each text, a version without marking was constructed, a version with one objective marker, and a version with one subjective marker. Marking was a within subjects factor, counterbalanced by means of three different lists. This means that of all 24 texts that a participant reads, 8 are implicit versions, 8 are objectively marked and 8 are subjectively marked.

Objective marking can only be manipulated at a point in the text where two segments are indeed connected by an objective causal relation (for instance, cause and effect). Subjective marking can only be manipulated at a point in the text where two segments are connected in a subjective causal relation (for instance, claim and argument). This means that we cannot compare objective marking and subjective marking directly. What we can do, though, is to compare the objective marking version to the no marking version, and the subjective marking version to the no marking version. Subsequently we can compare the objective-implicit pattern to the subjective-implicit pattern to see if there are any differences. This leads to the following specific hypothesis:

H 1+2: The objective-implicit comparison yields different results than the subjective-implicit comparison. In the latter, an extra inference effect occurs.

The following text is an example of the texts that were used in the on-line experiment.

10. 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./Therefore,/an exit exam/for college students/ needs to be implemented/as soon as possible./All students will benefit from such a system./

In the example (10), we see two markers, *consequently* and *therefore* (objective marking, *subjective marking*, / =new segment, target segment.1, target segment.2). In the objective version, only the marker *consequently* is present. The subjective marker *therefore* is omitted and replaced by a full stop. In the subjective version, *consequently* is not present, but the subjective marker *therefore* is present. In the no-marking version, neither marker is present. By doing so, the content, style, and choice of words is kept constant between all text versions. Only one word or 'lexical chunk' is different between the versions: the marker. Examples of objective markers that were used in the experiment are *as a result*, *thereby*, *consequently*. Examples of subjective markers in the experiment are *therefore*, *that is why*, *in conclusion*. Connectives and lexical cue phrases were balanced over texts and type of relation (objective and subjective).

Reading times were collected for four segments. Target segment 1 is the text segment that immediately follows the marker, then there are two middle segments, and finally target segment 2 is the last segment of the same sentence (for an example, see example 2 and Table 7.3).

These texts were constructed by adapting original persuasive texts from brochures and websites, because they had to be as natural as possible. The original texts were changed slightly so they all had the same structure: one paragraph long, presenting as much as possible an objective view on the situation but concluding with one clear standpoint. By keeping the texts as objective as possible, I wanted to trigger forewarning primarily with the marker. If forewarning would be caused by other elements in the text, it would be impossible to pinpoint the forewarning effect in reading times. The texts were both pre-tested by experts in persuasion and by readers from the target group (in this case, psychology students). They all considered the texts to be natural, although some terms were removed since they were not part of the average student's vocabulary.

7.5.5 Procedure

Participants read the 24 texts by means of a self-paced moving window paradigm (constructed in the software program E-prime), 8 of them in the no-marking version, 8 with an objective marker and 8 with a subjective marker. The texts were presented in segments of three or four words, depending on logical segmentation of the text. After each text, participants answered

verification statements (see below, example 11 and 12), enabling us to check whether they had made the necessary inference and persuasion statements to see whether they agreed or disagreed with the text standpoint (see below, example 13). In addition, they had to indicate their perception of the author's intent (as in Experiment 5). The following statements are examples from the experiment.

11. Objective verification statement: A college exit exam ensures that students all possess the same basic skills.
12. Subjective verification statement: Implementing exit exams will have a positive effect on the quality of education.
13. Persuasion statements: College exit exams need to be implemented. / If I have to vote, I will vote against college exit exams / College exit exams are a bad idea.

7.5.6 Participants

Forty-eight psychology undergraduate students from Florida State University participated in this experiment for course credit. Data from three participants were omitted from the analyses, because their overall accuracy on the verification test was below chance level. The average age of the 45 remaining participants was 18.8 years. Thirty participants were female, 15 were male. They were all native speakers of American English.

7.6 Results

First, the on-line results are presented, then the off-line results and, finally, a comparison is made between these two types of effects. The effects were analyzed by participants and by items.

7.6.1 On-line effects of coherence marking

As noted earlier, it is not possible to use an objective marker to mark a subjective relation and vice versa. The relation between two text segments determines what type of marker can be used. Therefore, data from the objective version are only compared to the implicit version. The same holds for the subjective version, which is only compared to its implicit counterpart. Comparisons concern the exact same sentence, in the objective and subjective condition preceded by a marker, in the no-marking condition preceded by a full stop.

TESTING THE FOREWARNING HYPOTHESIS - 157

Reading times were collected for the sentence after the marker in three parts: the segment that immediately followed the marker (target segment 1), the middle part of the clause (consisting of two segments), and the last segment before the sentence end (target segment 2). Each segment consisted of three or four words, depending on the logical segmentation of the text. The following target sentence in Table 7.3 illustrates the segmentation.

A person who drinks and drives is three times more likely to be involved in a crash than a sober driver.

Because of this,	the safe thing	is to ask	someone to be	the designated driver
Marker	Target segm 1	Middle segm.	Middle segm.	Target segm.2
Word 0	Words 1,2,3	Words 4,5,6	Words 7,8,9	Words 10,11,12

Table 7.3 *Segmentation of a target sentence in Experiment 2*

Outliers (more than two standard deviations above or below the mean per participant per condition) were removed from the sample (Ratcliff, 1993), which constituted less than 3% of all data. In Table 7.3, the reading times for the segments after the objective marker are compared to the same segments in the no-marking text. Also, reading times for the segments following the subjective marker are compared to the same segments when there was no marker present. This means that in Table 7.4, we can compare row 1 to 2 and row 3 to 4, but we can not compare them directly, given that the objective markers and the subjective markers mark different relations in the text. Their content differs and they cannot be compared directly. Statistical comparisons were performed by means of separate 2 (segments) * 2 (marking conditions: no marking versus either objective or subjective) repeated measures ANOVA's.

	RT Segment 1	RT Segment 2	RT segm 1 + middle segments + segm 2
No objective marking	1061 (310)	1279 (484)	4418 (1396)
Objective marking	1003 (402)	1130 (301)	4067 (1310)
No subjective marking	1273 (311)	1077 (270)	4304 (1988)
Subjective marking	1000 (241)	1162 (335)	3889 (1469)

Table 7.4 *Mean reading times and SD (RT) per condition: no marking, objective and subjective marking.*

These results show a different pattern of on-line processing for the same sentences after objective marking than after no marking. Also, the pattern of on-line processing for the same sentences after subjective marking differs from processing after no marking. Figures 7.1 and 7.2 illustrate these two patterns.

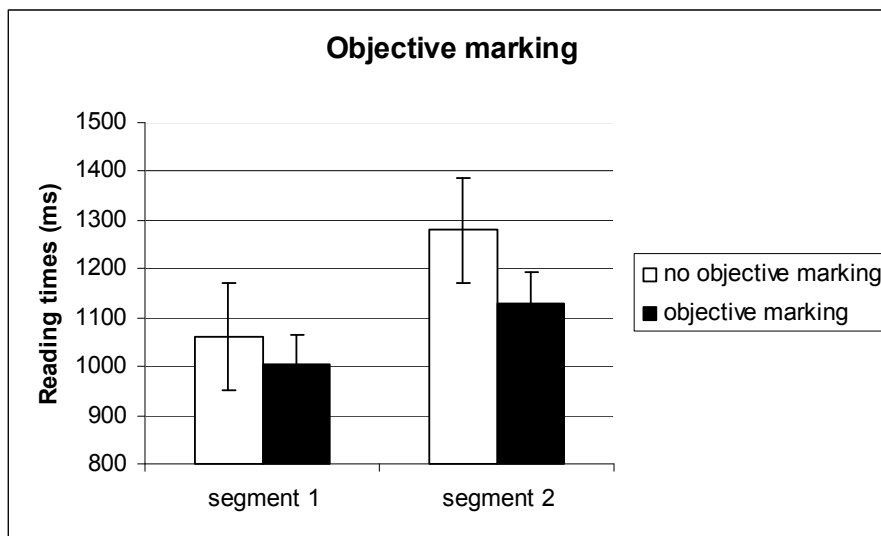


Figure 7.1: *Mean reading times for the no marking and the objective marking condition, for the immediately following segment (1) and the end of the sentence (2)*

The objective marker caused the next sentence to be processed more quickly in its entirety ($F_1(1,44)=2.23$, $p=.07$, $\eta^2=.05$, $F_2(1,22)=3.49$, $p<.01$, $\eta^2=.35$). The two target segments separately show the same tendency. We do find evidence for the integration effect, but not for the inference effect (see hypothesis 1).

For subjective marking, there is also an overall speeding up effect for the sentence in its entirety ($F_1(1,44)=4.31$, $p<.05$, $\eta^2=.09$, $F_2(1,22)=2.63$, $p<.06$, $\eta^2=.1$). But when we look at the target segments separately, we see that there are two different effects: a speeding up effect (integration) and a slowing down effect (inference). Immediately after the marker, processing was faster than in the no marker condition, but at the end of the second clause, the marker actually slowed down processing (interaction marking * segment: $F_1(3,41)=28.65$, $p<.01$, $\eta^2=.24$; $F_2(1,23)=12.19$, $p<.05$, $\eta^2=0.35$).

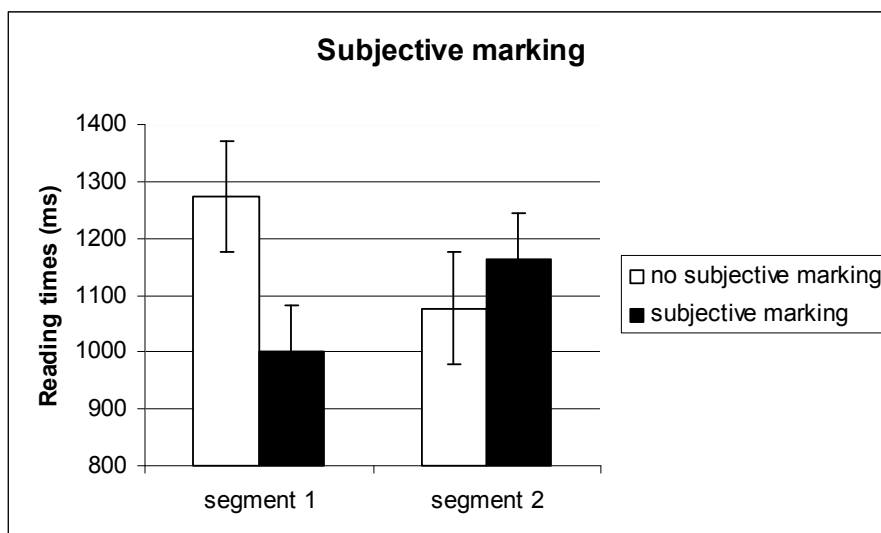


Figure 7.2: *Mean reading times for the no marking and the subjective marking condition, for the immediately following segment (1) and the end of the sentence (2)*

As predicted by hypothesis 2, there is an additional inference effect compared to the objective condition. In the objective condition, no inference effect occurred, whereas in the subjective condition, we do see evidence for such an inference effect.

7.6.2 Off-line effects of coherence marking

The off-line data may be able to explain why this inference effect occurs only in the subjective condition and not in the objective condition. The current experiment also contained verification statements (hypothesis 3), a question regarding the author's intent (hypothesis 4) and persuasion statements (hypothesis 5), intended to measure the off-line effects of coherence marking.

Verification latencies

The results on the verification statements *accuracy* were not influenced by coherence marking. All conditions yielded an accuracy of approximately 75 percent. However, the verification *latencies* did differ between conditions. Table 7.5 compares response times for the objective inference verification for two conditions (where there was no objective marker present and when the readers

read the objective marker) and the same comparison is made for the subjective inference verification statements for two conditions (no subjective marking vs. the subjective marker).

	Objective inference verification	Subjective inference verification	Author's intent question
No marking text	5444 (1725)**	5497 (1683)**	5198 (2048)
Text objective marker	4115 (1391)**		5085 (1917)
Text subjective marker		4599 (1557)**	4426 (1460)*

Table 7.5 *Mean reaction times and standard deviations to answer verification statements (for the objective inference and the subjective inference separately) and rate the author's intent (* $p < .05$, ** $p < .01$)*

Table 7.5 shows that readers are faster in responding to the verification statements if a marker is present. This is the case for both the objective and the subjective marker. A 2 (objective versus subjective verification latencies) * 2 (no marking versus marking) repeated measures ANOVA ($F_1(1,44)=24.23$, $p < .001$, $\eta^2=.36$; $F_2(1,23)=14.57$, $p < .01$, $\eta^2=.39$) shows that participants are faster when answering a verification question when they have read the marked text version. These results suggest that a marker causes the inference to be made on-line, whereas the absence of a marker does not lead to such an immediate inference. In the latter case, readers are capable of making this inference after reading, but only when asked a question concerning this inference. Taken together, these results seem to confirm hypothesis 3: when a marker is present, readers make causal inferences on-line. This seems to be in contradiction with the reading time data. We only saw an inference effect for the subjective marker, not for the objective marker. This issue is addressed in the Discussion.

Author intent

Participants were also asked to rate the intent they believed the author had with every text. Their answers and their reaction times were analyzed. Outliers more than two standard deviations above or below the average per participant and per text were removed, which involved 2% of all data. We can compare these

TESTING THE FOREWARNING HYPOTHESIS - 161

results for all three versions, since the text content was kept constant over text versions.

Coherence marking had no effect on the outcome of these ratings. However, there was an effect on the reaction times. The results in the last column in Table 7.5 show that participants answered this question on intent faster after having read a text with a subjective marker than after a text without a marker and a text with an objective marker ($F_1(2,86)=2.80$, $p<.05$, $\eta^2=.06$, $F_2(2,23)=8.83$, $p<.01$, $\eta^2=.27$). A Tukey's posthoc analysis shows that the subjective marker text is the only text that causes significant differences on the intent latencies. The results on the author intent question suggest that the inference concerning author intent was made on-line only in the case of the subjective marker. These results support hypothesis 4: subjective marking leads to an on-line inference concerning the author's intent.

Persuasion

As in Experiment 5, the experiment included persuasion statements at the belief, attitude and intention level. Reliability between the three statements was calculated for each text. Three texts had to be omitted from this analysis because the alpha of the persuasion statements was lower than 0.5. The average Cronbach's alpha for the persuasion statements of the other texts was .63. Therefore, the statements are analyzed together.

The three text versions (subjective, objective and implicit) were identical, only one marker differed per version. In the subjective version, there was one subjective marker present, in the objective version, there was one objective marker present and in the implicit version, I used neither of these two. The content in these three versions was exactly the same. So, if we see effects on persuasion, they can only be caused by the marker. The first column of Table 7.6 shows the results on persuasion for the three different types of marking.

Text version	Mean persuasion
No marking	3.17 (.61)
Objective marking	3.22 (.46)
Subjective marking	3.31 (.45)

Table 7.6 *Mean scores and standard deviations on persuasion (on a 7-point scale, 1 = totally agreeing with the text, 7= completely disagreeing with the text)*

These results are not significant ($F_1(1,43)=1.92$, $p=.08$, $\eta^2=.04$, $F_2(2,23)=2.19$, $p=.12$). There is a tendency for the subjective version to be less persuasive than the no-marking version. Although this could be an indication for forewarning, it is not significant and it is also not the same pattern as in Experiment 5, where the objective version had the strongest positive effect on persuasion.

An explanation for these different findings might lie in another variable, namely involvement. The meta-analysis conducted by Wood and Quinn (2003) provides evidence for the role of involving topics. When readers are highly involved with the text topic, they build up resistance by bolstering their own views. When the text topics are less involving, they change their opinion before the actual appeal, to avoid threat to their own attitudes. I performed a post hoc analysis on the data of Experiment 6, this time only including texts discussing topics to which American students could relate. I made a distinction between high involvement topics, such as parking around campus and college exit exams, and low-involvement topics, such as the political situation in Israel and hybrid car technology. This yielded the results in the last Table 7.7.

	Low involvement	High involvement
No marking	3.15 (.72)	3.45 (.75)
Objective marking	3.25 (.65)	3.36 (.54)
Subjective marking	3.36 (.54)	3.46 (.44)

Table 7.7 *Mean scores and standard deviations on persuasion for high involvement texts and low involvement texts (on a 7-point scale, 1 = totally agreeing with the text, 7 = completely disagreeing with the text)*

Interestingly, the pattern for high involvement replicates the results of Experiment 5. The objective version seems to enhance a text's persuasive power. The only significant difference is between the objective and the subjective marking conditions ($t(14)=1.68$, $p=.05$). For low involvement texts, the significant difference lies between the subjective and the implicit version: $t(14)=2.52$, $p<.05$. This suggests an interaction with involvement that I do not yet fully control. The explanation could be that in the case of low involvement texts, individuals are not inclined to spend much effort on the text. The objective marker makes understanding the text easier and results in more

persuasion. However, in the case of high involvement, people are inclined to spend more energy on the text. The objective markers lose their positive effect, because people compensate with their motivation and effort. Now the subjective marker signals the author intent, mentioned above, which causes resistance and results in diminished persuasion. It seems worthwhile to include the variable involvement systematically in follow-up experiments.

Relations between on-line and off-line evidence

A very important asset of Experiment 6 is that it offers the opportunity to link on- and off-line evidence. This could help explain the on-line processing patterns for objective and subjective marking from Figures 7.1 and 7.2.

There seems to be an apparent paradox in the objective marking results between reading times and verification latencies. The pattern of reading times is consistent with the integration effect, but does not provide evidence for the inference effect. However, the verification data provide evidence that suggests this inference *has* been made on-line. At first sight, this appears to be a paradox. However, this is not the first study to find no slowing in the reading times, but that nevertheless finds an effect on the off-line measures of inference-making (Maury & Teisserenc, 2005; Millis & Just, 1994; Sanders & Noordman, 2000). There are three possible explanations for these findings. It is possible that the inferences were made so quickly and automatically that they do not require significantly more processing time. Another possibility is that the used method (a moving window paradigm) is not sensitive enough to show the increase in processing effort that is generally called the inference effect. For instance, maybe readers slow down on several segments instead of on one segment. A third possibility is in line with Millis and Just's (1994) finding that coherence marking does not induce inference making, but quite the opposite: coherence marking requires fewer inferences and therefore it only causes an integration effect and no inference effect.

The situation for subjective marking is clearly different from the situation for objective marking. Here, we do see evidence for the integration and the inference effect of the subjective coherence marker, as I expected to find for the objective marking condition. The increase in reading times at the end of the sentence does not seem to be related to making an inference about the causal relation expressed. A regression analysis with reading times as predictor and verification latencies as dependent variable shows no effect. What

then is causing readers to slow down? Could this be another type of inference, for instance the predicted author's intent inference? The correlation between the reading times on the last segment and the reaction times for the question on the author's intent indicates that this might be the inference that readers make ($r(43) = .23, p < .01$). Readers who slow down at the end of the sentence after a subjective marker, are faster when asked to rate the author's intent. Following the same reasoning as for causal inferences, this could be an indication that they have inferred the author's intent on-line, during reading. Because of this on-line inference, readers can answer this question more quickly after reading. Hence, when readers slow down after a subjective marker, they appear to be making an inference concerning the intent of the author. This is a clear indication for the forewarning effect of subjective coherence marking.

7.7 Discussion

In Experiment 6, on- and off-line evidence for the forewarning effect of subjective marking was found. When looking at low-involvement texts, the subjective versions were less persuasive than the no marking versions, thereby providing direct evidence for the forewarning mechanism of subjective marking. This result solves the problem encountered in Experiment 5, where the subjective version only differed significantly from the objective version and not from the no-marking version. Experiment 6 not only solves this problem, but also offers an explanation for this, namely involvement. When we only include low-involvement texts, we see that the subjective condition has a negative effect on persuasion, compared to the implicit version. And this was the effect that we were looking for (see Experiment 5) if we want to show a forewarning effect. In sum, reading times and reaction times to verification statements indicate that subjective markers seem to cause readers to make an inference concerning the author's intent, whereas objective markers do not.

However, the effects of coherence marking on persuasion are not as pronounced as in Experiment 5; all persuasion scores for the American students – the participants in Experiment 6- range from slightly left of the middle of the scale to slightly right of the middle. In Experiment 5, the Dutch participants were much more pronounced in their opinions and attitudes. This might be a cultural difference concerning the involvement of students in political and global issues, as the split in high-involvement versus low-involvement suggests. The difference could also be due to the use of 7 point

Likert scales. Standard deviations are a little larger in Experiment 5, and this could be an indication that Dutch students are more likely to circle extreme points at the 7-point scale. Most American students circled a position in the middle of the scale. Another explanation could be the fact that in Experiment 6, participants had to give their opinion on 24 different issues, compared to only two issues in Experiment 5. Answering persuasion questions 24 times could cause participants to reflect less on their actual opinion and translation into the Likert scale, and circling more often the middle option on the scale. Still, the results on persuasion in Experiment 6 are significant, even though they are small.

The on-line evidence from Experiment 6 shows that objective marking causes faster processing, whereas subjective marking causes an instant speeding up effect, but at the end of the sentence, it causes processing to slow down. How do these findings relate to the two functions of coherence marking, integration and inference?

Traxler, Bybee and Pickering (1997) compared reading times after an objective marker to reading times after a subjective marker and found the same slowing effect as in Experiment 6, at the end of the sentence after a subjective marker. They explain their finding by hypothesizing that subjective relations require an inference to establish the nature of the causal consequence, usually a belief about events in the world. However, our regression analyses show that the slowing effect is not related to this particular inference, but rather to the inference concerning the author's intent. In the Traxler et al. experiment, there was no strong context to elicit inferences concerning the author's intent. The sentences were presented in isolation. Therefore, it is very unlikely that their results can be explained by a forewarning mechanism. Possibly, the slowing effect they report is caused by the relative complexity of the subjective relations: it is often harder to follow the line of reasoning in subjective relations than in objective relations (Sanders, 2005). This can also be the case in the current Experiment. This would explain the effects found in the reading times, but not the regression analysis that showed that reading times and author intent latencies were correlated. Therefore, it is more likely that, when slowing down, readers seem to be inferring something related to the author's intent.

However, there is a possible methodological problem with the operationalization of perceived author intent. When such a question is asked, the participant might become aware that this is apparently of interest to the

researcher, and this may cause the inference to be made. This could influence the processing of the next text, and the one following, and so forth. Only processing of the first text (when the question concerning the author's intent has not been asked yet) can be considered totally 'natural'. Still, because reading time experiments typically involve within subjects designs, I have measured author's intent in a within subjects design. There appeared to be no systematic difference between the first time participants were confronted with the questions and the subsequent occasions, suggesting that there is no need to worry too much about this methodological problem: readers do seem to infer authorial intent when they are slowing down at the end of the second sentence. Obviously, there is no way of knowing whether this is the *only* thing they are doing. Another possibility, already mentioned before, is that the inference that is prompted on-line by the subjective marker is simply more problematic to validate than an objective one, in line with Traxler et al.'s findings. Under such an interpretation, validating an objective inference could be simpler, because this relation exists in external reality. Validating a subjective relation would take more cognitive effort and thus more processing time. One way in which subjective relations are more difficult to validate than objective relations could be the presence of the author: not only do readers have to check the causal relation, but also the author's intent. The crucial new finding in this experiment is the evidence for an authorial intent in the case of subjective markers only: readers verified authorial intent faster and showed additional processing time at the end of the sentence where this inference is most likely to be made. I take this as evidence in favor of the forewarning effect of subjective marking. Whether or not other factors play an additional role remains to be investigated.

7.8 General Discussion

Although forewarning is a well-known and well-accepted phenomenon in social psychology, there has not been much psycholinguistic research on the actual text characteristics that cause such a reaction of resistance or the effects of forewarning on text processing. In this respect, the two experiments reported here constitute an important step forward. In marketing contexts as well as in public information contexts, causing the least resistance possible when you want to persuade people is a vital question to many writers and document designers.

The present experiments have shown specific influences of text characteristics on persuasion. Subjective markers of coherence, which make the author clearly visible as an argumentative language user, seem to cause a forewarning effect, whereas objective markers, 'only' expressing coherence relations in external reality, do not. The distinction between objective and subjective coherence marking affects the effectiveness of a (persuasive) text. Although discourse analysts have repeatedly argued that coherence markers influence the interpretation of persuasive text, the actual effect on on-line processing and text interpretation had never been previously empirically demonstrated. The most convincing text version for involved readers possesses the objective markers. For all readers, the least convincing version was the version with subjective marking. The results of both experiments show that carefully choosing your markers can influence the persuasive power of a text.

Nevertheless, there are several factors that require further investigation, such as involvement. Research on forewarning indicates that this is an important variable and in the present experiments, it seems that the factor involvement plays a role as well. Prior knowledge did not play a role in the experiments in this chapter, but this could be due to the operationalization. Individual differences, such as involvement, prior knowledge and motivation all need to be investigated in a systematic way.

Also, the apparent contradiction in reading times and verification latencies requires more attention. The reading times indicated an inference effect for subjective marking only, not for objective marking. However, the verification latencies indicated an inference effect for both coherence marking conditions. It could be that making an inference on-line costs time, but that we are simply unable to measure this delay. This is not likely, since we did find an inference effect in the reading times for the subjective condition. Maybe causal inferences are made so quickly and automatically, that there is no delay to measure. This is also not likely, since we did find a delay for one of our conditions. Another option would be that coherence markers cause fewer inferences to be made instead of more. The data from the experiments in this chapter are not conclusive. In any case, other reading time studies have found the same results (Maury & Teisserenc, 2005; Millis & Just, 1994; Sanders & Noordman, 2000); moreover, using a more sensitive method such as eye-tracking could provide more insight into this matter (Cozijn, 2000).

7.9 The four steps of the forewarning process

Finally, there is the question of the theoretical explanation of the differential effects of objective and subjective markers on on-line processing. What processes take place when readers slow down after the subjective marker? In Chapter 6, I proposed a four step model of this process. In this last paragraph of Chapter 7, I will review the model in the light of the data from both experiments. Below, the four steps are repeated.

Step 1: Detection of Subjectivity
Step 2: Recognizing and Identifying the Author's Intent
Step 3: Reactions to the Identified Intent: warning induced resistance
Step 4: Maintaining the Original Attitude

Step 1 was clearly confirmed by the data. Subjective and objective marking has different effects on the reader and the reader must detect the subjectivity in order for this difference to exist. This does not imply that readers consciously decide to detect subjectivity. When asked, they are able to reflect on these issues, as we have seen in previous chapters. This step need not be a conscious decision by the reader in order to be an important step in the model of processing instructions proposed in this chapter.

Step 2 concerns the Author's Intent. Subjective marking has shown to be an explicit textual trigger for the author's intent inference. The on-line data clearly show that an inference is made in the subjective condition, compared to the implicit condition. This is clear evidence for step 2, as well as for the assumption that an informative, neutral intent of the author is the default assumed by the reader: no inference occurred in the objective condition. If readers inferred the author's intent regardless of textual clues, but simply all the time, there would have been an inference effect in the objective condition as well.

Step 3 cannot be confirmed with the current data. The current experiments did not provide any information on the cognitive processes of elaboration and so forth that are assumed to cause resistance. It is a challenge to do so, because these processes may not be experienced consciously and in order to measure them, we would have to intervene in the natural reading and persuasion process, for instance by means of think aloud protocols. Interesting as these data would be, they pose a threat to the persuasion process: once

TESTING THE FOREWARNING HYPOTHESIS - 169

readers become aware of the researchers interest in their opinion, they tend to stick to their opinion and become somewhat resistant to change.

Step 4 concerns the result of the forewarning process: diminished persuasion. Here, the model is in need of some adjustments. We clearly saw that involvement plays a role and that involved readers react differently than un-involved readers. Involved readers benefit from objective marking: this version results in the most favorable attitude change. Probably, this effect is caused by the facilitating effect that coherence markers have on the comprehension process. Involved readers *want* to understand the pros and cons of a point of view before they make a decision. Objective markers help them to do so. Uninvolved readers do not fully process the information. Objective markers do not have a positive effect. However, the subjective markers seem to have a negative effect: they cause resistance and inhibit persuasion. In sum, step 4 differs for involved and uninvolved readers.

In conclusion, three out of four steps of the forewarning process, described in Chapter 6, are confirmed by the data. An additional factor, involvement, needs to be included. More research is needed to confirm step 3: Future research is needed to determine exactly what is increasing readers' processing time in the subjective marking condition. So far, the on-line evidence from Experiment 6 suggests that the subjective marking prompts an inference concerning the author's intent, and that this inference causes the increase in reading times. This is a strong indication of the forewarning effect that results from the subjective coherence markers.

In this chapter, I summarize the findings of the studies presented in the previous chapters, by discussing three key notions in this dissertation: Situation Model Representations (8.1), Coherence Marking (8.2) and Forewarning (8.3). Then, I describe the implications of these results for Document Design (8.4). On the basis of these conclusions, I finally discuss limitations and directions for future research (8.5).

This dissertation aimed at investigating the effects of linguistic coherence marking on the reader and the reading process. I showed how the effect of coherence marking on comprehension depends on the reader's prior knowledge: high knowledge readers understand the text better when coherence marking is scarce (an implicit text), low knowledge readers comprehend the text better when coherence marking is used often in the text (an explicit text). For both groups of readers, an explicit text leads to a more positive opinion about the text, referred to in this dissertation as *higher appraisal*. Readers estimate their own understanding of a text as higher after reading an explicit text, referred to in this dissertation as *higher Feeling of Knowing*. For effects on *persuasion*, the reader's attitude toward the text topic, it is important to distinguish between objective and subjective marking. When a marker signals an objective causal relation, this relation exists in external reality, for instance a CAUSE-CONSEQUENCE relation. When a marker marks a subjective causal relation, this relation is constructed by the speaker or writer, for instance a CLAIM-ARGUMENT relation. Objective marking can have a positive effect on persuasion, whereas subjective marking can have a negative effect on persuasion.

In this chapter, I discuss these results in more detail by highlighting key concepts from this dissertation: coherence marking (8.2) and forewarning (8.3). Before doing so, it is important to consider how the empirical results were obtained. Operationalizations of both the situation model representation and the variable prior knowledge (8.1) have been investigated in this dissertation, since they are essential for any investigation of the role of coherence marking in a text. That is why I want to start this final chapter with discussing these issues of validation.

8.1 Situation model representations and prior knowledge

Situation model representations are a crucial level of text comprehension. This is the level of interest we need to address when we want to know whether or not readers have really understood the information in a text. It is at the level of the situation model representation that we learn whether or not the communicational purposes of a text, such as conveying information or convincing the reader, have been realized. The various experiments in this dissertation have confirmed the importance of the situation model

representations. Two aspects have been essential for my approach: 1) the scope and 2) the operationalization of the situation model representation.

8.1.1 Scope of the situation model representation

Concerning the first question, the scope of situation model representations, there has traditionally been a strong focus on comprehension alone. Although this is undoubtedly an important factor in communication, there are clearly more effects that a text can have on a reader than simply conveying the textual information. For one, texts are likely to influence opinions as well. As an example, I presented in Chapter 2 a text on the history of the Eiffel Tower, describing that it was originally considered to be hideous. Although the main goal of this text was to convey historical information, it is very unlikely that such a text does *not* influence people's opinions. Even without ever having seen the Eiffel Tower, people necessarily form an opinion. This intuitive assumption makes sense in the light of Kintsch's (1998) elaborations on how situation model representations are constructed. Kintsch claimed that readers make extensive use of their prior experiences, their background knowledge and their opinions when they construct a situation model representation of a text. Therefore, I argued in Chapter 2 that in order to get a maximally valid picture of the situation model representation, more communicative effects should be investigated than comprehension alone. I proposed to include *appraisal* (the reader's attitude concerning the text), *meta-cognition* (the reader's own assessment about the cognitive processes involved in processing the text), and *persuasion* (the reader's attitude toward the text topic). By doing so, the situation model representation consists of cognitive (comprehension, appraisal and persuasion) and meta-cognitive aspects.

8.1.2 Situation model representation methodology

The second question concerning situation model representations is a methodological one: how can we optimally assess this level of representation? Methods to assess appraisal, meta-cognition and persuasion were already well-developed and tested in previous research. However, it is not easy to simply select a standard method for measuring text comprehension from previous research, because many different methods are used. Recently developed methods such as the sorting task and the mental model task seem promising methods to assess the richness and complexity of situation model

representations, but it seems worthwhile to validate these methods. This led to the following research question, presented in Chapter 2:

RQ: What methodology is most suitable for the assessment of situation model representations?

The validity of situation model representation methodology was investigated in Chapter 4. A large experiment was conducted on the validity of several methods that are generally assumed to assess situation model representations: bridging inference questions, sorting tasks and mental model task. Also, cloze tasks were included to provide a text base measurement to which we compare the other methods. The results showed that the sorting task was the most valid method to operationalize situation model representations: it scored well on all of our criteria, such as internal reliability, convergent validity, divergent validity and known-group validity. All other methods posed problems on one or more of our criteria. The sorting task therefore seems to be more valid to assess situation model representations than the other methods included in the experiment.

Aside from text comprehension, the influence of reader characteristics on text comprehension also needed validation research. When readers construct situation model representations, they use their prior knowledge. Readers with more prior knowledge are expected to process a text differently than readers who lack this prior knowledge. It is therefore important to be able to distinguish between these two groups of readers. Globally, researchers have two different options: creating two groups by providing the necessary background (an instruction paradigm), or selecting participants in such a way that they are expected to differ in the amount of background knowledge that they have (an expert-novice paradigm). In the first experiments in this dissertation, an expert-novice paradigm was used to operationalize prior knowledge. When comparing the results from these experiments, it became clear that, although the novices always behaved in the same way, the experts' reading behavior differed considerably from one experiment to another. An instruction paradigm seemed to have solved this problem (in Chapter 5): by giving half of the participants the necessary background knowledge by means of a movie clip, they became 'experts'. Contrary to the previous experiments,

we had optimal experimental control over the exact knowledge that readers possessed.

The conclusion that the instruction paradigm is the most successful operationalization of prior knowledge is contrary to our intuitive expectation: one would expect the expert-novice paradigm to be more reliable, since it relies on extensive knowledge structures that have been established over time. Our conclusion is the more remarkable because it is contrary to earlier findings by McNamara and Kintsch (1996), where instruction did not have the expected effects²⁰. How can operationalizations that simply provide instruction and knowledge during 30 minutes be more reliable? The most plausible answer lies in the activation of relevant concepts (Anderson, 1984). The movie clips did not *literally* provide the same information as the text. They presented analogies and parallel situations to the ones in the text. For instance, when the text explained genetic manipulation in plants and food, the movie clip explained the process of human genetic manipulation. By providing equivalent but not identical information, it is plausible that the relevant concepts were activated in the reader's knowledge structure. Activation may play a lesser role when an expert-novice paradigm is used: then, the researcher relies on the expert's elaborated knowledge structures that novices do not have. This may be the reason why the operationalization of prior knowledge was most reliable when an instruction with a movie clip was given: the relevant concepts were activated in the reader's knowledge base. Clearly, further research is needed to confirm this tentative explanation.

In summary, the experiments in this dissertation have shown the sorting task to be the most valid situation model representation task, and tentatively an instruction paradigm the most valid method to operationalize prior knowledge in experiments on text comprehension.

²⁰ There is one very important difference between the instruction of McNamara and Kintsch (1996) and the one we used; the medium. McNamara and Kintsch used a pretraining booklet before having participants read the text, whereas we used a movie clip. According to the modality principle (Moreno & Mayer, 1999), the use of audio or video to explain a textual presentation enhances understanding. In our experiments, we used video to provide knowledge about the subsequent texts: video and textual information. The movie clip did in fact create a learning effect, whereas the pretraining booklet does not seem to cause such an effect.

8.2 Coherence marking

Now that we have defined the scope of situation model representations and established how these representations can optimally be assessed, we can go back to the original focus of this dissertation: the effect of a textual characteristic, coherence marking, on the reader.

8.2.1 Definition of coherence marking

Coherence marking signals relations between discourse segments. It provides the reader with processing instructions on how to integrate incoming information with previously processed information and on what inferences are needed in order to understand the text fully. Examples of coherence markers are connectives such as *because*, lexical cue phrases such as *the reason for this is* and organizer phrases such as *There are two causes for this problem*.

In this dissertation, only purely linguistic coherence marking was investigated (see Chapter 1). This means that markers that add any additional content to the text, such as elaboration or an added example, were not considered linguistic coherence markers. In previous research, these additions to the text have often been considered coherence markers, but here, they fall outside of the scope of purely linguistic coherence marking.

The central question in this dissertation was: How does linguistic coherence marking influence the reader's mental representation? Not only comprehension was investigated, but also appraisal, meta-cognition and persuasion (see 8.1). Also, the question was addressed for two text genres, informative and persuasive, and two types of readers, low knowledge and high knowledge readers. This led to the following research questions:

RQ: In what way does linguistic coherence marking influence the situation model representation of a text?

RQ: Does the reader characteristic prior knowledge influence the effect of linguistic coherence marking on the situation model representation?

These questions were investigated in Chapters 3 and 5. Three experiments were conducted in order to establish the exact nature of the effects of coherence marking and the interacting reader characteristic prior knowledge. We can conclude that when both comprehension and prior knowledge are operationalized in a valid and reliable way, there clearly is an interaction effect

of coherence marking and prior knowledge on comprehension in the informative text genre. For high knowledge readers, the text without markers causes better understanding of the text. For low knowledge readers, the text with markers improves text comprehension. This confirms the findings by McNamara, Kintsch, Songer and Kintsch (1996) and McNamara and Kintsch (1996). The theoretical explanation for this interaction effect is that readers, who do not possess the necessary knowledge, cannot make the causal inferences needed to understand the implicit text fully. They lack the background knowledge to do so. The coherence markers are needed in order to make the inferences. For readers who do possess this necessary background knowledge, coherence markers are no longer necessary: readers are capable of making the necessary inferences themselves because of their knowledge. In fact, the data have shown that they even benefit from the absence of coherence marking. This is assumed to be caused by a different processing strategy. When markers are absent, high knowledge readers have to work harder in order to understand the text: they have to make the inferences themselves. This leads to deeper processing and results in better comprehension. In the following two flow charts (Figures 8.1 and 8.2), these different effects of coherence marking on comprehension for low and high knowledge readers are represented. These flow charts represent *global* processing of a text, not the *local* processing of one relation in the text that can be marked or not.

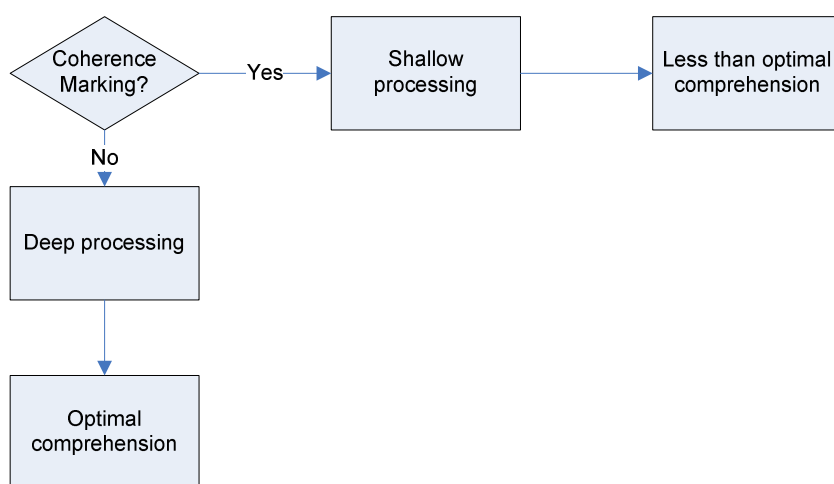


Figure 8.1: *Schematic Model of the Role of Coherence Marking in Comprehension for High Knowledge Readers*

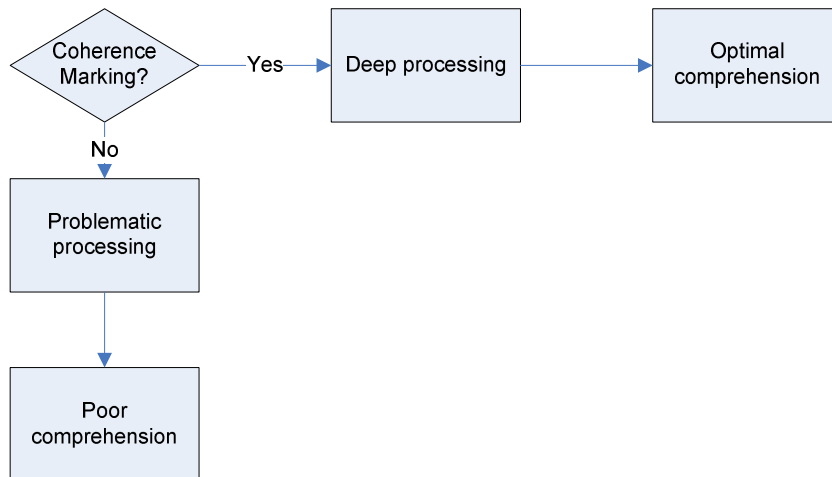


Figure 8.2: *Schematic Model of the Role of Coherence Marking in Comprehension for Low Knowledge Readers*

As I mentioned above, I proposed in this dissertation to extend the scope of situation model representations. In the experiments in this dissertation, appraisal and meta-cognition were also assessed. Appraisal concerns the reader's opinion or attitude about the text, the central question being: is it a good text or not? In Chapter 2, I explained what dimensions of appraisal are relevant in this context. Appealingness (how appealing is the text?) and accessibility (how easy is it to process the text?) are the dimensions of appraisal that I expected to be affected by coherence marking in a text. Meta-cognition was operationalized by means of *Feeling of Knowing*, which concerns the reader's impression of having understood the text. This impression can be justified (readers think they have understood the text and they in fact have) or it can be an illusion (readers think they have understood the text but in fact they have not, or vice versa). This led to the following research questions:

RQ: What are the effects of coherence marking on text appraisal? More specifically, what are the effects of coherence marking on appraisal dimensions as appealingness and accessibility?

RQ: What is the effect of coherence marking on the Feeling of Knowing that readers report after having read the text?

CONCLUSION AND DISCUSSION- 179

Effects of coherence marking on appraisal and Feeling of Knowing are clear: coherence marking improves appraisal and Feeling of Knowing. In the experiments in Chapters 3 and 5, participants find texts with marking to be easier, clearer and more pleasant to read. Also, coherence marking leads to a higher Feeling of Knowing. This impression is not an illusion: readers seem overall justified in their Feeling of Knowing: They do indeed perform adequately on comprehension tasks.

I also proposed to include persuasion in the situation model representation. Even for informative texts, such as the Eiffel Tower example, attitudes are formed or changed and attitude change could very well occur. For texts that are written explicitly with persuasion in mind, chances of such an attitude change are even higher. Assuming that coherence markers act at the situation model representation level, they are also expected to have the potential to cause attitude change.

RQ: What effect does coherence marking have on persuasion?

In the Pilot Experiment in Chapter 3, it became clear that coherence markers can influence persuasion. I only found effects in the informative genre, where the explicit version was more persuasive for low knowledge readers and the implicit version was more persuasive for high knowledge readers. Effects in the persuasive genre did not occur, leading us to look further for explanations. In Chapters 6 and 7, we have seen that it is necessary to make the distinction between marking of objective relations and marking of subjective relations. The former can improve persuasion, whereas the latter can cause resistance. I have called this a *forewarning* mechanism, which I will address in more detail in 8.3.

RQ: Are there differences concerning the effects of coherence marking in informative texts and in persuasive texts?

Aside from including appraisal and meta-cognition when studying coherence marking, another innovative aspect in this dissertation is the fact that I included two different genres, informative and persuasive, to check whether the knowledge that we have on coherence marking and its role could be generalized over text genre. To start with, there were no clear expectations concerning the role of coherence marking in both genres: the assumption was that coherence

marking would roughly have the same effects, whether it occurred in an informative or a persuasive context. Table 8.1 provides an overview of the effects in both genres.

In some cases, the observed effects of coherence marking were similar in both genres; for instance, for appraisal and Feeling of Knowing. In both genres, coherence marking leads to a more positive appraisal of the text and a higher Feeling of Knowing. In other cases, the observed effects differ from one genre to another. There are two such contrasts in particular in Table 8.1 that I would like to address in more detail.

	Comprehension	Appraisal and FOK	Persuasion
Informative genre	For readers with more prior knowledge, better performance after implicit text. For readers with less prior knowledge, better comprehension after explicit text.	Marking causes more positive appraisal and higher FOK.	Explicit texts are more persuasive for low knowledge readers, implicit texts are more persuasive for high knowledge readers.
Persuasive genre	For all readers, better comprehension after explicit version.	Marking causes more positive appraisal and higher FOK.	Subjective marking causes forewarning, objective marking does not.

Table 8.1 *Overview of effects of coherence marking and prior knowledge on comprehension, appraisal and persuasion per genre*

The first contrast becomes visible if we look at the column that presents the effects of coherence marking on comprehension in both genres. The effects of coherence marking on comprehension depend on the *genre* in which the coherence markers occur. In the informative texts, an interaction effect with prior knowledge was observed, whereas in the persuasive texts, there was no interaction but a main effect of coherence marking. In the General Discussion in chapter 5 (5.4), I already addressed the issue of how coherence marking can have different effects on comprehension depending on the genre. The analyses showed that the different effects on comprehension depending on genre are mainly caused by the high knowledge readers, who seem to ‘use’ coherence markers differently in an informative context than in a persuasive context. I speculated that this difference can be explained by Cognitive Load Theory (for

CONCLUSION AND DISCUSSION- 181

instance Paas, Renkl, & Sweller, 2003). In short, this theory states that optimum learning occurs when the load on working memory is kept to a minimum. Such a theory would predict that readers allocate their resources differently, depending on the genre. A comparable effect has been observed before. Readers have been shown to allocate their attention differently depending on the text genre, for example by Zwaan (1994). He manipulated genre by presenting all participants with an identical text, but giving half of the participants the instruction that they would be reading literary texts, while the other half was told to expect news paper articles. The results showed that readers who expected newspaper articles allocated their attention differently than the readers who expected the literary texts. Newspaper expectations led to a focus on the situation model representation, whereas literacy expectations led to a focus on text base characteristics.

In the experiments in this dissertation, attention of the high knowledge readers might also have been allocated differently in the two genres, in this case informative and persuasive texts. Here is a speculative view on what is happening. High knowledge readers have more cognitive energy available than low knowledge readers. It is simply easier for high knowledge readers to integrate the incoming information with what they already know, and to make the necessary inferences. In informative texts, they use their 'extra' cognitive energy for optimal comprehension. They do so, because they know that comprehension is the main goal of informative text. In order to benefit from an implicit text, high knowledge readers need to work hard. This is very likely to have been the case in the experiments in this dissertation, because high knowledge participants report negative appraisal of the implicit text, although they perform better on comprehension after an implicit text than after an explicit text.

In persuasive texts, the high knowledge readers use their extra cognitive energy for persuasion, because they know that persuasive texts try to change their opinions. This means that they spend energy comparing the point of view of the text to their own, weighing the arguments, and so on. In the implicit informative texts, high knowledge readers seem to allocate their attention to deep processing and 'making up for' the lack of coherence, which results in high comprehension. In the implicit persuasive texts, the high knowledge readers no longer allocate their attention to comprehension, but to persuasion. Possibly, they spend their extra available energy on deciding what

their opinion is on the issue, what arguments they have for their point of view, comparing them to the arguments put forward in the text, checking whether or not they agree with this, etcetera.

What role does coherence marking play in this allocation of attention? Coherence marking is known to make processing easier, thereby making processing resources available for other communicative effects of a text than just comprehension. When no coherence markers are present, all energy needs to be used in order to construct coherent meaning from the text. When coherence markers are present, processing is easier and extra resources become available for persuasion. Even though we do not have on-line data to support this idea, this might be why we find an interaction effect in the informative texts and a main effect in the persuasive texts: the crucial factor would be allocation of processing resources. Such a tentative explanation as presented here is in urgent need of on-line investigation.

A second contrast in Table 8.1 that I would like to emphasize is the fact that the effects of coherence marking on persuasion depend on the *type of coherence relation* (either a subjective or an objective relation). Many theories provided indications for the fact that different types of coherence relations and their marking could have different effects, but, to my knowledge, this had never been demonstrated empirically before. Evidence from categorizations (Sanders, Spooren, & Noordman, 1992) and processing (Noordman & de Blijzer, 2000) had already been presented, but with the experimental results in this dissertation, the categorization of coherence relations is shown to be not only cognitively plausible, but has also been established empirically: this distinction exists in psychological reality.

Two types of coherence relations were marked in Chapter 7: objective and subjective relations. When an objective relation is marked, this relation exists in external reality, for instance a CAUSE- CONSEQUENCE relation. When a subjective relation is marked, this relation is constructed by the speaker or writer, for instance a CLAIM-ARGUMENT relation. In the experiments in this dissertation, I showed that objective and subjective marking does not only influence persuasion differently, but also processing. In Chapter 7, on-line reading data showed that an objective marker causes the reader to speed up for the whole immediately following sentence, compared to the implicit objective relation. After a subjective marker, compared to the implicit subjective relation, readers initially speed up in the first segment of three words, but at the end of

the sentence during the last three words, they slow down. This looks like a typical inference effect: this is the point where readers are expected to make a causal inference (Cozijn, 2000; Noordman & Vonk, 1997). The verification statements showed that in both the objective and subjective condition, the causal inference had been made on-line. However, there was a slowing down effect in the reading times for the subjective condition only. In this dissertation, I have proposed that this is an effect of *forewarning*. In section 8.3, I address this effect in more detail.

8.2.2 Implications for coherence marking theory

The central focus in this dissertation was on linguistic coherence marking and its effects on the reader. How do the results in this dissertation add to existing coherence marking theory? In general, this dissertation has shown that coherence marking clearly influences the discourse representation that a reader constructs of a text. The assumption that coherence markers can be taken as processing instructions (Chapter 1 and 6) seems therefore accurate.

More precisely, we can draw four important conclusions about the way in which coherence marking can influence the representation that a reader constructs. A first conclusion is that we have seen that coherence marking acts at the level of situation model representations. Coherence marking has been shown to interact with prior knowledge on the text topic that readers may or may not possess. This interaction between prior knowledge and text information occurs at the level of the situation model representations.

A second conclusion is that the research in this dissertation has shown that the effect of coherence marking not only interacts with the reader characteristic prior knowledge, but also with genre. In the informative genre, coherence marking interacted with prior knowledge. In the persuasive genre, there was a main effect of coherence marking. In this final chapter, I have suggested that this difference may be caused by a different allocation of attention in the two different genres.

A third conclusion is that coherence marking not only affects text comprehension, but also appraisal, meta-cognition and persuasion. For such a 'small' and relatively subtle textual characteristic, it is an important finding that it can affect cognitive and meta-cognitive reactions that readers may have to a text.

A fourth conclusion concerns the difference between objective and subjective marking. This difference is even smaller and more subtle than the difference between an implicit and an explicit text. And even this smaller and more subtle textual characteristic influences persuasion, as we will see in more detail in the following section.

8.3 Forewarning

Forewarning theory is a social psychological theory that predicts that when people recognize an attempt to influence them, they build up resistance and they become difficult to persuade. So far, no study has examined text linguistic factors that could play a role in the forewarning mechanism. In typical forewarning experiments, participants were literally told before they would hear or read a message that ‘this message will try to persuade you’. Forewarning theory was more concerned with describing the psychological reaction that people may have when they become aware of the attempt to influence them, than by precisely describing factors that cause such a reaction. Of course, coherence marking is much more subtle as an indicator that an attempt is made to influence the reader. The experiments in Chapter 7 have provided evidence for subjective marking to cause resistance, whereas objective marking does not (see also Table 8.1). This forewarning mechanism is reflected in scores on persuasion (objectively marked text is more persuasive than subjectively marked text), in scores on author intent (the author’s goal is perceived to be more persuasive after having read a subjectively marked text than after the objectively marked text) and in the correlation between these effects (Exp 5). In an on-line experiment (Exp 6), we have seen how an inference is made at the end of the sentence with the subjective marker, and how this inference seems to concern the author’s intent. This on-line experiment also showed that the reader characteristic involvement seems to play a crucial role in the forewarning mechanism.

A very important step forward in forewarning theory is that we now know that specific text characteristics can cause such a psychological reaction. In this dissertation, I have proposed four processing steps of the forewarning process. Step 1, Detection of Subjectivity, means that readers have to detect the subjectivity in order for forewarning to occur. In the current context, readers needed to detect subjectivity from the subjective markers, not from the objective markers. Step 2, Recognizing and Identifying the Author’s Intent,

means that readers recognize that the author of a text has a clear goal, and subsequently readers ask themselves what this goal might be. In the experiments, we have seen that an author's inference was only made when there was a strong context for it, in this case subjectivity. These author intent inferences cost time and are reflected in the reading time data. Also, there seems to be a default expectation that readers have: when there are no clues to the author's intent, they seem to assume that the intent is to inform. This can be concluded from the comparison between the reading time data and the author's intent latencies in Chapter 7: the author's intent inference is only made when a specific clue is present. Step 3 that I proposed in the forewarning process, the actual Reaction that readers have to this Identified Author's Intent, has not been confirmed yet. Possible reactions are counter-argumentation and building up resistance. These reactions do not have to be consciously experienced. They have not been measured in the experiments in this dissertation in order to keep the persuasion process as natural as possible. Step 4, Maintaining the Original Attitude and Making it Stronger, has partly been confirmed. The subjective condition was less persuasive than the objective condition. Does this mean that in the subjective condition, readers stick to their initial attitude? In the experiments in this dissertation, I never measured initial attitudes, only post-text attitudes. When investigating persuasive effects of a text, it is tempting to simply measure the attitude toward a certain topic before reading a text on that topic, and measuring that same attitude again after participants have read the text. However, this method is in itself a forewarning mechanism. Once readers have consciously formulated and reported on their attitudes, they become resistant to attitude change (Hoeken, 1995). Therefore, in all the experiments in this dissertation, only post-text attitude measurements were collected. Participants were assigned randomly to one condition or the other. If differences between conditions occur, they have to be caused by the text. Step 4, maintaining the *original* attitude, is difficult to confirm if we take it literally, because the initial attitude was not measured. But we can say with certainty that subjective marking has negative effects on persuasion. Also, the data have shown that Step 4 seems to differ for involved and uninvolved readers. Involved readers benefit from objective marking: this version results in the most favorable attitude change. Uninvolved readers do not process the information fully. For them, objective markers do not have a positive effect.

However, the subjective markers seem to have a negative effect for uninvolved readers: they cause resistance and inhibit persuasion.

The model for the forewarning process, proposed in Chapter 6 and tested empirically in Chapter 7, is represented in the two following tentative processing models of forewarning (see Figure 8.3 below and 8.4 on the next page). Two different models were constructed, because of the apparent interference of the reader characteristic involvement in the process of forewarning. The process of forewarning is elicited by a signal of subjectivity, in this case the coherence marker. This signal triggers an author's intent inference. This inference is made by both involved and uninvolved readers. However, they react differently: for involved readers, this does not cause forewarning to occur, whereas for uninvolved readers, it seems to be the case that they *do* experience forewarning. A plausible explanation would be that according to the Elaboration Likelihood Model (Petty & Cacioppo, 1981), involved readers want to make sure that they hold the correct attitude. They consequently process deeper and weigh advantages and disadvantages carefully (the *central route*). In this case, forewarning does not occur. Whether or not involved readers detect subjectivity, they still process deeply and weigh advantages and disadvantages carefully. When readers are not involved in the text topic, they do not feel the need to be sure that they hold the correct attitude. This results in shallow processing: not weighing the advantages and disadvantages carefully, but finding a fast way out (the *peripheral route*). Here, forewarning may occur: detection of subjectivity can be interpreted by the reader as a signal to ignore the rest of the text, because clearly someone is trying to persuade the reader. I propose two different flow charts to illustrate the steps in the forewarning process: Figure 8.3 for high involvement readers and Figure 8.4 for low involvement readers, in which I have included the expectations about processing routes (ELM).

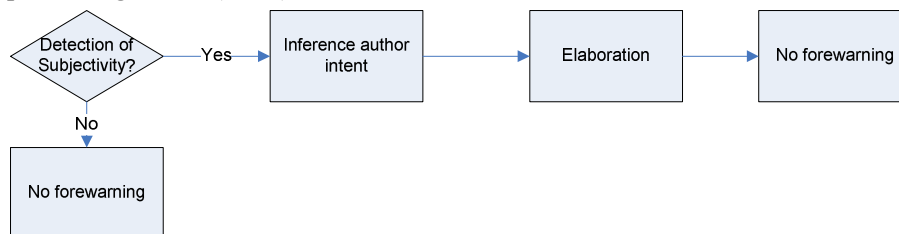


Figure 8.3 *Tentative Schematic Model of Steps in the Forewarning Process, for Highly Involved Readers*

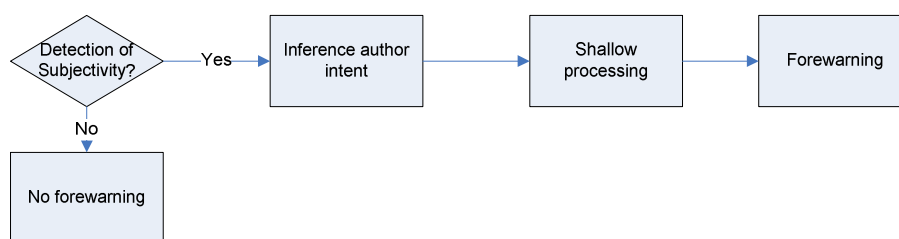


Figure 8.4 *Tentative Schematic Model of Steps in the Forewarning Process, for Low Involved Readers*

These two models are both based on the four steps of processing that I outlined in Chapter 6: Step 1 is the Detection of Subjectivity, Step 2 is Recognizing and Identifying the Author's Intent, Step 3 concerned the psychological reaction of counter-argumentation and elaboration, and Step 4 concerned the maintaining of the original attitude. In Figure 8.3, the processing steps for the involved readers, Steps 1, 2 and 3 are present. But because of the elaboration and deep processing that the ELM would predict, forewarning seems to be overruled. Involved readers want to be certain that they hold the correct attitude. Therefore, they are not influenced by the signal of subjectivity as much as the low involved readers (Figure 8.4). In Figure 8.4 for low involvement readers, steps 1, 2 and 4 are present. Whether or not step 3 (the psychological counter-argumentation) occurs has not been demonstrated in this dissertation.

In sum, the experiments in this dissertation have provided evidence for the forewarning effect. Also, a first step was made in describing the steps in the forewarning process. However, the picture of the forewarning process is far from complete. Some of our insights were developed in a post-hoc study and therefore not obtained with a careful manipulation of involvement. In section 8.5, directions for future research are discussed and I propose to use experimental techniques that may help to find out more about the forewarning effect.

This dissertation takes forewarning theory in a new direction by not just focusing on psychological implications, but by combining the psychological reaction with the possible linguistic *cause* for the phenomenon to occur. For social psychologists, the consequences of the forewarning reaction are probably more interesting than its cause. That is why, in these studies, the causes are explicit, literal messages of persuasive intent. For linguists, the situation is

reversed; *the reason why* it happens might be more interesting. Which textual feature causes forewarning and which one does not? I believe that only by observing exactly under what circumstances the forewarning mechanism occurs, we can fully understand how textual features can cause this phenomenon. In this dissertation, a beginning was made to do so.

8.4 Implications for Document Design

The research reported in this dissertation does not only have implications for methodology and theory in cognitive linguistics and psychology, as we have seen in the previous sections, but also in the field of document design. The experiments have clearly shown that coherence marking affects text effectiveness. This important text characteristic influences comprehension, appraisal, feeling of knowing and persuasion. In this paragraph, these implications are discussed and illustrated with very concrete examples. What is the best solution in each situation, to add coherence markers or to leave them out?

8.4.1 Educational context

One advice for educational document design might be summarized as: ‘Estimate what the reader knows’. We have seen that the effects of coherence marking differ from low knowledge readers to high knowledge readers. Low knowledge readers need the coherence markers in order to construct the best possible situation model representation. High knowledge readers, on the other hand, seem to perform better after the deep processing that occurs with the reading of the implicit version. Depending on whether or not an educational text is explaining something new, or something that the pupils already have knowledge about, the advice on coherence marking differs. The main goal of educational texts is to teach pupils new information. Therefore, I would say that we can assume that they will mostly behave like low knowledge readers, and thus benefit from coherence marking in their schoolbook texts.

Suppose that we are writing a French History book and we need to describe the life and reign of Louis XIV. A good option would be the following text passage²¹:

²¹ Adapted and translated from Rivière, D. (1995) *Histoire de la France*. Paris, Hachette Education. Chapter 12 deals with the reign of Louis XIV. I added the coherence marking.

CONCLUSION AND DISCUSSION- 189

From 1661, when Louis XIV was 22 years old, he was the King of France. Louis's reign can be characterized by the remark attributed to him, '*L'état, c'est moi*' [I am the state]. Also, Louis wanted to be called '*Le Roi Soleil*' [the Sun King]. One of the reasons for this was that he considered the King to be the lieutenant of God on earth. Louis XIV eliminated the position of 'principal minister', so the King became the only ruler of the country. His main goal was to pursue greatness for the French country. Consequently, France became a reformed country, but the economy suffered greatly. Louis XIV had passionate relationships with women. This is why he had various secret galleries built near his apartment to allow his many courtesans to reach him. The result was that he had 6 children with his wife, Queen Marie-Thérèse, and 12 more bastard children with his mistresses. Therefore, the court sometimes almost resembled a harem.

The underlined coherence markers might, as we saw in Chapters 3 and 5, be beneficial for text comprehension. Causal connectives such as *therefore*, *consequently* and *but*, and lexical cue phrases such as *the result was that* and *one of the reasons for this was*. Unless the expected reader has much prior knowledge about Louis XIV: then, the preferred option would be to remove the coherence markers and let the readers establish coherence for themselves. However, in the context of a school book, one can expect the pupils to be expected to *learn* something they did not know beforehand, making them low knowledge readers. In this case, I would expect them all to be more or less unfamiliar with French History.

The advice based on findings in this dissertation would be to put the coherence markers in and to thereby help students where possible. In general, this will make it easier for them to understand the text and learn from it, they will then appraise the text more positively. However, there are limits. Probably, marking *every* relation with a coherence marker would irritate the reader more than anything else. Like all good things in life, coherence markers are to be used in moderation.

The same advice appears to hold for other informative contexts, such as newspaper articles. Most people are low knowledge readers, since the experiments have shown that *exactly* the right knowledge structures are necessary in order to benefit from an absence of coherence marking. Again, the advice would be: get to know the reader, but if you do not know exactly what they know, the safe option is to use coherence markers in the text, especially on crucial places in your text structure.

8.4.2 *Marketing context*

Implications for persuasive document design could be summarized in the following adage: ‘Hide your intentions when you want to persuade!’

It became clear from the research in this dissertation that people were more likely to accept information as true when they perceived it to be neutral and objective. This is not a new finding; Petty and Cacioppo already showed a similar effect in 1979. However, the fact that such a subtle manipulation as a coherence marker can have this effect *is* a new finding. Petty and Cacioppo’s manipulations are not so subtle: they announce that an attempt will be made to influence people’s opinions and as a result, people are less likely to be persuaded. In this dissertation, another step is added: a subtle linguistic manipulation in the text itself leads to the recognition of the persuasive intent and then again, people are less likely to be persuaded.

Suppose that we are writing a brochure about Versailles Castle, the castle that Louis XIV ordered to be expanded²². The leaflet aims at attracting as many visitors as possible: a true marketing goal.

The Palace and Gardens of Versailles are worth a visit, because they count among the most prestigious of the world’s heritage sites. ~~Another reason to visit Versailles is that~~ it represents the finest and most accomplished achievement of 17th century French art. The Castle of Versailles was transformed and enlarged by Louis XIV, ~~thereby~~ offering France the most prestigious symbol of all his power. Plan your day out discovering all the splendours of Versailles thanks to the range of possibilities available. You can visit all the different parts of the Palace freely. Therefore, remember to visit Castle Versailles!

The original version of this leaflet did not contain the coherence markers. This is a good strategy: the markers added here mark subjective relations and this decreases persuasiveness, as we have seen in Chapter 7. Objective markers can be used without any negative effects on persuasion: there might even be positive effects. But for this positive effect, readers would have to be involved and motivated to read this leaflet about Versailles.

²² Based on the original English leaflet from Versailles Castle, www.chateauversailles.fr and on information from www.francemonthly.com. Coherence markers were added by me.

8.5 Discussion

In this paragraph, I would like to go back to the scope of situation model representations, by asking: how far does the scope of the situation model representation reach (8.5.1)? Then, the limitations of the work in this dissertation (8.5.2) and possible directions for future research are presented (8.5.3).

8.5.1 Implications for situation model theory

The term situation model representation has been reserved to indicate the deepest and richest level of mental representation. At what level of mental representation do factors such as appraisal, meta-cognition and persuasion play a role? It makes sense to assume that if we want to include them in the mental representation, this has to be at the richest and deepest level of representation. Clearly, these factors do not act at the level of the text base representation or the surface code, but at the situation model representation level.

Does this imply that *every* possible relevant factor needs to be integrated in the situation model representation? Where are the boundaries of this theoretical construct? How many factors need to be integrated in situation model representations to explain for the liveliness and detail that we find in our situation model representations of books we read and enjoy? When we read, we construct a very rich world inside our heads. Obviously, this is not done by simply integrating textual information and prior knowledge. Therefore, I argued that we need to consider prior experiences and opinions as well. But even if we do so, does that account for all the information in such a rich mental representation as the situation model representation? Or are there more factors that need to be included? My position is that the situation model representation is so rich and vivid, that it may very well include aspects that we have been overlooking so far. Here, I would like to address one particular debate about the nature of situation model representations. Specifically, the question is: is situation model theory adequate to account for the richness of our representation?

Recently, there has been a lively debate that opposes the ‘traditional’ situation model theory to the view of embodied cognition (EC). The main claim of EC is that the contents and operation of the mind are grounded in an agent’s physical characteristics and embodied experience. This can also be applied to language in general, or reading in particular. Zwaan (1999a) gives the

following examples: while reading, people can almost feel ‘cold wind blowing in our face, the smell of beer, a kiss on our lips, and a hot piece of pizza sticking to the roof of our mouths’ (p.83). Zwaan argues that although it is not unreasonable to hold situation model theory accountable for this, EC would be a better explanation.

The following example taken from Zwaan and Taylor (2006) explains the difference between the situation model and the embodied view. Suppose a reader comes across the sentence ‘Eric turned down the volume’. Classical situation model theory predicts that this sentence activates the relevant propositions (text base) and then, the reader integrates the information from the text with prior knowledge in a network (situation model representation). The embodied cognition view predicts something completely different: an activation of a motor program for counter clockwise manual rotation in the reader. This reaction can be explained by the phenomenon of motor resonance: simply put, when people observe someone else perform an action, their brain reacts as if they were performing that action themselves. Psychologists such as Zwaan, Glenberg, Anderson and others have indeed provided evidence for such claims. In contrast, situation model theorists claim that the richness of situation models can also be explained by a larger number of propositions that are all part of the situation model representation.

Can we apply this idea of EC to the texts discussed in this dissertation? One of the reactions that readers experience in reading persuasive texts is resistance. In the experiments in this dissertation, this reaction has been referred to as *forewarning*. We might argue that experiencing resistance in persuasion, sticking to one’s own beliefs, could very well be analyzed as an embodied reaction. The metaphor that is often used to describe this reaction is ‘to dig in one’s heels’, or in Dutch ‘de hakken in het zand zetten’. These metaphors describe a *mental* form of resistance by referring to *physical* acts. Lakoff and Johnson (1980) suggested that metaphors not only make our thoughts more vivid and interesting, but that they actually structure our perceptions and understanding. Judging from these metaphors, we might argue that resisting an attempt to be influenced could resemble an almost physical experience of resisting. Therefore, such a reaction could be represented in an embodied experience.

Whether or not situation model representations include sensor motor information is beyond the scope of this dissertation. We have seen that it

definitely includes aspects such as people's opinions and their experiences. They all play a role at the formation of the situation model representation and therefore, it makes sense that they are part of such a representation. But how exactly these aspects are represented in the brain, whether this is by means of propositions, networks, or by means of motor resonance, is another issue.

My preliminary conclusion here is that the boundaries of the situation model representation remain to be investigated. However, I do want to stress that the traditional view of situation model representations, including the network between propositions from a text and prior knowledge, is definitely too simplistic and does not account for the richness of situation model representations. By including factors such as appraisal, meta-cognition and persuasion, we are getting closer to the required level of text representation that we refer to as the situation model representation.

8.5.2 Limitations

An advantage of all the studies reported in this dissertation is that they include different text and reader variables, in order to form as complete a picture of coherence marking and its functioning as possible: Prior knowledge, coherence marking, comprehension, appraisal, meta-cognition and persuasion. As we have seen in the contrastive analysis in Table 8.1, it is important to include these different variables to get a valid picture of the situation. However, there is also a downside to this advantage: some variables impose certain conditions on others. For example, the fact that in Chapter 7, processing effects were investigated in a repeated measures design, led to a design in which the forewarning effect was also examined in a repeated measures design. And this might have caused problems, since asking participants to rate an author's intent 24 consecutive times probably influences their reaction: the 20th time, the question is hardly unexpected anymore. Still, I carefully checked in the data whether this actually occurred. The data do not show that this problem has influenced the results (Chapter 7); the results on the first text that readers read do not differ systematically from the results on the following texts. But even if it seems that no problems occurred, I think that on a more global level this issue certainly needs more attention: is a repeated measures design the most suitable one to test for forewarning effects? Is there a methodological way around this?

Another limitation for generalizability of the findings in this dissertation could be that there are some possible confounding variables. Some factors are simply not constant between experiments. *Language*, for one: all experiments are conducted in Dutch (1-5) except Experiment 6, which was conducted in English. I seem to find the same effects across these two languages, but this factor was not systematically accounted for. Also, although we included one very important reader characteristic, namely prior knowledge, there are indications that another factor may also be of importance: *involvement*. I will get back to this factor in the next section, since this seems to be a very important reader characteristic that was not taken into account systematically. Also, the reader characteristic that we chose to include, prior knowledge, was not manipulated consistently throughout the experiments. This was done, as I described in Chapter 5, in order to find out what operationalization of this reader variable was most successful.

8.5.3 Indications for future research

In this paragraph, I will first address the effects of coherence marking on comprehension and appraisal, and the work that still needs to be done in this area. Then, I will have a more detailed look at effects on persuasion and formulate directions for further research.

Comprehension and appraisal

The effects of coherence marking on comprehension and appraisal have shown to be quite robust. There are many studies that have focused on comprehension effects of coherence marking and this dissertation seems to provide a quite clear answer to the question whether or not coherence markers are beneficiary for comprehension. We know that reader and text characteristics interact. This dissertation added new empirical results to this the interaction between genre and text characteristics on the reader's comprehension. However, some questions remain to be answered, concerning processing and reader characteristics.

A first, very important matter is that of processing data. Effects on comprehension combined with on-line data to account for these effects are necessary in order to take the next step. This would help explain why the effects of coherence marking on comprehension differed from the informative to the persuasive genre. I suggested that this difference might be caused by a

different allocation of attention and cognitive resources. On-line data could test this explanation. Furthermore, we have seen how in previous research, some studies reported an inference effect in on-line reading time data and other studies did not. In Chapter 7, this occurred within one experiment: an author's intent inference effect was found in the subjective but not in the objective condition. We need to know more precisely what is happening in the different conditions and on-line data are needed to answer these questions.

A second matter concerns reader characteristics. Is the interaction between prior knowledge and coherence marking the only relevant interaction with reader characteristics? Or do other reader characteristics have to be taken into account as well? Even though we did not explicitly address this issue, it seems that involvement is another important factor that needs to be taken into account.

Persuasion

Concerning the effects of coherence marking on persuasion, one of the most interesting new findings in this dissertation concerns the forewarning effect. More research needs to be done on this topic. We have seen that subjective marking causes an inference to be made concerning the author's intent, and this diminishes persuasion. Although the forewarning effect was demonstrated in both experiments in chapter 7, several questions remain unanswered. The most important ones are discussed below.

Eliminating other explanations

Although strong indications for the forewarning effect were found in Chapter 7, they do not definitely eliminate other explanations. Subjective relations are known to be more complicated than objective relations. This means that subjective relations are more difficult to validate and check against the prior knowledge base (Traxler, Bybee, & Pickering, 1997). In other words, both objective and subjective relations require the making of a causal inference, but this inference is more complex in nature in subjective relations. This process of validating can also be an explanation for our finding that subjective marking slows processing down at the end of the sentence. This is the point in time where the causal inference is made (Noordman & Vonk, 1998), and this inference is more complex in the case of a subjective relation than in the case of an objective relation. Consequently, if a process is difficult, one possible

effect is that it diminishes persuasiveness. This could be an alternative explanation for our findings.

However, this alternative explanation does not explain for the correlation between the inference effect we observe and the author's intent latencies. This is the reason why I do not believe that the alternative explanation of a more complex causal inference can explain for *all* effects that were found. The difficulty of the causal inference could contribute to the effect, but it is certainly not the only reason for the slowing down effect. Rather, I am confident that the results favor the forewarning explanation. There are possible ways to strengthen this forewarning explanation.

Circumstances of the forewarning effect

One possible way to prove the forewarning hypothesis in a more definite way would be to create conditions in which it is expected to disappear and check whether this indeed influences processing data. One might think of explicit instructions. If participants are given an identical text, but half of them are told from the beginning of a text that an attempt will be made to change their opinions (compare Petty & Cacioppo, 1979) and the other half are not, I would expect the forewarning effect after a subjective marker to disappear in case of an explicit instruction. After all, readers know from the beginning of the text that they are going to be persuaded. There is no need to infer the author's intent anymore: it was explicitly mentioned. In other words, the forewarning effect would have to occur at the beginning of the text, meaning that the actual author's inference would have to be made at that point in the text, and no longer at the point of the subjective marker. Reading time patterns after subjective marking would then look differently, with no clear inference effect.

Reader characteristics

Prior knowledge and coherence marking interact on comprehension. However, prior knowledge did not have any (interacting or other) effects on persuasion, contrary to the expectations. Is this due to the fact that I did not *provide* the prior knowledge, as I advocate in Chapter 5? Is it due to an influence of yet another reader characteristic: involvement? Chapter 7 provides indications for this. Future research needs to include these reader characteristics in a more systematic way, in order to analyze the interaction between prior knowledge and coherence.

Other categories of markers

In the first experiments in this dissertation, all markers of coherence were analyzed together. The aim was to investigate their collective effects, not the effect of one single manipulation. Further on in this dissertation, this aim shifted and in the last experiment in Chapter 7, only causal markers were included, in the form of connectives and lexical cue phrases. They either signaled an objective or a subjective relation. In a more precise way, the effects of one single marker on persuasion, on verification statements and on author's intent ratings were examined. I feel that this combination of collective and precise studies has proven to be successful: it first provided us with a general picture, and the more precise measurements were needed in order to fill in the blanks. However, a logical step to take in the future would be to examine in more detail the effect of other categories of markers. How about contrastive markers such as *although* or *however*? Do they also cause forewarning and resistance to persuasion? More research is needed that includes other types of markers.

Eye-tracking

Methodology evolves rapidly. Inference effects of connectives might be better visible when using eye-tracking than with a simpler moving window paradigm (Cozijn, 2000). This might help pinpoint the exact moment in time where readers infer author's intent.

Also, eye tracking technology might enable us to answer the question of causal inference making after a coherence marker: do coherence markers cause the causal inference to be made at the end of the sentence or do they present information in such an explicit way that the inference has become redundant? It is difficult to decide from the data in this dissertation which one of these is the right explanation and I believe that the more precise the measurements, the more reliable the answers we will find.

Cognitive Load

Our (speculative) explanation for the influence of genre on comprehension effects is that of a cognitive load effect (see 8.2.1). If Cognitive Load is a correct explanation, this too has to be visible at very precise moments in time: when processing a marked relation, readers would have to have more cognitive energy available than when processing an implicit relation. An excellent way to test this is by way of a secondary task paradigm (for instance Britton, Glynn,

Meyer, & Penland, 1982; and more recently Spooren, Mulder, & Hoeken, 1998), where participants have to read texts and perform a secondary task at the same time.

Other genres

The fact that effects of coherence marking differed from one genre to another is an important clue to the importance of genre and the influence it can have on reader's behavior. Therefore, I argue that experimental studies on textual characteristics should go beyond genre boundaries and include at least two.

If the best thing writers of persuasive texts can do is to hide the persuasive intent, it might be a smarter idea to convey the information through informative texts, and let readers make up their own minds. Of course, the information present in the text can still be chosen to 'push' the reader in a certain direction. This strategy is hardly new: we have become used to so-called advertorials (Waltzer, 1988), advertisements written in the form of an objective opinion editorial, designed to look like a legitimate and independent news story. Another example is the infomercial (Haley, 1993), often made to closely resemble actual television programming with minimal acknowledgement that the program is actually an advertisement. Still, there might be other interesting genres that could be used as a 'disguise'. Persuasive text could then, as a strategic decision, very well be disguised as a narrative (Appel & Richter, accepted for publication). Would that cause less forewarning, thereby making the text more persuasive? Would coherence marking of subjective relations still have its negative effect on persuasion?

Correlation comprehension – forewarning

So far, I have investigated the forewarning mechanism independent of comprehension. But it is not unlikely that these effects of coherence marking are linked. For involved readers, the forewarning effect does not occur, maybe because of deep processing. This deep processing would also have to result in better comprehension, as we saw for the high knowledge readers: when coherence markers are absent, they work harder to 'make up for' this lack of coherence in the text. This leads to deeper processing, resulting in better comprehension. Therefore, my prediction would be that the results on comprehension and on persuasion or forewarning are linked. They both result from the deeper processing that this type of reader can do to compensate for an apparent lack of coherence in the text.

The same reasoning applies to the low knowledge readers. They need the coherence markers in order to be able to make the required inferences. This results in shallow processing, because these readers do not have the necessary knowledge to check these inferences against their knowledge base. These same readers are sensitive to subjectivity in a text: when they infer the author's intent, they processed the text in a shallow manner, resulting in the maintaining of their original attitude. Both the effects on comprehension and the effects on persuasion are linked because of the shallow processing strategy. Further research can elaborate on this, to verify my hypothesis that comprehension and persuasion are strongly correlated, and disentangle the causal nature of the relation: which factor is the cause and which one the consequence? Are they correlated because of an interacting variable, the processing strategy? Answering these questions can clarify the picture of influences of coherence marking on both comprehension and persuasion.

8.6 Epilogue

In conclusion, this dissertation combines three lines of research in different disciplines. First, it relies on research in text linguistics and discourse processing that considers coherence to be a crucial characteristic of discourse (Hobbs, 1979; Sanders & Spooren, 2001). Second, the research in this dissertation integrated insights from text linguistic studies defining objective and subjective differences between types of coherence relations and their markers (Knott & Sanders, 1998; Pander Maat & Sanders, 2001; Sanders, Spooren, & Noordman, 1992). Third, the experiments elaborate on the phenomenon of forewarning and persuasion, studied in social psychology for decades (Petty & Cacioppo, 1979).

Combining these three research traditions has proven fruitful. Such an interdisciplinary account contributes to the further development of a cognitive theory of discourse coherence, at the interface of linguistics and discourse psychology. It has helped to explain more of the complex process that takes place when people read texts. This process appears to be influenced by coherence markers; text characteristics that are often considered 'no more' than subtle text-analytical differences. Therefore, in terms of text effectiveness, choosing coherence markers requires careful consideration of the marker's subjectivity.

References

- Alderson, J. C. (1983). The cloze procedure and proficiency in English as a foreign language. In J. W. Oller (Ed.), *Issues in Language Testing Research* (pp. 205-217). Rowley, MA: Newbury House.
- Anderson, R. C. (1984). Role of the reader's schema in comprehension, learning and memory. In R. C. Anderson, J. Osborn & R. J. Tierney (Eds.), *Learning to read in American schools: Basal readers and knowledge texts*. (pp. 243-258). Hillsdale, NJ: Erlbaum.
- Anscombre, J.-C., & Ducrot, O. (1983). *L'argumentation dans la langue*. Bruxelles: Pierre Mardaga.
- Appel, M., & Richter, T. (accepted for publication). Persuasive Effects of Fictional Narratives Increase over Time. *Media Psychology*.
- Bachman, L. F. (2000). Modern language testing at the turn of the century: assuring that what we count counts. *Language Testing*, 17(1), 1-42.
- Baker, L. (1985). How do we know when we don't understand? Standards for Evaluating Text Comprehension. In D. L. Forrest-Pressley, G. E. MacKinnon & T. Gary Wallen (Eds.), *Metacognition, cognition and human performance* (pp. 155-205): Academic Press.
- Barton, E. L. (1995). Contrastive and non contrastive connectives. Metadiscourse functions in argumentation. *Written Communication*, 12(2), 219- 240.
- Ben-Anath, D. (2005). The role of connectives in text comprehension. *Working Papers in TESOL and Applied Linguistics*, 5(2), 1-27.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Birkmire, D. P. (1985). Text processing: The influence of text structure, background knowledge, and purpose. *Reading Research Quarterly*, 314-325.
- Blanc, N., & Tapiero, I. (2001). Updating Spatial Situation Models: effects of prior knowledge and task demands. *Discourse Processes*, 31(3), 241-262.
- Boogaard, M., Sanders, T., & Admiraal, W. (1996). Vullen connectieven de open plekken? Over coherentie en tekstbegriptoetsing. In P. J. M. Groot & L. J. A. Nienhuis (Eds.), *Leesvaardigheid in de eerste en in een tweede taal. Een vergelijking*. (pp. 7-20).
- Boscolo, P., & Mason, L. (2003). Topic Knowledge, Text Coherence, and Interest: How they Interact in Learning From Instructional Texts. *Journal of Experimental Education*, 71(2), 126-148.
- Britton, B. K. (1994). Understanding Expository Text. Building Mental Structures to Induce Insights. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 641-674). San Diego: Academic Press.
- Britton, B. K., Glynn, S. M., Meyer, B. J. F., & Penland, M. J. (1982). Effects of text structure on use of cognitive capacity during reading. *Journal of Educational Psychology*, 74(1), 51-61.
- Britton, B. K., & Gülgöz, S. (1991). Using Kintsch's Computational Model to Improve Instructional Text: Effects of Repairing Inference Calls on Recall and Cognitive Structures. *Journal of Educational Psychology*, 83(3), 329-345.
- Brown, P., & Levinson, S. C. (1987). *Politeness. Some universals in language usage*. (Vol. 4). Cambridge: Cambridge University Press.

REFERENCES- 201

- Brown, S. P., & Stayman, D. M. (1992). Antecedents and consequences of attitude towards the ad: a metaanalysis. *Journal of Consumer Research*, 19, 34-51.
- Chen, H. C., Reardon, R., Rea, C., & Moore, D. J. (1992). Forewarning of Content and Involvement: Consequences for Persuasion and Resistance to Persuasion. *Journal of Experimental Social Psychology*(28), 523-541.
- Connor, U., & Upton, T. (2003). Linguistic Dimensions of Direct Mail Letters. In P. Leistyna & C. F. Meyer (Eds.), *Corpus Analysis, Language Structure and Language Use* (pp. 71-86). Amsterdam: Rodopi.
- Cozijn, R. (2000). *Integration and inference in understanding causal sentences*. Tilburg University: Doctoral dissertation.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 280-302.
- de Jong, M. (1998). *Reader feedback in text design; Validity of the plus-minus method for the pretesting of public information brochures*. Amsterdam Doctoral dissertation.
- de Jong, M., & Lentz, L. (1996). Expert Judgments versus Reader Feedback: A comparison of text evaluation techniques. *Journal of Technical Writing and Communication*, 26(4), 507-519.
- de Jong, M., & Schellens, P. J. (1994). Towards a valid design for pretesting and revising leaflets. In L. v. Waes, E. Woudstra & P. v. d. Hoven (Eds.), *Functional Communication quality*. Amsterdam: Rodopi.
- Degand, L., Lefevre, N., & Bestgen, Y. (1999). The impact of connectives and anaphoric expressions on expository discourse comprehension. *Document Design*, 1(1), 39-51.
- Degand, L., & Sanders, T. (2002). The impact of relational markers on expository text comprehension in L1 and L2. *Reading and Writing*, 15(7-8), 739-757.
- Dixon, P., & Bortolussi, M. (2001). Text is Not Communication; A Challenge to a Common Assumption. *Discourse Processes*, 31(1), 1-25.
- Duke, N. K. (2005). Comprehension of What for What: Comprehension as a Nonunitary Construct. In S. G. Paris & S. Stahl (Eds.), *Children's Reading Comprehension and Assessment* (pp. 93-104). Mahwah, NJ: Lawrence Erlbaum Associates.
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth: Hartcourt Brace Jovanovich.
- Epstein, W., Glenberg, A. M., & Bradley, M. M. (1984). Coactivation and comprehension: contribution of text variables to the illusion of knowing. *Memory and Cognition*, 12(4), 355-360.
- Erickson, T. A., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, 20, 540-552.
- Ericsson, K. A., & Smith, J. (1991). Prospects and limits of the empirical study of expertise: an introduction. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise. Prospects and limits* (pp. 1-38). Cambridge: Cambridge University Press.
- Fauconnier, G. (1984). *Espaces mentaux. Aspects de la construction du sens dans les langues naturelles*. Paris: Editions de Minuit.
- Fellenz, M. R. (2004). Using assessment to support higher level learning: the multiple choice item development assignment. *Assessment and Evaluation in Higher Education*, 29(6), 703-718.

202 - REFERENCES

- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Fletcher, C. R., & Chrysler, S. T. (1990). Surface Forms, Textbases, and Situation Models: Recognition Memory for Three Types of Textual Information. *Discourse Processes, 13*, 175-190.
- Friestad, M., & Wright, P. (1994). The Persuasion Knowledge Model: How people cope with Persuasion Attempts. *Journal of Consumer Research, 21*, 1-30.
- Garbarino, E. C., & Edell, J. A. (1997). Cognitive effort, affect and choice. *Journal of Consumer Research, 24*, 147-158.
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition 68*, 1-76.
- Gilbert, R., Martínez, G., & Vidal-Abarca, E. (2005). Some good texts are always better: text revision to foster inferences of readers with high and low prior background knowledge. *Learning and Instruction, 15*, 45-68.
- Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self assessment of comprehension. *Memory and Cognition, 10*, 597-602.
- Goldstein, H. (2003). *Multilevel Statistical Models 3rd Edition*. London: Arnold.
- Gordon, C. J., & Braun, C. (1985). Metacognitive Processes: Reading and Writing Narrative Discourse. In D. L. Forrest-Pressley, G. E. MacKinnon & T. Gary Wallen (Eds.), *Metacognition, cognition and human performance* (pp. 1-6): Academic Press.
- Graesser, A. C., McNamara, D. S., & Louwerse, M. M. (2003). What Do Readers Need to Learn in Order to Process Coherence Relations in Narrative and Expository Text? . In A. P. Sweet & C. E. Snow (Eds.), *Rethinking Reading Comprehension* (pp. 82-98). New York: Guilford Publications.
- Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48*, 163-189.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing Inferences During Narrative Text Comprehension. *Psychological Review, 101*(3), 371-395.
- Haberlandt, K. (1982). Reader expectations in text comprehension. In J. F. Le Ny & W. Kintsch (Eds.), *Language and comprehension* (pp. 239-249): North-Holland Publishing Company.
- Haley, K. (1993). The infomercial begins a new era as a marketing tool for top brands. *Advertising Age, 64*.
- Halliday, M. A., & Hasan, R. (1976). *Cohesion in English*. New York: Longman.
- Heller, E., & Areni, C. S. (2004). The effects of conditional indicative language on the comprehension and acceptance of advertising claims. *Journal of Marketing Communications, 10*, 229-240.
- Hidi, S., & Baird, W. (1986). Interestingness- a neglected variable in discourse processing. *Cognitive Science, 10*, 179-194.
- Hobbs, J. R. (1979). Coherence and coreference. *Cognitive Science, 3*, 67-90.
- Hoeken, H. (1994). Evaluating persuasive texts: the problems of how and what to measure. In L. v. Waes, E. Woudstra & P. v. d. Hoven (Eds.), *Functional Communication Quality* (pp. 76-87). Amsterdam: Rodopi.
- Hoeken, H. (1995). *The design of persuasive texts: Effects of content, structure, and style on attitude formation*. Tilburg University: Doctoral dissertation.

REFERENCES- 203

- Hoeken, H. (1998). *Het ontwerp van overtuigende teksten. Wat onderzoek leert over de opzet van effectieve reclame en voorlichting*. Bussum: Coutinho.
- Hoffman, R. R., Shadbolt, N. R., Burton, A. M., & Klein, G. (1995). Eliciting Knowledge from experts: a methodological analysis. *Organizational Behavior and Human Decision Processes*, 62(2), 129-158.
- Holleman, B. (2000). *The forbid/allow asymmetry. On the cognitive mechanisms underlying wording effects in surveys*. (Vol. 16). Amsterdam: Rodopi.
- Horn, J. L. (1969). *Priifsystem für Schul-und Bildungsberatung*. Göttingen: Verlag für Psychologie Hogrefe.
- Johnston, W. A., & Heinz, S. P. (1978). The flexibility and capacity demands of attention. *Journal of Experimental Psychology: General*, 107, 420-435.
- Kamalski, J. (2004). How to Measure the Situation Model. On the operationalization of a crucial level of text representation. In A. Kerkhoff, J. de Lange & O. Sadeh Leicht (Eds.), *Yearbook Utrecht Institute of Linguistics* (pp. 121-132).
- Kamalski, J., Lentz, L., & Sanders, T. (2004). Coherentiemarkering in informerende en persuasieve teksten. Een empirisch onderzoek naar cognitieve en affectieve effecten. *Tijdschrift voor Taalbeheersing*, 26(2), 85-104.
- Kamalski, J., Lentz, L., & Sanders, T. (2006, July 26-29). *Effects of Coherence Marking on the Comprehension and Appraisal of Discourse*. Paper presented at the Cognitive Science Conference, Vancouver, Canada.
- Kamalski, J., Lentz, L., & Sanders, T. (submitted). Coherence marking, prior knowledge and comprehension of informative and persuasive texts: Sorting things out.
- Kamalski, J., Lentz, L., Sanders, T., & Zwaan, R. A. (submitted). The fore-warning effect of coherence markers in persuasion: off-line and on-line evidence.
- Kamalski, J., Sanders, T., Lentz, L., & van den Bergh, H. (2005). Hoe kun je het beste meten of een leerling een tekst begrijpt? Een vergelijkend onderzoek naar vier methoden. *Levende Talen*, 4, 3-9.
- Kamalski, J., Sanders, T., Lentz, L., & van den Bergh, H. (submitted). How to assess situation model representations. On the validity of text comprehension tasks.
- Kamalski, J., van den Bergh, H., Lentz, L., & Sanders, T. (2005, July 6-9). *How to measure a crucial level of text representation: the situation model*. Paper presented at the Fifteenth Annual Meeting of the Society for Text and Discourse, Amsterdam.
- Kintsch, W. (1988). The role of knowledge in Discourse Comprehension: A Construction-Integration Model. *Psychological Review*, 95(2), 163-182.
- Kintsch, W. (1998). *Comprehension. A paradigm for cognition*. Cambridge: Cambridge University Press.
- Knott, A., & Dale, R. (1994). Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, 18, 35-62.
- Knott, A., & Sanders, T. (1998). The classification of coherence relations and their linguistic markers: an exploration of two languages. *Journal of Pragmatics*, 30, 135-175.
- Koelen, M., & Martijn, C. (1994). Persuasieve Voorlichting. In N. G. Röling, D. Kuiper & R. Janmaat (Eds.), *Basisboek voorlichtingskunde* (pp. 217-239). Amsterdam/Meppel: Boom.
- Kok, G., & Damoiseaux, V. (1996). Beïnvloeding van attitudes en gedrag. In B. Klandermans & E. Seydel (Eds.), *Overtuigen en activeren*. Assen: Van Gorcum.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

204 - REFERENCES

- Land, J., Sanders, T., Lentz, L., & Bergh, v. d., H. (2002). Coherentie en identificatie in studieboeken. Een empirisch onderzoek naar tekstbegrip en tekstwaardering op het vmbo. *Tijdschrift voor Taalbeheersing*(4), 281-302.
- Langacker, R. W. (1990). Subjectification. *Cognitive Linguistics*(1), 5-38.
- Langer, J. A. (1984). Examining background knowledge in text comprehension. *Reading Research Quarterly*, 468-481.
- Lentz, L., & De Jong, M. (1997). The evaluation of text quality: expert-focused and reader-focused methods compared. *IEEE Transactions on Professional Communication*, 40(3), 224-234.
- Lentz, L., & de Jong, M. (2003). In welke termen denken lezers over tekstproblemen? *Tijdschrift voor Taalbeheersing*.
- Levenston, E. A., Nir, R., & Blum-Kulka, S. (1984). Discourse Analysis and the testing of reading comprehension by cloze techniques. In A. K. Pugh & J. M. Ulijn (Eds.), *Reading for professional purposes. Studies and practices in native and foreign languages*. London: Heinemann.
- Leventson, E. A., Nir, R., & Blum-Kulka, S. (1984). Discourse Analysis and the testing of reading comprehension by cloze techniques. In A. K. Pugh & J. M. Ulijn (Eds.), *Reading for professional purposes. Studies and practices in native and foreign languages*. London: Heinemann.
- Linderholm, T., Everson, M. G., van den Broek, P., Mischinski, M., Crittenden, A., & Samuels, J. (2000). Effects of Causal Text Revisions on More- and Less-skilled Readers' Comprehension of Easy and Difficult Texts. *Cognition and Instruction*, 18(4), 525-556.
- Long, D. L., Wilson, J., Hurley, R., & Prat, C. S. (2006). Assessing text representations with recognition: the interaction of domain knowledge and text coherence. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 32(4), 816-827.
- Lorch, R. F., & Lorch, E. P. (1986). On-line Processing of Summary and Importance Signals in Reading. *Discourse Processes*, 9, 489-496.
- Lorch, R. F., Lorch, E. P., Ritchley, K., McGovern, L., & Coleman, D. (2001). Effects of Headings on Text Summarization. *Contemporary Educational Psychology*, 26, 171-191.
- Lord, F. M., & Novik, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading Massachusetts: Addison-Wesley.
- Mann, W. C., & Thompson, S. A. (1988). Rhetorical Structure Theory. Toward a functional theory of text organization. *Text: an Interdisciplinary Journal for the Study of Discourse*, 8, 243-281.
- Maury, P., & Teisserenc, A. (2005). The role of connectives in science text comprehension and memory. *Language and Cognitive Processes*, 20(3), 489-512.
- McGuire, W. J., & Papageorgis, D. (1962). Effectiveness of Forewarning in Developing Resistance to Persuasion. *The Public Opinion Quarterly*, 26(1), 24-34.
- McKeown, M. G., Beck, I. L., Sinatra, G. M., & Loxterman, J. A. (1992). The contribution of prior knowledge and coherent text to comprehension. *Reading Research Quarterly*, 27, 78-93.
- McNamara, D. S. (2001). Reading both High-Coherence and Low-Coherence Texts: Effects of Text Sequence and Prior Knowledge. *Canadian Journal of Experimental Psychology*, 55(1), 51-62.

REFERENCES- 205

- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1-43.
- McNamara, D. S., & Kintsch, W. (1996). Learning from texts: effects of prior knowledge and text coherence. *Discourse Processes, 22*, 247-288.
- Meyer, B. J. F., Brandt, D. M., & Bluth, G. J. (1980). Use of top level structure in text: key for reading comprehension of ninth grade students. *Reading Research Quarterly, 16*(1), 72-103.
- Meyers-Levy, J., & Malaviya, P. (1999). Consumers' Processing of Persuasive Advertisements: An integrative framework of Persuasion Theories. *Journal of Marketing, 63*, 45-60.
- Miller, G. A. (1969). A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology, 6*, 169-191.
- Millis, K. K., & Just, M. A. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language, 33*, 128-147.
- Moeschler, J., & Reboul, A. (1994). *Dictionnaire encyclopédique de pragmatique*. Paris: Editions du Seuil.
- Moreno, R., & Mayer, R. E. (1999). Cognitive Principles of Multimedia Learning: The role of Modality and Contiguity. *Journal of Educational Psychology, 91*(2), 358-368.
- Murphy, P. K. (2001). What makes a text persuasive? Comparing students' and experts' conceptions of persuasiveness. *International Journal of Educational Research, 35*, 675-698.
- Murray, J. D. (1995). Logical Connectives and Local Coherence. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of Coherence in Reading* (pp. 106-125). Hillsdale, NJ: Erlbaum.
- Nolen, S. B. (1995). Effects of a Visible Author in Statistical Texts. *Journal of Educational Psychology, 87*(1), 47-65.
- Noordman, L. G. M., & de Blijzer, F. (2000). On the processing of causal relations. In E. Couper-Kuhlen & B. Kortmann (Eds.), *Cause, Condition, Concession, Contrast* (pp. 35-56). Berlin: Mouton de Gruyter.
- Noordman, L. G. M., & Vonk, W. (1997). The different functions of a conjunction in constructing a representation of the discourse. In J. Costermans & M. Fayol (Eds.), *Processing Interclausal Relationships. Studies in the production and comprehension of text*. (pp. 75-93). Mahwah, NJ: Erlbaum.
- Noordman, L. G. M., & Vonk, W. (1998). Memory-based Processing in Understanding Causal Information. *Discourse Processes, 26*(2), 191-212.
- O'Keefe, D. J. (1990). *Persuasion. Theory and Research*. (Vol. 2). Newbury Park, California: Sage Publications.
- O'Keefe, D. J. (1997). Standpoint explicitness and persuasive effect: a meta-analytic review of the effects of varying conclusion articulation in persuasive messages. *Argumentation and Advocacy, 34*, 1-12.
- Olsen, J. R., & Biolsi, K. J. (1991). Techniques for representing expert knowledge. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise. Prospects and limits*. (pp. 240-285). Cambridge: Cambridge University Press.
- Ortony, A., Clore, G. L., & Collins, A. (2001). *The cognitive structure of emotions*. Cambridge: Cambridge University Press.
- Oversteegen, L., van Keulen, R., & van Wijk, C. (2002). Afwisseling versus herhaling in woordgebruik. *Tijdschrift voor Taalbeheersing*(4), 303-331.

206 - REFERENCES

- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist, 38*(1), 1-4.
- Paas, F., Tuovinen, J. E., Tabbers, H., & van Gerven, P. W. M. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist, 38*(1), 63-71.
- Pander Maat, H., & Degand, L. (2001). Scaling Causal Relations and Connectives in terms of Speaker Involvement. *Cognitive Linguistics, 12*(3), 211-245.
- Pander Maat, H., & Sanders, T. (2001). Subjectivity in causal connectives: An empirical study of language in use. *Cognitive Linguistics, 12*(3), 247-273.
- Paris, S. G., & Stahl, S. (2005). *Children's Reading Comprehension and Assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Pearson, P. D., & Hamm, D. N. (2005). The Assessment of Reading Comprehension: A review of Practices - Past, Present, and Future. In S. G. Paris & S. Stahl (Eds.), *Children's Reading Comprehension and Assessment* (pp. 13- 69). Mahwah, NJ: Lawrence Erlbaum Associates.
- Petty, R. E., & Cacioppo, J. T. (1979). Effects of forewarning of persuasive intent and involvement on cognitive responses and persuasion. *Personality and Social Psychology Bulletin, 5*(2), 173-176.
- Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion: Classic and Contemporary Approaches*. Dubuque, Iowa: Wm. C. Brown Company Publishers.
- Pit, M. (2003). *How to express yourself with a causal connective. Subjectivity and causal connectives in Dutch, German and French*. Amsterdam: Rodopi.
- Quené, H., & van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: a tutorial. *Speech Communication, 43*, 103-121.
- Ratcliff, R. (1993). Methods for Dealing with Reaction Time Outliers. *Psychological Bulletin, 114*(3), 510-532.
- Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*(1), 69-80.
- Roebben, N. (2004). *Etude sur l'impact des connecteurs de cause sur la comprehension de textes expositifs*. Universite catholique de Louvain: Doctoral dissertation.
- Roller, C. M. (1990). The interaction between knowledge and structure variables in the processing of expository prose. *Reading Research Quarterly, 25*, 79-89.
- Romero, A. A., Agnew, C. R., & Insko, C. A. (1996). The Cognitive Mediation Hypothesis Revisited: An Empirical Response to Methodological and Theoretical Criticism. *Personality and Social Psychology Bulletin, 22*(7), 651-665.
- Roseman, I. J., Antoniou, A. A., & Jose, P. E. (1996). Appraisal Determinants of emotions: constructing a more accurate and comprehensive theory. *Cognition and Emotion, 10*(3), 241-277.
- Royer, J. M., Cisero, C. A., & Carlo, M. S. (1993). Techniques and Procedures for Assessing Cognitive Skills. *Review of Educational Research, 63*(2), 201-243.
- Rymenans, R. (2004). *Onderzoek naar kenmerken van effectieve scholen*. Utrecht University: Doctoral dissertation.
- Sanchez, R. P., Lorch, E. P., & Lorch, R. F. (2001). Effects of Headings on Text Processing Strategies. *Contemporary Educational Psychology, 26*, 418-428.
- Sanders, T. (1997). Semantic and Pragmatic Sources of Coherence: On the Categorization of Coherence Relations in Context. *Discourse Processes, 24*, 119-147.

REFERENCES- 207

- Sanders, T. (2005). Coherence, Causality and Cognitive Complexity in Discourse. In M. Aurnague, M. Bras, A. L. Draoulec & L. Vieu (Eds.), *Proceedings of the First International Symposium on the Exploration and Modelling of Meaning SEM-05* (pp. 31-44). Biarritz, France.
- Sanders, T., & Noordman, L. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes, 29* (1), 37-60.
- Sanders, T., & Spooren, W. (2001). Text representation as an interface between language and its users. In T. Sanders, J. Schilperoord & W. Spooren (Eds.), *Text representation: linguistic and psycholinguistic aspects*. Amsterdam: John Benjamins Publishing.
- Sanders, T., & Spooren, W. (2006). Discourse and Text Structure. In H. Cuyckens & D. Geeraerts (Eds.), *Handbook of Cognitive Linguistics*. Oxford: Oxford University Press.
- Sanders, T., Spooren, W. P. M., & Noordman, L. G. M. (1992). Toward a Taxonomy of Coherence Relations. *Discourse Processes, 15*(1), 1-35.
- Sanford, A. J., & Garrod, S. C. (1994). Selective Processing in Text Understanding. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 699-719). San Diego: Academic Press.
- Schellens, P. J., & de Jong, M. (2004). Argumentation Schemes in Persuasive Brochures. *Argumentation, 18*, 295-323.
- Scherer, K. R. (1999). On the sequential nature of appraisal processes: indirect evidence from a recognition task. *Cognition and Emotion, 13*(6), 763-793.
- Schmalhofer, F., & Glavanov, D. (1986). Three components of Understanding a Programmer's Manual: Verbatim, Propositional, and Situational Representations. *Journal of Memory and Language, 25*, 279-294.
- Scouller, K. M., & Prosser, M. (1994). Students' experiences in studying for multiple choice question examinations. *Studies in Higher Education, 19*(3), 267-280.
- Shanahan, T., & Kamil, M. L. (1984). The relationship of the concurrent and construct validities of cloze. In J. A. Niles & L. A. Harris (Eds.), *Changing perspectives on research in reading/ language processing and instruction*. (pp. 252-256). Rochester, NY: National Reading Conference.
- Singer, M. (1990). *Psychology of Language. An introduction to sentence and discourse processes*. Hillsdale, NJ: Erlbaum.
- Snoeck Henkemans, F. (2001). Argumentation, explanation and causality. An exploration of current linguistic approaches to textual relations. In T. Sanders, J. Schilperoord & W. Spooren (Eds.), *Text representation: linguistic and psycholinguistic aspects*.
- Spooren, W., Mulder, M., & Hoeken, H. (1998). The role of interest and text structure in professional reading. *Journal of Research in Reading, 21*(2), 109-120.
- Spyridakis, J. H., & Standal, T. C. (1987). Signals in expository prose: Effects on reading comprehension. *Reading Research Quarterly, 28*, 285- 298.
- Stukker, N. (2005). *Causality marking across levels of language structure. A cognitive semantic analysis verbs and causal connectives in Dutch*. Utrecht University: Doctoral dissertation.
- Sweetser, E. (1990). *From etymology to pragmatics. Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.
- Taylor, W. (1953). Cloze procedure: a new tool for measuring readability. *Journalism Quarterly, 30*, 414-438.

208 - REFERENCES

- Tierney, R. J., LaZansky, J., Raphael, T., & Cohen, P. (1987). Author's intentions and reader's interpretations. In R. J. Tierney, P. Andes & J. Mitchell (Eds.), *Understanding reader's understanding* (pp. 205-226). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Tormala, Z. L., & Petty, R. E. (2002). What doesn't kill me makes me stronger: The effects of resisting persuasion on attitude certainty. *Journal of Personality and Social Psychology*, *83*(6), 1289-1313.
- Traxler, M. J., Bybee, M. D., & Pickering, M. J. (1997). Influence of Connectives on Language Comprehension: Eye-tracking Evidence for Incremental Interpretation. *The Quarterly Journal of Experimental Psychology*, *50*(a)(3), 481-497.
- Traxler, M. J., Sanford, A. J., Aked, J. P., & Moxey, L. M. (1997). Processing Causal and Diagnostic Statements in Discourse. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *23*(1), 88-101.
- van den Bergh, H. (1990). On the Construct Validity of Multiple Choice Items for Reading Comprehension. *Applied Psychological Measurement*, *14*(1), 1-12.
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- van Oostendorp, H., & De Mul, S. (1990). Moses beats Adam: a semantic relatedness effect on a semantic illusion. *Acta Psychologica*, *75*, 35-46.
- van Wijk, C. (1996). Persuasieve effecten van presentatiestijl; Een toepassing op overheidsvoorlichting. *Tijdschrift voor Taalbeheersing*, *18*, 368-382.
- van Wijk, C. (2000). *Toetsende Statistiek: Basistechnieken. Een praktijkgerichte inleiding voor onderzoekers van taal, gedrag en communicatie*. Bussum: Coutinho.
- Verhagen, A. (2005). *Constructions of Intersubjectivity. Discourse, Syntax and Cognition*. Oxford: Oxford University Press.
- Vivanco, V. (2005). The absence of connectives and the maintenance of coherence in publicity texts. *Journal of Pragmatics*, *37*, 1233-1249.
- Voss, J. F., Fincher-Kiefer, R., Wiley, J., & Ney Silfies, L. (1993). On the processing of arguments. *Argumentation*, *7*(2), 165-181.
- Voss, J. F., & Ney Silfies, L. (1996). Learning from history text: the interaction of knowledge and comprehension skill with text structure. *Cognition and Instruction*, *14*, 45-68.
- Waltzer, H. (1988). Corporate advocacy: Advertising and political influence. *Public Relations Review* *14*, 41-55.
- Wiley, J. (2005). A fair and balanced look at the news: What affects memory for controversial arguments? *Journal of Memory and Language*, *53*, 95-109.
- Wood, W., & Quinn, J. M. (2003). Forewarned and Forearmed? Two Meta-analytic Syntheses of Forewarnings of Influence Appeals. *Psychological Bulletin*, *129*(1), 119-138.
- Zuwerink, J. R., & Devine, P. G. (1996). Attitude Importance and Resistance to Persuasion: It's Not Just The Thought That Counts. *Journal of Personality and Social Psychology*, *70*(5), 931-944.
- Zwaan, R. A. (1994). Effect of genre expectations on text comprehension. *Journal of Experimental Psychology*, *20*(4), 920-933.
- Zwaan, R. A. (1999a). Embodied Cognition, Perceptual Symbols, and Situation Models. *Discourse Processes*, *28*(1), 81-88.
- Zwaan, R. A. (1999b). Situation Models: The Mental Leap Into Imagined Worlds. *Current Directions in Psychological Science*, *8*(1), 15-18.

REFERENCES- 209

- Zwaan, R. A., & Radvansky, G. A. (1998). Situation Models in Language Comprehension and Memory. *Psychological Bulletin*, *123*(2), 162-185.
- Zwaan, R. A., & Taylor, L. J. (2006). Seeing, Acting, Understanding: Motor Resonance in Language Comprehension. *Journal of Experimental Psychology : General*, *135*(11), 1-11.

Samenvatting in het Nederlands

Dit proefschrift gaat over het effect dat zogenaamde **coherentiemarkeringen** hebben op de mentale representatie die de lezer vormt van een tekst. De volgende voorbeelden illustreren het begrip coherentiemarkering.

1. Elsevier heeft een succesvol jaar achter de rug. De wetenschappelijke uitgever boekte een groei van bijna 6 procent tot ruim 1,5 miljard pond.
2. Elsevier heeft een succesvol jaar achter de rug, *want* de wetenschappelijke uitgever boekte een groei van bijna 6 procent tot ruim 1,5 miljard pond.

Het eerste voorbeeld komt uit een krantenbericht in de Volkskrant van 15 februari 2007. Voorbeeld 2 koppelt de twee zinnen aan elkaar met behulp van het woord *want*. *Want* is een zogenaamde **coherentiemarkering**: het markeert de coherentierelatie die bestaat tussen de eerste en de tweede zin, in dit geval een bewering-argument relatie. De relatie tussen deze twee tekstdelen blijft niet impliciet. De lezer hoeft niet helemaal zelf af te leiden wat de relatie is tussen deze twee tekstsegmenten; dit wordt al aangegeven door de coherentiemarkering. In dit proefschrift staan coherentiemarkeringen en hun effect op de lezer centraal.

Behalve connectieven, zoals *want* of *omdat*, zijn er andere soorten coherentiemarkeringen: lexicaal signaalzinnen (*de reden hiervan is dat, we kunnen nu concluderen dat*), organizers (*in de volgende paragraaf worden twee oorzaken genoemd voor dit probleem*) en kopjes (*oorzaken*). Dit zijn allemaal voorbeelden van het markeren van *relationele coherentie*. Een ander type coherentie dat ook gemarkeerd kan worden, is *referentiële coherentie*. Als een tekst expliciet referentieel coherent is, dan wordt vaak gerefereerd aan hetzelfde woord en dezelfde term. In een implicietere tekst, worden referenten zoals *hij*, *zij*, *het* gebruikt. In het bovenstaande voorbeeld gebruiken beide teksten de eerste keer Elsevier om naar het bedrijf te verwijzen, en de tweede keer 'de wetenschappelijke uitgever'. De lezer moet dan zelf de link leggen dat dit twee keer hetzelfde bedrijf betreft.

Coherentie speelt een cruciale rol bij het construeren van de mentale representatie van de informatie uit de tekst. Met andere woorden, een lezer heeft een tekst pas begrepen als verbanden gelegd zijn tussen verschillende concepten of gebeurtenissen in een tekst. Of er nu coherentiemarkering

212 – SAMENVATTING

gebruikt wordt in een tekst of niet, de lezer zou in beide gevallen in staat moeten zijn de relatie tussen twee tekstsegmenten te infereren. Of de schrijver nu *want* gebruikt zoals in voorbeeld 2, of dit connectief weglaat zoals in voorbeeld 1, de coherentierelatie tussen de twee zinnen is en blijft dezelfde. De markering verandert die relatie niet, maar expliciteert en signaleert.

Hoofdstuk 1: coherentiemarkering

In hoofdstuk 1 laat ik zien dat coherentiemarkering een ruim begrip is. Vaak schaart men ook extra voorbeelden, uitleg van afkortingen en additionele informatie over het tekstonderwerp onder het kopje coherentiemarkering. Dat is in dit proefschrift duidelijk niet het geval. Naar mijn idee is er alleen sprake van coherentiemarkering als geen additionele inhoudelijke informatie toegevoegd wordt aan de inhoud van de tekst, maar slechts de tekststructuur en coherentierelaties tussen zinnen geëxpliciteerd worden. In dit proefschrift staat daarmee het effect van puur linguïstische coherentiemarkering op de lezer centraal. De belangrijkste vraag is: Als in een tekst de coherentie expliciet gemarkeerd wordt, of juist niet, wat voor effecten heeft dat dan op de lezer?

Hoofdstuk 2: Effecten van coherentiemarkering op de lezer

Hoofdstuk 2 bespreekt voorbeelden van eerder onderzoek naar de effecten van coherentiemarkering op de mentale representatie die de lezer construeert van de informatie in de tekst. Dit onderzoek richt zich vaak alleen op tekstbegrip: zorgen coherentiemarkeringen ervoor dat de lezer de tekst beter begrijpt, of juist niet? Het lijkt erop dat coherentiemarkering een positieve invloed heeft op tekstbegrip als de lezer weinig van het onderwerp van de tekst afweet. De lezer heeft de markeringen dan als het ware nodig om de relaties tussen de tekstsegmenten te leggen, want er is niet genoeg inhoudelijke voorkennis om dit zelf te kunnen doen. Voor lezers die veel voorkennis hebben over het onderwerp van de tekst ligt de situatie anders. Zij hebben de markeringen niet nodig, sterker nog, onderzoek lijkt erop te wijzen dat deze markeringen hun tekstbegrip niet ten goede komen, maar zelfs hinderen. De verklaring die hiervoor dan vaak gegeven wordt is dat lezers met voorkennis de tekstuele informatie dieper verwerken als ze zelf relaties en verbanden moeten leggen. Als ze daarbij geholpen worden door een markering, verwerken ze de informatie uit de tekst op een meer oppervlakkige manier.

Coherentiemarkering beïnvloedt de situatiemodel representatie die lezers construeren van de informatie in een tekst. In hoofdstuk 2 geef ik een voorbeeld van dit type representatie aan de hand van de zin: *de prins houdt de prinses vast*. Om van deze zin een situatiemodel representatie te vormen, is het nodig de informatie te integreren met eerdere informatie, voorkennis en context. De representatie van de ene lezer kan een stripboekachtige situatie zijn van een innig verliefde prins en prinses voor een kasteel met een draak erin. Maar, als deze zin in de krant staat onder de titel *Alexia geboren!*, dan is een representatie van Willem Alexander die zijn dochter trots in zijn armen houdt meer van toepassing.

Dit voorbeeld geeft aan hoe rijk en ingewikkeld situatiemodel representaties zijn. Ze hangen af van voorkennis, van eerdere ervaringen en meningen die de lezer kan hebben. Daarom stel ik in hoofdstuk 2 voor om deze factoren ook mee te nemen als we onderzoeken hoe coherentiemarkering de lezer kan beïnvloeden. Het is zeker voorstelbaar dat een lezer meningen en eerdere ervaringen *gebruikt* bij het construeren van een situatiemodel representatie. Deze aspecten maken dus ook deel uit van die representatie. Een voorbeeld uit dit proefschrift: als je een tekst leest over de Eiffeltoren, dan gebruik je je eigen mening, je eventuele eerdere ervaring en herinnering om je een beeld te vormen van de informatie in de tekst. Het is haast onmogelijk, zelfs al heb je de Eiffeltoren niet gezien, om geen mening te hebben na het lezen van die tekst. Daarom stel ik voor om niet alleen begrip te meten als we willen weten hoe de situatiemodel representatie van een lezer eruit ziet, maar ook aspecten als waardering en attitudes. Dit wil zeggen dat ik de invloed van coherentiemarkering en voorkennis op de lezer onderzoek door te na te gaan wat voor effect ze hebben op tekstbegrip, waardering en overtuigingskracht.

Hoofdstuk 3: een verkenning van factoren in pilot-experiment 1

In hoofdstuk 3 rapporteer ik een eerste pilot-experiment waarin ik deze factoren combineer: de onafhankelijke variabelen coherentiemarkering en voorkennis, plus de afhankelijke variabelen tekstbegrip, waardering en overtuigingskracht. Daarnaast bekijk ik of dit effect gegeneraliseerd kan worden over genres. Onderzoeken naar coherentiemarkering betreffen bijna altijd informerende teksten. In dit eerste pilot-experiment gebruik ik naast informerende teksten ook persuasieve teksten. De meest gebruikte definitie van deze twee genres maakt gebruik van de auteursintentie: als de auteur de intentie

214 – SAMENVATTING

heeft om informatie te verschaffen aan de lezer, betreft het een informerende tekst. Als de auteur de intentie heeft de lezer te overtuigen, dan betreft het een persuasieve tekst. Vaak wordt getracht de lezer te overtuigen door middel van pragmatische argumentatie: het benadrukken van de (on)wenselijkheid van de gevolgen van bepaald gedrag. Hoewel deze definitie in theorie volstaat, is in de praktijk de auteursintentie niet altijd even transparant. Daarom geef ik hoofdstuk 2 een gedetailleerde lijst met tekstkenmerken die kenmerkend zijn voor het persuasieve genre. Deze lijst is ook gebruikt bij de constructie van de experimentele materialen in dit proefschrift.

Het pilot-experiment in hoofdstuk 3 laat zien dat er in het informatieve tekstgenre inderdaad een interactie optreedt tussen voorkennis en coherentiemarkering. Lezers zonder voorkennis presteren beter op tekstbegripvragen na het lezen van een expliciet gemarkeerde tekst. Lezers met voorkennis beantwoorden de begripvragen even goed na het lezen van een impliciete tekst als na het lezen van een expliciete tekst. In het persuasieve genre wordt er een hoofdeffect gevonden van coherentiemarkering op begrip en geen interactie tussen voorkennis en coherentiemarkering. Voor beide groepen lezers, met of zonder voorkennis, leidt de expliciet gemarkeerde tekst tot beter tekstbegrip.

In dit pilot-experiment zijn tevens vragen gesteld over tekstwaardering. Proefpersonen gaven hun mening over de tekst door middel van 7-punts semantische differentiaal (bijvoorbeeld: deze tekst vind ik makkelijk 1 2 3 4 5 6 7 moeilijk). Lezers blijken de aanwezigheid van markeringen in beide genres positief te waarderen: de expliciete tekst werd makkelijker en duidelijker gevonden. Naast begrip en waardering, werd lezers ook gevraagd hun eigen begrip in te schatten (ook wel *Feeling of Knowing* genoemd). Uit de resultaten blijkt, dat lezers door markeringen het gevoel krijgen dat ze de tekst goed begrijpen. Correlatie-analyses geven aan dat lezers in staat zijn hun eigen tekstbegrip correct in te schatten.

De resultaten op overtuigingskracht blijken niet eenduidig. Daarom kom ik hierop in hoofdstuk 6 terug. Een ander probleem is de meetmethode van tekstbegrip. In dit experiment zijn open vragen gesteld, die achteraf niet optimaal betrouwbaar bleken te zijn. Daarom is hoofdstuk 4 gewijd aan de zoektocht naar een betere manier op tekstbegrip te meten.

Hoofdstuk 4: meten van situatiemodel representaties in validatie experiment 2

In hoofdstuk 4 vergelijk ik vier verschillende methoden om tekstbegrip te meten: bridging inference vragen (multiple choice of open vragen), een cloze test (een ‘gatentekst’), een sorteertaak (het groeperen van begrippen uit de tekst), en een mental model taak (het in een schema plaatsen van begrippen uit de tekst). Zowel in hoofdstuk 4 als in Appendix 2 staan voorbeelden van deze taken. De vragen en de cloze taak zijn traditionele tekstbegripmethoden, de sorteertaak en de mental modeltaak zijn pas recentelijk gebruikt om tekstbegrip te meten.

Op diverse aspecten van validiteit zijn deze methodes vergeleken: interne betrouwbaarheid, convergente validiteit, divergente validiteit en known-group validiteit. In een groot experiment op vier verschillende middelbare scholen in Nederland hebben 200 4vwo-ers en 200 6vwo-ers ieder 4 teksten gelezen uit het eindexamen Nederlands en daar één van de taken over beantwoord.

De resultaten van dit experiment lijken erop te wijzen dat de sorteertaak de meest geschikte taak is om situatiemodel representaties te meten: op geen van de validiteitscriteria heb ik deze taak kunnen afwijzen. De vraag welke methode op de tweede plaats komt, is moeilijker te beantwoorden. De andere taken hebben allemaal voor- en nadelen. De cloze taak heeft als voordeel dat de taak intern erg betrouwbaar is, maar als nadeel dat hij in 6vwo de verschillen tussen leerlingen niet goed kon weergeven. De mental modeltaak heeft als nadeel een relatief onbetrouwbaar resultaat, maar is wel geschikt voor beide groepen leerlingen. De (open en gesloten) tekstbegripvragen komen bij dit experiment als slechtste optie uit de bus. Ze scoorden het laagst op interne betrouwbaarheid, op samenhang met andere methoden en weinig van de variatie tussen de leerlingen bleek toe te wijzen aan de capaciteiten van de leerling. Samengevat betekent dit dat de sorteertaak de meest valide en betrouwbare methode lijkt te zijn om situatiemodel representaties te meten.

Hoofdstuk 5: terug naar het effect van coherentiemarkering op begrip in experiment 3 en 4

Nu vastgesteld is dat de sorteertaak de meest valide methode is om het situatiemodel niveau te meten, is het noodzakelijk om pilot-experiment 1 in hoofdstuk 3 opnieuw uit te voeren met deze meer valide methode. Dit gebeurt

216 – SAMENVATTING

in experiment 3 in hoofdstuk 5. Opnieuw wordt onderzocht hoe coherentiemarkeringen het tekstbegrip beïnvloeden. De resultaten van dit experiment tonen geen interactie tussen coherentiemarkering en voorkennis: voor zowel lezers met als lezers zonder voorkennis leidt de expliciete tekst versie tot beter tekstbegrip. Dit resultaat verschilt van het eerdere resultaat in experiment 1. De vraag is nu waardoor dit verschil veroorzaakt wordt: het gebruik van een meer valide tekstbegripmethode, of een andere factor in het experiment? Als we in meer detail naar de resultaten kijken, dan zien we dat de lezers zonder voorkennis op dezelfde manier reageren als in het pilot-experiment: voor hen zijn teksten met markeringen beter te begrijpen. De experts daarentegen vertonen ander gedrag in experiment 3 dan in experiment 1: ook voor hen lijken nu de markeringen voordelig te zijn. Dit resultaat is moeilijk te interpreteren. Waarom vertonen de experts in experiment 3 ander gedrag dan in experiment 1? Het is niet uit te sluiten dat dit effect veroorzaakt wordt door de manipulatie van voorkennis, of liever gezegd, de afwezigheid van die manipulatie. Proefpersonen in beide experimenten waren geselecteerd op basis van de kennis die ze geacht werden te hebben en zo werden twee groepen proefpersonen gevormd: met veel (studenten medicijnen of biologie) en weinig voorkennis (studenten geschiedenis of rechten) over het tekstonderwerp (genetische manipulatie en orgaandonatie).

In experiment 4 in hoofdstuk 5 wordt de voorkennis wél direct gemanipuleerd door middel van filmpjes uit actualiteitenprogramma's. Deze filmpjes verschaffen voorkennis en activeren tevens voorafgaand aan het lezen de concepten in de kennisstructuur van de lezer die relevant zijn voor het begrijpen van de teksten. Opnieuw wordt begrip gemeten aan de hand van de sorteertaak.

In het informatieve genre treedt een interactie op tussen voorkennis en coherentiemarkering, precies zoals eerder beschreven in deze samenvatting (hoofdstuk 2 en 3). Het is inderdaad zo dat lezers met weinig voorkennis over het tekstonderwerp baat hebben bij een expliciet gemarkeerde tekst, terwijl lezers met meer voorkennis juist beter presteren op tekstbegrip na het lezen van een impliciete tekst. In het persuasieve genre is wederom sprake van een hoofdeffect: voor beide groepen lezers is de tekst met markering voordelig voor het begrip ervan. Op waardering worden de effecten uit experiment 1 gerepliceerd: coherentiemarkering heeft een positief effect op waardering.

Het theoretisch plausibele interactie-effect tussen voorkennis en coherentiemarkering is door meerdere onderzoekers onderzocht, maar vaak niet gerepliceerd. In dit onderzoek is deze interactie wel gerepliceerd, met zorgvuldig geconstrueerde teksten die slechts verschillen in de mate waarin zij linguïstische coherentiemarkeringen bevatten. Ook heb ik aangetoond dat genre van invloed is op het al dan niet optreden van deze interactie. In de discussie kom ik hier uitgebreider op terug.

Hoofdstuk 6: coherentiemarkering en overtuigingskracht: forewarning

Het *forewarning* principe, bekend uit de sociale psychologie, beschrijft hoe mensen weerstand opbouwen tegen een poging om hen te overtuigen wanneer zij zich bewust zijn van de anders intentie van de ander om te overtuigen. Vervolgens wordt het moeilijk, zo niet onmogelijk, om hen nog te overtuigen. In hoofdstuk 6 stel ik een model voor dat met behulp van het verschijnsel *forewarning* voorspelt welke effecten coherentiemarkering kan hebben op de overtuigingskracht van een tekst.

De theoretische voorspelling is dat *subjectieve* markeringen weerstand veroorzaken bij de lezer: als de lezer eenmaal door heeft dat de schrijver de intentie heeft de lezer te overtuigen, dan bouwt de lezer weerstand op. *Objectieve* markeringen zouden dit effect niet moeten hebben. Het volgende voorbeeld over een ijshockeywedstrijd illustreert het verschil tussen objectieve en subjectieve relaties:

5. Er werd flink gevochten op het ijs. Een van de spelers raakte gewond.
6. Er werd flink gevochten op het ijs. Het was een erg interessante wedstrijd.

In voorbeeld 5 bestaat er een objectieve relatie tussen het gewond raken en het vechten, die door de spreker wordt gerapporteerd. In voorbeeld 6 bestaat er een subjectieve relatie: de relatie tussen het vechten en de conclusie dat het interessant was bestaat, omdat de schrijver/spreker dit vindt. Beide relaties kunnen gemarkeerd worden: in voorbeeld 5 kunnen de zinnen aan elkaar gerelateerd worden door het objectieve connectief *daardoor*, in voorbeeld 6 kunnen de zinnen gerelateerd worden door het subjectieve connectief *dus*.

Het model dat ik voorstel in hoofdstuk 6 gaat uit van de aanname dat coherentiemarkeringen de lezer instructies geven over hoe de tekst verwerkt moet worden. De volgende stappen zijn daarbij van belang: het detecteren van

subjectiviteit, het herkennen en identificeren van de auteursintentie, het ervaren van weerstand en tot slot het behouden van de originele attitude.

Hoofdstuk 7: testen van de forewarning hypothese, experiment 5 en 6

In hoofdstuk 7 worden de forewarning hypothese en de vier voorgestelde stappen getest. Hiervoor is met Nederlandse proefpersonen een off-line lees experiment gedaan, waarbij lezers naar hun mening werd gevraagd nadat ze een tekst hadden gelezen met ofwel subjectieve, ofwel objectieve, ofwel weinig tot geen markerings. Uit de resultaten blijkt dat lezers eerder geneigd zijn het met de schrijver eens te zijn na het lezen van een objectief gemarkeerde versie dan na een subjectief gemarkeerde versie. Ook is hetzelfde verschil zichtbaar wanneer gevraagd wordt naar de auteursintentie: lezers vinden de intentie van de auteur persuasiever wanneer ze een subjectieve versie gelezen hebben dan als ze een objectieve versie gelezen hebben. Dit bevestigt de laatste stap in het model dat in hoofdstuk 6 voorgesteld is.

In een on-line leesexperiment, uitgevoerd onder Amerikaanse studenten, ontstaat een vergelijkbaar maar niet identiek patroon: na het lezen van een tekst met één enkele subjectieve markering, zijn mensen minder geneigd het met de schrijver eens te zijn dan na een tekst zonder marking. Dit resultaat verschilt van de uitkomst van het off-line experiment, want daar verschilden de objectieve en de subjectieve versie sterker van elkaar. Wel laten de resultaten opnieuw zien dat ook de gerapporteerde auteursintentie beïnvloed wordt door de markerings. De verschillen met het off-line experiment lijken verklaard te kunnen worden aan de hand van het lezerskenmerk betrokkenheid. Voor betrokken lezers lijkt een subjectieve markering niet tot weerstand te leiden: zij zijn dermate gemotiveerd om tot een in hun ogen correcte attitude te komen, dat een signaal van subjectiviteit een zorgvuldige afweging van argumenten niet in de weg staat. Voor minder betrokken lezers lijkt een signaal van subjectiviteit wel een reden om minder diep te verwerken: er treedt dan forewarning op en lezers blijven bij hun oorspronkelijke mening.

De leestijden die in dit experiment vastgelegd zijn tonen een interessant patroon. Ik heb leestijden vergeleken van twee zinnen die niet verbonden waren door middel van een coherentiemarkering met precies dezelfde zinnen die wel geëxpliciteerd waren door middel van een coherentiemarkering. Als die markering objectief is (bijv. *daardoor*), dan lezen lezers gemiddeld de tweede zin sneller als er wel een markering staat, dan als er

geen markering staat. Als de markering subjectief is (bijv. *dus*), versnellen lezers meteen na het lezen van de markering. Maar aan het einde van de zin vertragen ze ten opzichte van de ongemarkeerde zin. Dit effect is vaker gevonden en wordt een inferentie-effect genoemd. In eerder onderzoek betreft dit een causale inferentie: lezers checken of het causale verband ook klopt met hun eigen wereldkennis. In dit specifieke onderzoek lijkt het een auteursintentie-inferentie te betreffen: lezers herkennen en identificeren de auteursintentie. Dit bevestigt stap 1 en 2 van het model uit hoofdstuk 6.

Hoofdstuk 8: conclusie

In hoofdstuk 8 geef ik een overzicht van de verschillende effecten van coherentiemarkering zoals die geobserveerd zijn in de experimenten, zowel in het informerende als in het persuasieve genre. In vier stroomdiagrammen wordt de rol van coherentiemarkeringen visueel zichtbaar gemaakt. Belangrijke conclusies betreffende coherentiemarkering zijn dat: 1) coherentiemarkering de situatiemodel representatie beïnvloedt, het diepste niveau van tekstbegrip, 2) coherentiemarkering niet alleen interageert met voorkennis, maar ook met genre, 3) coherentiemarkering niet alleen effecten heeft op begrip, maar ook op waardering en overtuigingskracht en 4) het onderscheid tussen subjectieve en objectieve markeringen van belang is bij de overtuigingskracht van coherentiemarkeringen. Andere conclusies zijn dat 1) we kunnen laten zien dat forewarning veroorzaakt kan worden door specifieke tekstkenmerken en 2) de voorgestelde vier stappen in het model voor het grootste gedeelte bevestigd worden.

Ik bespreek in hoofdstuk 8 ook enkele onderwerpen die verder onderzoek vereisen. De precieze rol van het lezerskenmerk betrokkenheid is nog onduidelijk, omdat deze variabele niet systematisch in de experimenten was opgenomen. De vraag blijft tevens hoe groot de scope van een situatiemodel representatie eigenlijk is: wat hoort daar allemaal in? Door waardering en attitudes toe te voegen, is de scope breder geworden dan de traditionele betekenis van tekstbegrip en de integratie van tekstuele informatie en voorkennis. Maar waar houdt dat op? Hoeveel moet nog toegevoegd worden? Ik stel in hoofdstuk 8 dat het mogelijk is dat er nog andere aspecten toegevoegd moeten worden. Ik speculeer over de rol van *embodied cognition*, de theoretische stroming die veronderstelt dat kennis in het menselijk brein gerepresenteerd is

220 – SAMENVATTING

aan de hand van fysieke ervaringen. Nader procesonderzoek als eyetracking ligt hier voor de hand.

Tot slot kunnen we stellen dat de conclusies in dit proefschrift het belang van coherentiemarkeringen onderschrijven. Coherentiemarkeringen hebben niet alleen het vermogen lezers, afhankelijk van hun voorkennis, een tekst beter te laten begrijpen, maar ze beïnvloeden ook tekstwaardering en overtuigingskracht. Subjectieve coherentiemarkeringen kunnen weerstand veroorzaken, een zogenaamd forewarning effect. Hiermee is aangetoond dat forewarning veroorzaakt kan worden door specifieke tekstkenmerken, niet alleen door de expliciete boodschap dat er een poging gedaan zal worden de lezer te beïnvloeden. Ook is nader procesonderzoek nodig naar de leesprocessen die zich afspelen bij het veronderstelde forewarning effect als gevolg van het voorkomen van subjectieve tekstkenmerken. In dit proefschrift heb ik onderzoek uit drie verschillende disciplines gecombineerd: psycholinguïstisch onderzoek naar tekst verwerking met een taalkundige invalshoek, tekstlinguïstisch onderzoek naar verschillende typen coherentierelaties en de markering daarvan, en sociaal psychologisch onderzoek naar het forewarning effect. Door deze disciplines te combineren, hebben we meer inzicht gekregen in het complexe proces van het lezen van tekst. Coherentiemarkeringen lijken subtiele tekstkenmerken, maar dit proefschrift heeft laten zien dat ze verschillende duidelijke effecten kunnen hebben op de lezer.

Appendix 1:

(Part of) Materials from Pilot Experiment 1

1) Informative organ donation text, markers are underlined

Organ donation in The Netherlands

Since 1998, a new law exists in Holland with a Donor Register. This law didn't increase the number of available donors. There are already fewer donors per million inhabitants in Holland than Belgium, Austria or Spain. Moreover, the demand for organs still increases. Registration is voluntary, and because many people didn't register their choice, the decision of only one third of the Dutch people is known. 54% give permission for donation, 34 do not. The rest leave the decision to the surviving relatives. (...)

2) Persuasive organ donation text, markers are underlined

Organ donation: you can save lives!

Tim is only 16, but his heart seems worn out: he had leukaemia and the chemotherapy seriously damaged his heart. Tim has been on the waiting list since 1996. (...) These are only examples. At this moment, there's a long waiting list for donor organs. You can do something about this by giving permission for your organs to be donated after death. You can completely change people's lives. After all, after death you don't need your organs anymore. If you find this a difficult decision to make, the information in this brochure can help you.

Since 1998, a new law exists in Holland with a Donor Register. This law didn't increase the number of available donors. There are already fewer donors per million inhabitants in Holland than Belgium, Austria or Spain. Moreover, the demand for organs still increases. Registration is voluntary, and because many people didn't register their choice, the decision of only one third of the Dutch people is known. 54% give permission for donation, 34 do not. The rest leave the decision to the surviving relatives. (...)

Appendix 2: (Part of) Materials from Experiment 2

1) Text passage

Behind the façade [of politicians] should be professional, specialized knowledge. But also in this respect, something has changed. Better schooling and many information sources provide more knowledge on all kinds of professions, which is called protoprofessionalism. Lay people feel more like experts in for instance medical science, through medical programs and articles in the media. This has lead to a decrease in respect and an increase in mistrust and aggression towards doctors. The media have also caused political protoprofessionalism of Dutch citizens. This did not lead to more democracy, but it did lead to more mistrust and dissatisfaction.

2) Question Asking: Explain why protoprofessionalism could lead to mistrust (use 25 words maximum).

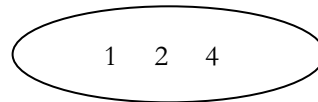
3) Cloze Task

Behind the façade [of politicians] should be professional, specialized _____. But also in this respect, something has changed. Better schooling and many _____ sources provide more knowledge on all kinds of professions, which is called proto-professionalism. Lay people feel more like experts in for instance _____ science, through medical programs and articles in the media.

4) Sorting Task

Here's a list of key words taken from the text. Make groups of words that you think should go together on the basis of the text. You can make as many groups as you want, and they can be of any size. Draw a circle for each group you want to make and put the right numbers in the circle.

- 1 Protoprofessionalism
- 2 Lay people have more knowledge
- 3 Increase in respect for experts
- 4 Mistrust of the government



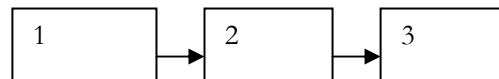
5) Mental Model Task

The text states that the Dutch people are not very happy with Dutch politics. In the text there are several chains from cause to consequence and they can be represented as follows:



Make chains on the basis of the text with the key words you find in the following list.

1. More attention for politics in the media
2. Lay people have more knowledge
3. Less respect for experts



Appendix 3:

(Part of) Materials from Experiments 3, 4 and 5

Passages from informative and persuasive text, with manipulation of coherence markers

- 1) Passage from informative genetic manipulation text, translated from Dutch (markers are underlined, [type of marker])

What are the consequences of genetic manipulation?[heading]

Different point of views exist on the pros and cons of genetic manipulation. [organizer]

The food industry says that genetic manipulation offers many advantages, since [connective] it enables us to start resolving world hunger. Moreover, [connective] genetic manipulation can help us fight illnesses in the future that are now incurable.

Environmental organizations say that genetic manipulation is an unacceptable change in nature, because [connective] it creates safety risks that the consumer is not fully aware of. An example of such a risk [lexical cue phrase] is public health. Consumers can accidentally swallow remainders of herbicide. (...)

In summary, [lexical cue phrase] genetic manipulation has good and bad sides to it. The future will show whether it/ genetic manipulation [referential coherence] is a beneficiary development or not.

224 – APPENDIX

2) Passage from persuasive genetic manipulation text (markers are underlined. [type of marker])

What are the dangers of genetic manipulation? [heading]

Different point of views exist on the pros and cons of genetic manipulation. [organizer]

The food industry says that genetic manipulation offers many advantages, since [connective] it enables us to start resolving world hunger and fight illnesses in the future that are now incurable.

Despite these advantages, genetic manipulation remains an unacceptable change in nature, because [connective] it creates safety risks that the consumer is not fully aware of. The food industry completely ignores these risks/ risks of genetic manipulation [referential coherence]. An example of such a risk [lexical cue phrase] is public health. Consumers can accidentally swallow remainders of herbicide. (...)

Therefore, there is [connective] only one clear answer to genetic manipulation: stop eating genetically manipulated food. Biological foods are safer for the environment, for agriculture and for the consumer.

3) Sorting Task

Here's a list of key words taken from the text. Make groups of words that you think should go together on the basis of the text. You can make as many groups as you want, and they can be of any size. Draw a circle for each group you want to make and put the right numbers in the circle.

solving the world hunger problem

crossing existing crops

moratoria

disease control

finances

resistance

food industry

health issues

allergy

1 7 4	5 6	8 9	3	2
-------	-----	-----	---	---

Appendix 4:

(Part of) Materials from Experiment 4 in Chapter 7

Examples of texts and corresponding statements: three verification statements (either true or false), three persuasion statements.

1) Even governor Bush/is driving a hybrid car,/combining electric and gas power./The gas engine works/when it is most efficient,/which is when the car is running at 20 mph or higher./At inefficient moments,/such as getting a car moving from a standstill,/the electric engine steps in./The gas engine operates/only in its near perfect window of efficiency,/thereby/burning substantially less/fuel than normal./That is why/these hybrid cars/are so interesting./They are environment-friendly/and at the same time/they save money on gas-expenses.

The electric engine only operates when the car is running 20mph or higher . F

When a car uses less fuel, it is good for our environment. T

When the gas engine only operates in efficient situations, it burns less fuel. T

Hybrid cars are environment-friendly.

I will seriously consider hybrid technology for my next car.

Hybrid cars cost less on fuel.

2) Florida politicians/have passed a law/letting people in Florida kill/in self-defence/on the street./As a consequence,/people do not have/to retreat when they/are being attacked./ But this new law/will bring a Wild West attitude/to Florida/magnet to hundreds of thousands of tourists./Now,/even more guns will be sold,/because/this law gives/gun owners/a license to kill./This could lead to very dangerous situations/and a change is needed in the way that/ Florida law deals with guns.

The new gun-law states that people under attack can use their guns. T

Because of this law, more guns will be sold. T

The new law is harmless. F

I think the new law is dangerous.

More accidents will happen now that this new law is passed.

There should be restrictions on gun use in Florida.

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/from state

226 – APPENDIX

universities 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/needs to be implemented/as soon as possible./All students will benefit from such a system./

A college exit exam ensures that students all possess the same basic skills. T

There are more disadvantages to a college exit exam than advantages. F

Implementing exit exams will have a positive effect on the quality of education. T

I think college exit exams are a bad idea.

If we have to give our opinions, I will vote in favor of college exit exams.

College exit exams need to be implemented.

4) Record labels are worried/that the rise of home CD-burners/has eaten into album sales./ Their efforts to protect CDs/against digital copying/are beginning to draw scrutiny/from lawmakers/concerned that the plans might violate the law/since/a 1992 law/allows music listeners/to make some personal digital copies/of their music./In return,/recording companies collect royalties/on the blank media/used for this purpose./That is why/copying your CDs/should be permitted./It is not the copying, but the copy protection that is illegal.

The 1992 law states that copy protection is forbidden. T

Since record companies are paid for all blank CDs, copying CDs is legal. T

You are allowed to copy your CDs for your friends. F

I think copying CDs should be permitted.

Copying CDs is the same thing as stealing.

I will no longer make copies for other people.

Curriculum Vitae

Judith Kamalski was born on September 2nd 1979 in Geldrop, The Netherlands. She attended *Philips van Horne Scholengemeenschap* in Weert, where she obtained her VWO diploma. In 1997, she started studying Corporate Communication at the *Radboud University Nijmegen*. In 1999-2000, she studied French language and literature (*Lettres Modernes*) at the *Sorbonne University* in Paris. After returning to Nijmegen, she graduated in 2002 in both Corporate Communication and French Language and Literature (cum laude).

In 2002, Judith started her PhD research at the *Utrecht Institute of Linguistics*. During her affiliation with the UiL-OTS, she taught a yearly course entitled Persuasive Documents for the *Department of Dutch Language and Culture*. In 2005, the Netherlands Organization for Scientific Research N.W.O. funded the research that she conducted at *Florida State University*, Tallahassee, U.S.A. Judith is currently working as a management trainee for the scientific publisher *Elsevier*.