TURUN YLIOPISTON JULKAISUJA
ANNALES UNIVERSITATIS TURKUENSIS

SARJA - SER. B OSA - TOM. 323

HUMANIORA

THE IMPORTANCE OF SEMANTICS

VISUAL WORLD STUDIES ON DRAWING INFERENCES AND RESOLVING ANAPHORS

by

Pirita Pyykkönen

TURUN YLIOPISTO
UNIVERSITY OF TURKU
Turku 2009

From the Department of Psychology
University of Turku, Finland

Supervised by

Jukka Hyönä, Ph.D
University Lecturer
Department of Psychology
University of Turku, Finland

Roger P.G. van Gompel, Ph.D
Senior Lecturer
School of Psychology
University of Dundee, UK

Reviewed by

Gerry T.M. Altmann, Ph.D
Professor
Department of Psychology
University of York, UK

Martin J. Pickering, Ph.D
Professor
School of Philosophy, Psychology and Language Sciences
University of Edinburgh, UK

Opponent

Gerry T.M. Altmann, Ph.D
Professor
Department of Psychology
University of York, UK

ISBN 978-951-29-4110-0 (Print)
ISBN 978-951-29-4111-7 (Pdf)
ISSN 0082-6987
Painosalama Oy – Turku, Finland 2009

THE IMPORTANCE OF SEMANTICS:

Visual World Studies of Drawing Inferences and Resolving Anaphors

Pirita Pyykkönen
Department of Psychology
University of Turku, Finland

ABSTRACT

The present thesis investigated the importance of semantics in generating inferences during discourse processing. Three aspects of semantics, gender stereotypes, implicit causality information and proto-role properties, were used to investigate whether semantics is activated elaboratively during discourse comprehension and what its relative importance is in backward inferencing compared to discourse/structural cues. Visual world eye-tracking studies revealed that semantics plays an important role in both backward and forward inferencing: Gender stereotypes and implicit causality information is activated elaboratively during online discourse comprehension. Moreover, gender stereotypes, implicit causality and proto-role properties of verbs are all used in backward inferencing. Importantly, the studies demonstrated that semantic cues are weighed against discourse/structural cues. When the structural cues consist of a combination of cues that have been independently shown to be important in backward inferencing, semantic effects may be masked, whereas when the structural cues consist of a combination of fewer prominent cues, semantics can have an earlier effect than structural factors in pronoun resolution. In addition, the type of inference matters, too: During anaphoric inferencing semantics has a prominent role, while discourse/structural salience attains more prominence during non-anaphoric inferencing. Finally, semantics exhibits a strong role in inviting new inferences to revise earlier made inferences even in the case the additional inference is not needed to establish coherence in discourse. The findings are generally in line with the Mental Model approaches. Two extended model versions are presented that incorporate the current findings into the earlier literature. These models allow both forward and backward inferencing to occur at any given moment during the course of processing; they also allow semantic and discourse/structural cues to contribute to both of these processes. However, while Mental Model 1 does not assume interactions between semantic and discourse/structural factors in forward inferencing, Mental Model 2 does assume such a link.

CONTENTS

ACKNOWLEDGEMENTS

Shortening my own text has always been one of the challenging tasks when polishing my articles. Here, I have tried to express my thanks as briefly as possible, but there are just too many people who deserve to be remembered and thanked for the part they played in bringing this journey to its end and even for getting it started in the first place.

To my supervisors, Dr. Jukka Hyönä and Dr. Roger van Gompel, I want to address my greatest thanks. I'm so extremely grateful for your guidance during these years. Jukka: I feel very honoured to have been able to learn eye-tracking research with such a great professional like you. Your endless patience for teaching and clarifying different things to me made my learning such an enjoyable experience. You also gave me space and I was willing to use it, as I felt secure knowing you were there and always ready to step in to help me whenever I needed any kind of advice. Roger: It's been so great to have had all those discussions with you during these years. I want to thank you for asking me to clarify my arguments over and over again, because without that, I wouldn't know as much about my own research as I do now. Warm thanks for always being there for me - whenever I needed advice, encouragement or new challenges (=hillwalking).

I want to thank Dr. Juhani Järvikivi and Dr. Danielle Matthews, my other collaborators on the studies. It's been such a great pleasure and joy to work with such inspiring and intelligent scientists like you are. I have learned a lot from you – and it's always been so much fun, too.

I'm really thankful to Prof. Gerry Altmann and Prof. Martin Pickering for agreeing to serve as reviewers of the thesis. It's hard to find a way to describe my feelings when I read those most beautiful and encouraging words I have ever heard about my research. I thank you for the insightful comments you made, which I know will also be a great help to me in my future research.

Thanks to the people who have discussed and commented on the papers and helped with the studies. I hope while reading this thesis you can be proud of how your contributions have become part of it: Prof. Matthew Crocker, Dr. Lynne Duncan, Dr. Martin Fischer, Dr. Yuki Kamide, Dr. Julie-Ann Marshall, Prof. Risto Näätänen and Dr. Ben Tatler. Thanks to Mark Buckley for reading through the whole thesis. I also thank all the participants who made the experiments possible.

I have had the great honour to be part of different research institutions during my Ph.D journey: Department of Psychology and Centre for Cognitive Neuroscience, University of Turku; School of Psychology, University of Dundee; School of Philosophy, Psychology, and Language Sciences, University of Edinburgh; and the Department of Computational Linguistics, Saarland University. I thank the people in all these institutions for always making it so easy and enjoyable for me to settle in to new places. Special thanks to Matt for believing in me and offering me my current post, even though I wasn't able to give him a more concrete promise about when I would complete my Ph.D other than "hopefully soon".

I thank the three-professor gang, Prof. Kimmo Alho, Prof. Heikki Hämäläinen and Prof. Risto Näätänen, for guiding me into the mysterious world of academia. Without all of those discussions I would be much less educated now, and this journey would have been much less fun without all those jokes, too.

I also want to express my gratefulness to Prof. Jussi Niemi who originally pointed me in the direction of psycholinguistics. The knowledge I learned while working with you and at your courses along with the courses offered by Prof. Pirkko Muikku-Werner and Prof. Stefan Werner in Joensuu has turned out to have been extremely important during these years.

There are many other colleagues who have been there in the office, at labs, conferences, summer schools, pubs, lunches, coffee breaks and in email/skype from a distance: Afra, Andriy, Berry, Christoph, Claudia, Emilia, Garance, Gunnar, Hamutal, Helene, Jens, Johanna, Judith, Juliane, Jyrki, Katja, Kumiko, Manabu, Manon, Maria, Maria-Nella, Mia, Outi, Richard, Robin, Sarah, and Seppo. My apologies for those people I haven't mentioned here: there are simply more dear friends and great memories than I could ever hope to be able to capture in this list.

During these years, I have presented my research at various meetings: AMLaPs, CUNYs, EPS, ESP (now ESLP), IASCL, SALC, SWAET and at the seminars in Dundee, Glasgow, Manchester, Saarbrücken, Trondheim, Turku and the MPI in Nijmegen. I am grateful to those audiences for their well-aimed questions and comments.

My Ph.D studies were financially supported by the Graduate School of Psychology in Finland and the grants from the Academy of Finland (no. 213731), the Finnish Cultural Foundation and Jenny & Antti Wihuri Foundation. Parts of the articles and the writing-up were carried out while working in Edinburgh and Saarbrücken. I received conference support from the CUNY conference, the Turku University Foundation and Psychology department at the University of Turku. Here, I want to express my thanks to these institutions.

Sydämelliset kiitokset äidille ja isille, Arja ja Seppo Pyykköselle ehtymättömästä kannustuksesta ja tuesta. Kiitokset tuesta ja hauskanpidosta myös Tuikulle, Jukalle, Terhille, Viivi-Lotalle, Jounille ja kouluajan ystäville. Juhani, miten paljon hienoja muistoja näiltä vuosilta onkaan kertynyt! Kiitos, että jaksoit kannustaa ja uskoa myös niinä hetkinä kun itse en nähnyt valoa.

Saarbrücken, November 2009.

Pirita Pyykkönen

LIST OF ORIGINAL PUBLICATIONS

The dissertation is based on four original articles. The articles are referred to in the text by the following Roman numerals:

I Pyykkönen, P., Hyönä, J., & Van Gompel, R.P.G. (2009). Activating gender stereotypes during online spoken language processing: Evidence from visual world eye-tracking. *Experimental Psychology*. Manuscript in press.*

II Pyykkönen, P., & Järvikivi, J. (2009). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*. Manuscript in press.*

III Pyykkönen, P., Matthews, D., & Järvikivi, J. (2009). Three-year-olds are sensitive to semantic prominence during online language comprehension: A visual world study of pronoun resolution. *Language and Cognitive Processes*. Manuscript in press.**

IV Pyykkönen, P., Van Gompel, R.P.G., & Hyönä, J. (2009). The importance of semantics and salience during anaphoric and non-anaphoric inferencing. *Manuscript.*

ABBREVIATIONS

ERP = Event related potentials
NP = Noun phrase
OBJ = Object
OVS = Word order: object-verb-subject
SUBJ = Subject
SVO = Word order: subject-verb-object
VP = Verb phrase

1 INTRODUCTION

The important question of how people manage to comprehend language quickly and effortlessly has attracted psycholinguists for a long time. It is commonly known that language comprehension does not only involve extracting the literal meaning of words in the text, but also involves drawing different types of inferences based on information that is not literally stated in the text (e.g., Clark, 1977; Haviland & Clark, 1974; Garnham, 2001). However, despite decades of research, there is still no clear consensus on what kinds of inferences are drawn during online language comprehension and how different factors contribute to and compete with each other during inference generation. Importantly, the current theories have failed to capture the relative importance of semantics compared to discourse/structural cues during inferencing and anaphoric processing. Limitations of the previously used methods may partly explain why these questions have remained unresolved.

In the research conducted for this thesis, I took advantage of a tight temporal coupling between spoken language comprehension and people's visual attention to scenes related to the sentences that they hear. I used visual world eye-tracking, that involves measuring listeners' eye movements to visually depicted pictorial stimuli while they are listening to spoken language, as a method to study what kinds of inferences are drawn online and how strong the effect of semantics is during inferential and anaphoric processes. I studied three dimensions of semantics: gender stereotypes, implicit causality and protoroles. Based on my findings, I argue that the language processor does not only operate in an optimal way by allowing both forward and backward inferencing at any given moment during the course of processing, but also weighs semantic and discourse/structural cues in such a dynamic way that has remained unexplained in current theories of inferencing.

The thesis consists of four original journal articles. In this overview, I set the research questions investigated in those studies in a wider theoretical context than was done in the original publications. Therefore, I will provide an extensive discussion of previous experimental literature on gender stereotypes, implicit causality and protoroles. In addition, I will review the experimental methods used in earlier studies of inferences and suggest reasons why these methods have failed to provide evidence for the hypotheses proposed by the current theories or have failed to disentangle competing hypotheses. Finally, after reviewing all four studies, I highlight what was learned about the importance of semantics during online inferencing and anaphor resolution, and what kinds of aspects should be taken into account when extending or revising the current processing theories. Finally, I propose two extended mental models to explain inference generation during online processing.

1.1 Bridging and elaborative inferences in discourse

In language comprehension, the definition of the term *inference* that is adopted in this thesis is a process in which a comprehender encodes information that is not

explicitly stated in the text (e.g. Clark, 1977; Haviland & Clark, 1974; McKoon & Ratcliff, 1989; Potts, Keenan, & Golding, 1988). Inferences can be divided into two groups based on the direction they are drawn during comprehension: *bridging inferences* are made in backward fashion while *elaborative inferences* are made in forward fashion (Clark, 1977; Clark & Haviland, 1977). For example, when people hear a pronoun *he* referring back to the earlier mentioned character in (1), the referent is assigned in backward fashion. Pronouns are one type of anaphors: the term *anaphora* refers to a relationship in which two linguistic elements refer to the same concept and the interpretation of one concept (*an anaphor*) is determined by the interpretation of the other concept (*an antecedent*) (e.g., Wasow, 1986). Thus, in example (1) the pronoun *he* and its antecedent *chimney sweep* refer to the same entity in a real world.

(1) The chimney sweep was climing onto the roof. He was worried whether it could be slippery.

Even though pronouns and anaphors in general are typical examples and the most widely studied types of bridging inferences, people also make inferences in backward fashion when such an explicit linguistic marker as an anaphor is not present in language. Clark (1977) described different types of bridging inferences in which the reference is justified as it is related to another entity or entities in the previous discourse context. Consider for example the sentences in (2). In the latter sentence, when people encounter the word *engine*, they might make a bridging inference that the person is talking about the same car as in the first sentence. This bridging inference is drawn by making use of the background knowledge that cars tend to have an engine (see also McKoon, Gerrig, & Green, 1996).

(2) John bought a new car. The engine needed tuning.

Similarly to backward inferences, there are different types of inferences that are drawn in forward fashion in language comprehension. Traditionally, elaborative inferences are used to define certain strict predictions people make based on constraining contexts (e.g., Keefe & McDaniel, 1993; McDaniel, Schmalhofer, & Keefe, 2001). For example, reading the first sentence in (3) may constrain people to make elaborative inference that *the report was burned,* and therefore, the latter sentence has been found to be easy to comprehend and integrate to the previous context.

(3) The spy threw the report in the fire. The ashes floated up the chimney.

In addition to the strict elaborative inferences, it is also suggested that people selaboratively activate other types of information such as gender stereotypes (e.g., Oakhill, Garnham, & Reynolds, 2002; Sanford, 1985). For example, in (1) people may make such an inference when they hear the word *chimney sweep* by inferring

that chimney sweeps stereotypically refer to males. In this case, the inference is not made to refer to any particular consequence of the action or event such as described in the example (3); instead, the activation is more generic in nature.

Based on the direction of inferences drawn in discourse, bridging inferences are also often called *backward inferences* and elaborative inferences *forward inferences*. In this thesis, I use the term 'bridging inference' or 'backward inference' to denote all kinds of inferences people draw in backward fashion and the term 'elaborative inference' or 'elaborative activation' to denote both types of elaborative inferences: strict predictions or general activation. The exact nature of the inference is specified when necessary.

1.2 Inferences and semantics in comprehension theories

In psycholinguistics, several different discourse comprehension theories have been proposed. One important difference between the theories is what kinds of inferences they assume are drawn during online comprehension. However, many theories are somewhat ambiguous on the question of what kind of information results in inferences. More specifically, what is the role of semantic cues and how strong are these cues compared to discourse/structural cues? In what follows, I will review theories that make assumptions about either forward activation or backward use of semantic information (or both), and the relative importance of semantics compared to discourse/structural cues during inferencing.

1.2.1 Mental model approaches

Different versions of the Mental Model theories make important claims about both inference generation and the role of semantics in inference generation. Mental Model approaches assume that language comprehension consists of constructing mental or imaginary models; thus, they aim to unite language processing with thinking and reasoning in general (Garnham, 1993, 2001; Johnson-Laird, 1983, 1994). These theories are built on the view that there is a direct link between perception and language: People use language to talk about their perception of the world (obtained through their primary sense, vision), and it is assumed that the mental representation of the perceived and the described worlds are the same (Miller & Johnson-Laird, 1976). In current forms of the theories, the representations are seen as mental models of real or imaginary situations of these worlds (Garnham, 2001). In other words, when people are talking or thinking about the world, they represent situations in the real world, and when people are talking or thinking about the fictional worlds or beliefs, they represent situations of imaginary worlds.

Mental Model theories assume that language comprehension occurs incrementally so that every new piece of information is continuously integrated into previous information during comprehension. In this respect, the modern Mental Model theories (described below) are closely related to earlier discourse comprehension theories such as Kintsch's (1988, 1998; Kintsch & Van Dijk, 1978) Construction-Integration

theory. This theory explains how people derive meaning from the discourse, i.e., what kinds of representations become encoded from the text and how that encoding happens during the course of comprehension. Unlike the other Mental Models however, this model assumes two processing stages in which mental representations are constructed from the literal propositions in the text. More precisely, the first stage, *construction*, includes inferences from local text-based information such as word, meanings, their associations and propositions derived from word meanings. This level allows several kinds of inferences, both elaborative and bridging inferences, to be created in parallel. Those elaborative inferences that turn out to be irrelevant are then deactivated in the second, *integration* stage that occurs after the connections described above are constructed. In the integration stage, those propositions and inferences that support the global discourse representation strengthen each other and other propositions and inferences are inhibited. When exactly integration occurs, can vary: According to the early version of the theory (Kintsch, 1988), this stage was assumed to take place at sentence boundaries and thus did not assume incremental word-by-word integration, as is the case with Mental Model theories (Garnham, 2001; Zwaan & Radvansky, 1998).

One part of Kintch's theorising is the situation model that comprehenders build from the linguistic input. This situation model is similar to the type of mental model that is assumed to be constructed by the current Mental Models approaches (Garnham, 1993, 2001; Zwaan & Radvansky, 1998). When reading or listening to language, comprehenders build a mental representation of the discourse by imagining how the events described in the discourse have occurred or would occur in the real or imaginary world. It should also be noted that this type of approach is similar to the kind of situation models that are assumed by embodied language comprehension models (Zwaan, 2004; Zwaan & Madden, 2004; Zwaan & Taylor, 2006; Zwaan & Radvansky, 1998; Barsalou, 1999). These models assume that people not only create a situational representation of the events during language comprehension but also simulate the events mentally. It is also suggested that such mental simulations are a compulsory part of language comprehension, and without such simulations comprehension processes are partly inhibited (see e.g., Fischer & Zwaan, 2008, for a recent discussion about the nature of simulations during comprehension).

Inferences and semantics. From the inferential processing perspective, Mental Model theories provide testable hypotheses to the questions of (1) what kinds of inferences are drawn online and (2) what the role of semantics is during these processes. They suggest that both backward and forward inferences are drawn online. While online drawing of backward inferences seems fairly uncontroversial, as will be shown in the next section, forward inferences are not assumed to be drawn online by theories opposing the Mental Model theories, such as the Minimalist Account (McKoon & Ratcliff, 1986, 1988, 1989, 1992).

Mental Model theories allow several types of forward inferences to be drawn online. For example, when people encounter the word *car* in a discourse, they may infer that cars have an engine. In addition, the background knowledge that is used in such inferencing can be either knowledge of specific things (such as this particular car has an engine) or a general knowledge (all cars have engines) (Clark, 1997; Garnham, 2001). In a similar vein, it is hypothesised that when people encounter a word such as

secretary or *chimney sweep*, they might infer that the former typically refers to female characters while the latter refers to male characters (Garnham, Oakhill, & Reynolds, 2002; Oakhill et al., 2005; Reynolds, Garnham, & Oakhill, 2006; Sanford, 1985). Analogically, it could be hypothesised that other types of semantic information can also be inferred elaboratively. For example, when people encounter interpersonal verbs such as *frighten* and *fear* that are found to bias resolution of the pronoun *he* in a sentence such as *John frightened/feared Bill, because he…* toward NP1 (frighten) or NP2 (fear), it may well be that such biasing information is elaboratively activated when hearing the verb, and the activation is not postponed until the pronoun is encountered and needs to be resolved (see Chapter 1.3.2 for a detailed description of these verbs and their biases).

According to Mental Model theories, forward activations occur because these models assume that interpretation is not only driven by explicitly stated information in the discourse and the propositions created between text parts, but instead, people also activate information related to each word in context by using general world knowledge and previous personal experience. After this information is activated, it is assumed to become quickly integrated into the current situation model, and the model is continuously updated during the course of processing based on this new information (Garnham, 2001; Johnson-Laird, 1983; Zwaan & Radvansky, 1998). This is in line with the incremental view of language comprehension, according to which every new piece of information is integrated into the previous context as soon as this information becomes available (e.g., McRae, Spivey-Knowlton, & Tanenhaus, 1994; MacDonald; Trueswell, Tanenhaus, & Garnsey, 1994). Importantly, the incremental approach held by Mental Model theories allows several types of information - including different dimensions of semantic information - to be used in inferences in order to continuously update the situation model. Unlike in some other processing models, as will be seen below, Mental Model theories assume that semantic information affects the generation of inferences.

1.2.2 Other theoretical approaches

Unlike the Mental Model theories, other theories have set more constraints on what kinds of inferences are drawn online and how different types of structural and semantic cues contribute to inference generation. Such theories include, for example, the Minimalist Account proposed by McKoon and Ratcliff (1986, 1988, 1989, 1992) and a set of more structurally motivated theories like Structure Building Theory (Gernsbacher, 1989, 1990; Gernsbacher & Hargreaves, 1988), Centering Theory and Discourse Prominence Theory (e.g., Gordon & Hendrick, 1998; Gordon, Grosz, & Gilliom, 1993; Grosz, Joshi, & Weinstein, 1995) as well as the Two-Stage Model of pronoun resolution (Corbett & Chang, 1983; Garrod & Sanford, 1994; Garrod & Terras, 2000).

Minimalist account. McKoon and Ratcliff (1986, 1988, 1992) have proposed the Minimalist Account of inference generation during reading. The model assumes, in contrast to Mental Models, that people do not fully represent the situation that can be inferred on the basis of the text. Instead, according to the Minimalist approach, only

those inferences that are necessary for local textual coherence are routinely made during online processing. Based on this assumption, it can be argued that forward inferences are not drawn online: at the point when they are drawn during comprehension, forward inferences do not contribute to the local coherence of the text: For example, inferring gender stereotype information of the word *secretary* in a sentence *John called his secretary* does not help to understand the meaning of the sentence. This assumption thus implies, for example, that gender stereotype information is not activated when people encounter a gender-stereotype word in the discourse. This assumption is clearly in contradiction to the assumptions of the Mental Model theories, as discussed above.

An important question is under what circumstances inferences are made online. According to the Minimalist Account, when people hear an anaphoric expression, they need to resolve it in order to build a coherent interpretation. Therefore, if the pronoun is ambiguous and activation of semantic information such as gender stereotypes of its potential antecedents may be helpful for resolving the pronoun, it is assumed that comprehenders activate and use this information during online language comprehension.

In addition, the Minimalist Account assumes that information that is readily available from short-term memory becomes activated during online comprehension. The term of 'readily available' is not clearly defined and thus somewhat open to interpretation as noted for example by Garnham (2001). However, McKoon and Ratcliff (1992) followed Kintsch and Van Dijk (1978) in assuming that information that is readily available in short-term memory contains the explicitly stated words in the current or immediately preceding context and the propositions derived from them. Based on this definition, it can be argued that this model assumes that information that is not part of the words' denotative meaning is not activated during online comprehension. One example of such information is gender stereotypes: they are not part of the word's denotative meaning and cannot be determined directly from the propositions derived from the discourse. Thus, the Minimalist Account does not predict elaborative activation of semantic information such as gender stereotypes.

Structurally motivated approaches. The following approaches are mainly focused on the pronoun resolution instead of dealing with inferences in general. In her Structure Building Theory, Gernsbacher (1988, 1989, 1990; Gernsbacher & Hargreaves, 1989) proposed that a coherent mental representation of the "structure" of a sentence or a discourse is built by first laying a foundation (a mental frame) on which all the subsequent information is mapped. In Structure Building Theory, the first-mentioned entity in a sentence or discourse is selected as the starting point for the first layer as it becomes available first and is thus accessible during the initial processing stage. Subsequent information is then mapped onto this representation. Thus, the first-mentioned entity carries a special role in structure building.

Although the privileged role of first-mentioned character is often seen as a general cognitive principle rather than a purely structural (linguistic) feature, I have categorised this approach as a "structurally motivated approach". The reason is that in this model the most relevant property is the surface structure: the first-mentioned entity has a privileged role regardless of other structural, semantic or thematic

properties (Gernsbacher & Hargreaves, 1989). The first-mentioned advantage is often found in pronoun resolution studies, as can been seen in Chapter 1.3: An ambiguous pronoun is often resolved to refer back to the NP1 rather than the NP2 (see also Gernsbacher, 1988; Corbett & Chang, 1983; Von Eckhardt & Potter, 1985).

A different structural approach is provided by Centering Theory, which aims to explain pronoun resolution based on how attention is centered around certain entities in a sentence. This approach builds upon a prominence hierarchy that is defined on the basis of surface features of a sentence (see (4), Gordon et al., 1993; Gordon & Hendrick, 1997a,b; Grosz & Sidner, 1986; Grosz et al., 1995).

(4) subject > object > oblique/other

Thus, unlike Structure Building Theory, Centering Theory assigns importance to grammatical roles rather than to the surface position of entities in a sentence. Grammatical role has been found to be an important factor when resolving pronouns. More precisely, people tend to resolve an ambiguous pronoun as referring back to the grammatical subject rather than to the grammatical object of the previous context (see Chapter 1.3). Later, Gordon and Hendrick (1998) modified the theory by proposing an extended model called Discourse Prominence Theory. This theory was created to integrate different referring expressions into the earlier proposed model. However, this model also took into account the earlier findings suggesting that grammatical role is not the only important surface feature that contributes to the pronoun resolution by including both order-of-mention and grammatical role into the prominence ranking. Based on a modified hierarchy, Discourse Prominence Theory assumes that entities mentioned earlier in a sentence are more preferred than those mentioned later in the sentence and that the grammatical subject is more preferred than the grammatical object.

Importantly, neither of the above structural approaches takes into account the semantics or the thematic role of potential antecedents presented in the preceding linguistic context. Thus, the effects of verb semantics in studies of pronoun resolution with sentences such as *John frightened/feared Bill, because he…* are difficult to explain with these approaches: These studies have shown that the pronoun is resolved based on the NP1- and NP2-biasing information of the verbs (see e.g., Gimenes, Caplan, & Rigalleau, 2006, for a recent review of these studies): Following the verb *frightened* (NP1-biased verb) the pronoun is resolved to refer to *John* while following the verb *feared* (NP2-biased verb) the pronoun is resolved to refer to *Bill* (see Chapter 1.3.2 for a more detailed description of these studies). These findings clearly show that semantic information presented prior to the pronoun plays an important role in searching for the antecedent for the pronoun. However, as noted above, none of the above structural approaches succeeds in explaining these findings.

Interestingly, another structural approach, the Two-Stage Model of pronoun resolution, may potentially explain (some of) these results. This model assumes that semantic information is activated and used late during the course of processing (Corbett & Chang, 1983; Garrod & Sanford, 1994; Garrod & Terras, 2000). The

model sets strict constraints on the type of information that is used during the first stage of processing, but not on the information used during the second stage. In the first stage, the candidate antecedents are activated based on constraints such as number, grammatical gender and other structural features. Only in the second stage, contextual information and general world knowledge is used to restrict the number of potential antecedents. Of these, the former stage is often called *bonding* and the latter *resolution* (see e.g., Garrod & Terras, 2000). This approach suggests that finding a correct referent for an unambiguous anaphoric expression occurs in the first stage by using the structural constraints described above as cues and is validated in the second stage, while processing of ambiguous anaphoric expressions always requires the resolution stage, during which the most plausible referent is selected. It is this stage that could explain semantic effects found in the resolution of an ambiguous pronoun following NP1- and NP2-biasing verbs: In the first stage the first-mentioned subject is preferred as an antecedent and only in the second stage the semantically most plausible antecedent is selected.

The structural theories described above set clear constraints on antecedent selection in backward inferences and the use of discourse/structural and semantic factors in these inferences. The main conclusion is that either semantic factors do not contribute to pronoun resolution at all or only during later stages of processing. These approaches face a challenge in explaining empirical findings of early semantic effects in pronoun resolution (these findings are detailed in the following section). Moreover, these approaches are intended to explain backward inferences rather than forward inferences. Thus, in what follows, I will refer back to these models only when discussing semantic effects in backward anaphoric processing.

Despite the differences between the theories presented above, both Mental Models Theory and the several alternative approaches are supported by experimental evidence. For example, one line of studies suggests that forward inferences of specific causes, continuations or consequences are drawn during online processing supporting Mental Models (Calvo & Castillo, 1996, 1998; Calvo, Meseguer, & Carreiras, 2001; Fincher-Kiefer, 1993, 1995, 1996; Keefe, & McDaniel, 1993; Klin, Murray, Levine, & Guzmán, 1999; Millis, Morgan, & Graesser, 1990; Murray, Klin, & Myers, 1993; Myers, Shinjo, & Duffy, 1987; Whitney, Richie, & Crane, 1992). However, although these studies suggest that forward inferences are drawn online, some studies suggest that although they are activated, they might also become quickly deactivated or are maintained in memory only for a limited time (Fincher-Kiefer, 1995, 1996; Keefe, & McDaniel, 1993), and their activation may be dependent on a highly constrained context (Klin, Murray, Levine, & Guzmán, 1999). Contrary to the assumptions of Mental Models Theory, another line of research suggests that forward inferences are not made online (Bloom, Fletcher, van den Broek, Reitz, & Shapiro, 1990; Duffy, 1986; Graesser, Haberlandt, & Koizumi, 1987; Magliano, Bagget, Johnson, & Graesser, 1993; Potts et al., 1988; Singer, & Ferreira, 1983) or the inferences are just minimally encoded (e.g., McKoon, & Ratcliff, 1986, 1989) or their activation is strategic in nature and thus may not necessarily be drawn online (Allbritton, 2004).

As becomes evident from above, there is no general agreement on what types of inferences are drawn online. The Mental Model approaches are most flexible in this manner, assuming that both forward and backward inferences are drawn, while the

Minimalist Account sets the most constraints on the types of inferences that are drawn online and is the most explicit about under what kinds of conditions they are made. Furthermore, the above theories make different predictions about the cues used for anaphor resolution (a form of backward inferencing), but what is common to all of them, they assume structural factors to have an early effect in anaphor resolution. Importantly, in most theories the role of semantic cues in both forward and backward inferencing has remained underspecified. Therefore, by examining the use of different aspects of semantics during inferencing and by comparing their use with discourse/structural factors, it is possible to take a step toward a more comprehensive picture of inference generation and the importance of semantics during language comprehension. In what follows, I review the experimental literature on semantic effects on forward and backward inferencing and show how important questions have remained unresolved.

1.3 Semantic factors in online inferences and anaphors

When people build coreference relationships between text parts, it has been noted that people use different aspects of semantic information during inferencing (e.g., Garnham, Oakhill, & Cruttenden, 1992; Garnham, Traxler, Oakhill & Gernsbacher, 1996; Koornneef, & Van Berkum, 2006; Stewart, Pickering, & Sanford, 2000), but the question of when during the course of processing the semantics is used has remained unresolved. One central question has been whether people use semantics only when processing backward inferences or whether semantic information is also activated elaboratively in forward fashion. In the thesis I tested the following dimensions of semantic information during backward inferencing and elaborative activation: gender stereotypes of role names and nouns, implicit causality bias of verbs and proto-role properties of agent and patient introduced by transitive verbs. As will be highlighted below, these dimensions provide interesting test cases in order to resolve the unanswered questions in the literature. Even though several hypotheses of the backward and forward use of semantic information in inferencing are proposed, the earlier studies have failed to answer these hypotheses. One reason has been the limitations of the selected methods used in those studies.

1.3.1 Gender stereotypes

The term *gender stereotype* refers to commonly shared information or belief that certain features are more common among males than females or among females than males. In language input, these beliefs can be inferred from the certain words; importantly however, gender stereotypes are not encoded in the denotative meanings of those words. For example, hearing the sentence *John called his secretary* does not state that the secretary is a woman, and assuming this is not needed in order to understand the sentence meaning. It is not yet known whether people activate elaboratively that secretaries are stereotypically female while hearing a sentence like the one above. However, when people subsequently read a sentence starting with a pronoun *she*, it is found that gender stereotype information is quickly activated and

used to resolve the antecedent of the pronoun (Duffy & Keir, 2004; Kennison & Trofe, 2003; Kreiner, Sturt, & Garrod, 2008; Osterhout, Bersick, & McLaughlin, 1997; Sturt, 2003a,b). This evidence suggests that gender stereotype information is activated when integrating an anaphor to the previous context in backward fashion. Importantly also, these studies are in line with the assumption that gender stereotypes can be activated elaboratively. However, as can be seen below, the methods used have not allowed unequivocal confirmation of this hypothesis.

Backward use of gender stereotypes. There is much evidence for backward use of gender stereotype information during inferencing. For example, using a word-by-word self-paced reading task in English, Kennison and Trofe (2003) found that when a female gender stereotype noun (*secretary*) was presented as an antecedent for the following pronoun, reading times were prolonged late in the second sentence when the pronoun was inconsistent with the stereotypical gender of the role name such as in Example (5), in comparison to a condition in which the pronoun was consistent with the gender stereotype (6). The study showed that gender stereotype information is used in pronoun resolution, but it failed to disentangle whether this information was inferred when encountering the word *secretary* or only when the pronoun needed to be resolved and integrated into the previous context.

(5) The secretary distributed an urgent memo. **He** made it clear that work would continue as normal.

(6) The secretary distributed an urgent memo. **She** made it clear that work would continue as normal.

Interestingly, the stereotype activation did not cause prolonged reading times at the mismatching pronoun itself, but instead later in the sentence. This finding suggests that gender stereotype information may require a certain time to become activated during online comprehension. Contrary to this finding, other studies have suggested that activation of gender stereotypes during backward inferences may be quick, and prolonged reading times for stereotypically inconsistent pronouns may be found at the pronoun itself (Duffy & Keir, 2004; Kreiner, Sturt, & Garrod, 2008; Sturt, 2003a,b). For example, in a study of English isolated sentences, Duffy and Keir (2004) found prolonged reading times for the reflective pronoun *herself* when it was inconsistent with the gender stereotype of the antecedent (e.g., the male-stereotype word *electrician*) compared to a stereotypically consistent role name. This result indicates that stereotype information is activated and used quickly in anaphor resolution.

Duffy and Keir (2004) also showed that when participants were presented with a context in which the gender of the role name was explicitly indicated (see Example 7), prolonged reading times in the stereotypically inconsistent condition were found already at the sentence-final noun (*woman*) in the second sentence compared to the stereotypically consistent condition (*The electrician was a cautious man***)** (see also Irmen, 2007). In addition, when the gender was specified by the context, stereotype

information did not affect reading times of the pronoun in the last sentence. This showed that when gender information was provided explicitly in the context, the earlier activated gender stereotype information was not used in later pronoun resolution.

(7) Jeff's power had been unreliable ever since the tornado. **The electrician was a cautious woman** who carefully secured **her** ladder to the side of the house before checking the roof. Jeff suspected that high winds had loosened the connection to the power lines. **The electrician taught herself** a lot while fixing the problem.

Activation of gender stereotypes during backward inferencing has also been found in electrophysiological studies by using for example event related potentials (ERP) that measure voltage differences of electric pulses on the scalp produced by a certain event such as the presentation of a target word in a sentence. For example, Osterhout et al. (1997) showed in a reading task in English that when a gender-marked reflexive pronoun was inconsistent with the gender stereotype information of its antecedent such as in (8), the brain produced a positive peak in the brain waves around 600 ms following the target onset (i.e., the so-called P600), when compared to a stereotypically consistent sentence (9).

(8) The doctor prepared **herself** for the operation.

(9) The doctor prepared **himself** for the operation.

Interestingly, the gender mismatch resulted in a P600 effect, which is typically found for ungrammatical structures, and not a N400 effect, which is typically found for semantically anomalous materials or for other types of semantic violations (see Kutas & Federmeier, 2007, for a recent overview of ERP components related to language comprehension). Later research has suggested that the ERP component associated with gender mismatch is not a result of the brain treating the violation as ungrammatical; instead the nature of the agreement violation should be seen as a violation of *syntactic preference* (Hagoort & Brown, 1999; Van Berkum, Brown, & Hagoort, 1999a,b). Violation of syntactic preference occurs in situations in which the sentence is syntactically well-formed, but where syntactic properties fail to satisfy the currently built, preferred analysis of that sentence. Regardless of the exact nature of this effect, the finding suggests that the brain activates gender stereotypes quickly during backward inferencing.

Taking both behavioural and neural evidence together, it can be hypothesised that gender stereotype information has an important role in backward inferencing. In addition, gender stereotype information may be inferred quickly when useful for backward inferencing (cf. Kennison & Trofe, 2003).

Forward activation of gender stereotypes. A challenging issue related to the activation of gender stereotypes has been the question of whether gender stereotypes

are activated elaboratively in a forward fashion. In some of the above studies it has been suggested that the rapid effects of gender stereotypes during integration may be a result of activation of gender stereotypes already at the stereotype word (see the discussion in e.g., Duffy & Keir, 2004). However, the methods used do not allow us to conclusively differentiate whether gender stereotype information becomes activated in a backward fashion at the point when needed for integration, e.g., when a pronoun or another referent needs to be attached to the previously mentioned character, or also elaboratively following the gender stereotype noun.

Sanford (1985) asked participants to read stories that included a male stereotype noun such as *surgeon* that later turned out to be a female person. He found that participants found it hard to understand the story. By using English stories such as (10), Reynolds et al. (2006) found that when readers encountered gender stereotypically inconsistent information, they slowed down their reading and sometimes even failed to interpret the identity of the surgeon (as indicated by an offline questionnaire administered after participants had read the story).

(10) This morning a **father** and his **son** were driving along the motorway to work, when they were involved in a horrible accident. The **father** was killed and the **son** was quickly driven to the hospital severely injured. When the boy was taken into the hospital a passing ***surgeon*** exclaimed: "Oh my god, that is my **son**!"

Reynolds et al (2006) argued that the increased reading times reflect elaborative inferencing. However, they measured reading times for the whole sentence, so it is unclear how rapidly such inferences are made. It is possible that the stereotype information is activated already when encountering the stereotype noun that would indicate elaborative activation. However, these studies have not been able to detect forward activation directly, as the stereotype effect was found only at the point where coherence in the discourse needs to be established: e.g., when the sentence containing the word *son* (10) has to be integrated to the preceding story.

In a similar vein, Carreiras, Garnham, Oakhill and Cain (1996) observed a reading difficulty when gender stereotype information was inconsistent with the morphological gender of the noun's determiner in Spanish. For example, sentences such as (11) that included words in which the gender-marked article mismatched with the stereotype nouns (*la carpintera* 'carpenter'; a male-biased word with the feminine article + ending) took longer to read than sentences such as (12) in which the word did not cause such mismatch with its determiner (*el carpintero*; a male-biased noun with the masculine article + ending).

(11) **La carpintera** tomó las medidas para hacer el armario. Era un encargo bastante urgente. Ella tenía que terminarlo en el plazo de una semana.

'The carpenter took measurements to make the cupboards. It was quite an urgent order. She had to finish it in the space of one week.'

(12) **El carpintero** tomó las medidas para hacer el armario. Era un encargo bastante urgente. El tenía que terminarlo en el plazo de una semana.

'The carpenter took measurements to make the cupboards. It was quite an urgent order. He had to finish it in the space of one week.'

Based on these findings, Carreiras et al. (1996) suggested that stereotype information is activated when encountering the stereotype noun. However, because they used sentence reading times as their measure, it was not possible to determine when exactly in the sentence a reading difficulty occurred. As the stereotype noun was located in the sentence-initial position, the reading difficulty may have occurred during this sentence-initial noun or shortly after it, but it may also have occurred later in the sentence. Because of the coarse-grained reading time measure, the study was not able to determine whether activation occurs at or shortly following the stereotype noun.

Interestingly, Kreiner et al. (2008) found only marginally significant late effects of gender stereotypes when using cataphoric expressions in which the gender-marked pronoun preceded the gender stereotype noun, such as *After reminding herself about the letter, the minister...* The stereotype effect was found only in the regression-path measures of the region that followed the gender-stereotype noun and not in any measures of the stereotype word region. When the stereotype noun was replaced by a definitional gender word such as *the king*, the mismatch effect was immediate. Kreiner et al (2008) argued that this finding shows that processing definitional and stereotypical gender is qualitatively different: while definitional gender is derived from the lexical entry itself, stereotypical gender is accessed via inferential processing.

Besides studies investigating online sentence and discourse processing, studies on lexical processing have suggested that gender stereotypes are activated elaboratively. For example, Banaji and Hardin (1996) found that when people were asked to judge the gender of the English pronouns *she* and *he*, judgements were faster when they were primed by stereotypically consistent (*doctor – he, nurse – she*) than inconsistent words (*doctor–she, nurse–he*) (see also Cacciari & Padovani, 2007, for converging evidence in Italian). A similar pattern of results was found when people were asked to judge whether the word appearing on the screen was a pronoun or not: when the pronoun was preceded by a stereotypically consistent prime word, judgements were faster than when it was preceded by a stereotypically inconsistent prime word. In addition, Oakhill et al (2005) found a similar pattern of results in a study in which people were asked to make an explicit link between the prime and target by asking whether terms such as *surgeon-brother* or *surgeon-sister* could refer to the same person. Even in a study in which participants were encouraged to pay attention to unambiguous lexical gender relations such as *landlady-sister* and were told that some occupations were not clearly marked for gender but could be held either by men or women (i.e., in the cases of gender stereotype occupations), they were less likely to accept stereotypically inconsistent than consistent pairs and response times for the inconsistent pairs were also prolonged compared to consistent pairs. Oakhill et al. therefore proposed that activation of gender stereotypes might be difficult or even

impossible to suppress. Although these studies clearly show that gender stereotype information is activated and has an effect on response accuracies and times, activation was studied with isolated words, and therefore, the results cannot directly be generalised to elaborative activation of gender stereotypes in a discourse context.

Relative importance of gender stereotypes. An additional aspect of backward use of gender stereotypes is their relative importance compared to other possible discourse/structural cues during inferencing. In the above examples (7-9), gender stereotype activation was studied by using structures that included only one possible antecedent for the pronoun - either stereotypically matched or mismatched. In addition, although it could be argued that prolonged reading times (or P600) would not have been found if gender stereotype information were not activated, these studies did not directly test how strong the gender stereotype cue is compared to other discourse/structural cues. This question can be addressed, for example, by using Finnish gender ambiguous anaphors (e.g., the third person pronoun *hän* 's/he'), which are used to refer to both male and female referents. In a context in which there are two possible antecedents, one structurally preferred and one stereotypically preferred, it is possible to shed light on the relative importance of gender stereotypes compared to discourse/structural cues during backward inferencing. It may well be that both discourse/structural and gender stereotype cues are used – as predicted on the basis of previous studies – but their relative importance may vary depending on the type of inference drawn; for example, in certain situations structural cues would be preferred over semantic cues, or vice versa (Clark & Haviland, 1977; Garrod, O'Brien, Morris, & Rayner, 1990; Greene, McKoon, & Ratcliff, 1992; Fincher-Kiefer, 1995).

In sum, although previous evidence has clearly shown that gender stereotypes are used during backward inferencing, it remains unclear how strong these cues are compared to discourse/structural cues (also which one is preferred earlier during online discourse processing) Moreover, previous studies have failed to provide conclusive evidence for elaborative activation of gender stereotypes during natural discourse processing. These issues were investigated in Studies I and IV.

1.3.2 Implicit causality

The term *implicit causality* refers to events described with interpersonal verbs like *frighten* and *fear*. These verbs carry interesting biasing information: when people are asked to continue sentences consisting of *NP1 VERBED NP2, BECAUSE…* people tend to continue the story by referring to NP1 after the verbs like *frighten* and by referring to NP2 after the verbs like *fear*, see examples (13) and (14), respectively (Garvey & Caramazza, 1974, Garvey, Caramazza, & Yates, 1976; Guerry et al., 2006). The continuations people typically produce reflect information about a cause associated with either NP1 (following the verb *frighten*) or NP2 (following the verb *fear*). As this cause is encoded implicitly in the event described by the verb in the first clause, the phenomenon is called implicit causality. The two groups of verbs can be named as NP1-biased and NP2-biased based on the direction of the verbs' biasing probabilities.

(13) John **frightened** Mary, because he… [NP1-biased verb]

(14) John **feared** Mary, because she… [NP2-biased verb]

Pickering and Majid (2007) suggest that implicit causality should be used to refer to reasons for events. For example, the continuation *was scared* following the above examples (13 and 14) would refer to the reason for the emotional reaction denoted by the verb. Following Garvey and Caramazza (1974) they mention that implicit causality "*provides an abstraction of the type of reason that is most likely to be provided for the event, and indicates which entity the reason tends to be about*" (Pickering & Majid, 2007: 785). Moreover, the bias is an inference drawn from the description of an event and can be determined probabilistically. This indicates that the implicit causality verbs do not produce equally strong biases (Garvey & Caramazza, 1974; see also Rudolph & Fösterling, 1997, for a meta-analysis of the biasing probabilities). The biasing strength can be best determined by using "neutral" NPs (e.g., John and Mary) combined with implicit causality verbs (*John frightened/feared Mary*) rather than using NPs that carry biasing information inferred from general knowledge. For example, it is generally assumed that a boy might be more afraid of his father than another way around. Thus, this general world knowledge could interfere with the biasing strength caused by the verbs themselves (*The father frightened/feared the boy*).

Crinean and Garnham (2006) defined implicit causality slightly differently from what is described above. They argued that this information is based on the semantic or thematic representation of the verb arguments (for similar accounts, see Au, 1986; Brown & Fish, 1983; Stevenson, Crawley, & Kleinman, 1994). Therefore, in their definition, implicit causality is not considered probabilistic, unlike in the above definition (Pickering & Majid, 2007). As pointed out by Pickering and Majid (2007), this leads to a situation in which implicit causality is an all-or-none property of the verbs. However, such an interpretation is not supported by earlier studies that clearly showed that implicit causality is probabilistic in nature and that the strength of the bias varies across verbs (Garvey & Caramazza, 1974; Rudolph & Fösterling, 1997).

In studies of implicit causality, important questions have been *when* during the course of processing implicit causality information becomes activated and how it affects inference generation. The Clausal Integration Account assumes that the use of implicit causality information in pronoun resolution occurs during integration and is therefore a late process (Garnham, 2001; Garnham et al., 1996; Stewart et al., 2000), whereas the Immediate Focusing Account assumes that comprehenders rapidly make use of implicit causality information (Greene & McKoon, 1995; Koornneef & Van Berkum, 2006; MacDonald & MacWhinney, 1995; McKoon, Greene, & Ratcliff, 1993; Van Berkum, Koornneef, Otten, & Nieuwland, 2007). Previous studies have failed to unequivocally decide between these hypotheses; again, the methods used have turned out to be unsatisfying to differentiate between the accounts.

Late use of implicit causality. Experimental studies that have provided support for either of the above accounts, clausal integration or immediate focusing, have approached the question of when implicit causality is used by investigating pronoun resolution. These studies have demonstrated a strong effect of verb-based implicit

causality information: people tend to interpret the pronoun referring to the character that is the cause of the event, i.e., NP1 is the preferred antecedent for the pronoun after NP1-biased verbs such as *frighten* and NP2 after NP2-biased verbs such as *fear* (Garnham et al., 1996; Greene & McKoon, 1995; Koornneef & Van Berkum, 2006; McDonald and MacWhinney, 1995; McKoon et al., 1993; Long & DeLey, 2000; Stewart et al., 2000; Van Berkum et al., 2007). However, as was already noted above, the studies fail to agree about timing of the use of implicit causality during pronoun resolution.

The Clausal Integration Account states that implicit causality information has an effect on comprehension only during a late semantic clausal integration phase, when implicit causality information in the main clause is integrated with the explicitly stated causal information in the subordinate clause (Garnham, 2001; Garnham et al., 1996; Stewart et al., 2000). For example, using a probe recognition task in English, Garnham and colleagues (1996) showed that the responses to the probe words that were presented at the end of the sentence were prolonged when the ending of the sentence was inconsistent with the implicit causality bias such as in (15). This effect was found irrespective of whether the probe word referred to the antecedent of the pronoun or not (i.e., for both probe words, Walter and Ronald, in Example 15). Interestingly, the study did not show the same effect for the probe words presented immediately following the pronoun. These findings were interpreted as evidence to support the claim that implicit causality takes considerable time before it exerts an effect.

(15) Walter apologized to Ronald this morning because he demanded an apology.
[N1 biased verb / Incongruent ending]

Similarly, Stewart et al. (2000) found no early effects of implicit causality in self-paced reading experiments, but instead, found a late facilitation for sentences with consistent implicit and explicit causes. In their experiment, participants read sentences similar to those in Garnham et al. (1996). In Experiment 2, the participants were presented the sentences in two fragments with the split between the fragments following the pronoun (or a full NP) in the second clause (*NP1 verbed NP2 because he*). The finding that implicit causality inconsistency prolonged reading times only in the second fragment was taken as evidence for late use of implicit causality information. However, because a coarse-grained reading time measure was used, it is possible that some of the differences found in the later fragment are spill-over effects related to the pronoun presented at the end of the first fragment. Therefore, although these studies found evidence for late use of implicit causality and supported Clausal Integration Account, it may well be that the experimental techniques were not ideally suited to find early implicit causality effects (see also the discussion about challenges of interpreting results from probe recognition tasks in Chapter 2.1).

Early activation/use of implicit causality. In contrast to the integration account presented above, the Immediate Focusing Account claims that implicit causality information is used much earlier to focus on one of the referents at the expense of another, and thus it has an effect on reference resolution immediately when a pronoun

is encountered (Greene & McKoon, 1995; Koornneef & Van Berkum, 2006; MacDonald & MacWhinney, 1995; McKoon et al., 1993; Van Berkum et al., 2007). For example, McDonald and MacWhinney (1995; see also McKoon et al., 1993; Long & DeLey, 2000) observed an early effect of implicit causality on probes presented immediately after the pronoun, thus providing evidence that contradicts the results of Garnham et al. (1996) and Stewart et al. (2000). However, the sentence materials used by McDonald and MacWhinney (1995) differed from those in Garnham et al. (1996) and Stewart et al. (2000), in that only congruent endings were presented. Therefore, their stories might have led participants to make strategic predictions of the upcoming discourse. These strategic predictions might have contributed to the observed prolonged responses to the incongruent probe words (to the name of NP1 following the NP2-biased verbs and to the name of NP2 following the NP1-biased verbs).

In accordance with the focusing account, Koornneef and Van Berkum (2006) also found an effect of implicit causality shortly after people read the gender-marked pronoun in experiments on Dutch. In stories such as (16), in which the gender-marked pronoun *hij* 'he' was inconsistent with the implicit causality bias of the verb (in the example the verb *bood zijn excuses aan* 'apologized' is a subject-biasing verb), prolonged reading times were observed compared to stories where the gender of the pronoun was consistent with the implicit causality bias. Koornneef and Van Berkum (2006) used two reading time methods: self-paced reading and eye-tracking. While the results of the self-paced reading experiment showed an implicit causality effect only after the pronoun, the "early" effect of implicit causality was found only in the regression path measure in eye-tracking. However, it could be argued that this measure does not reflect the earliest possible stage of comprehension, as it also includes fixations after making a regression from the region of interest (but before the eyes fixate words following the region of interest). The early eye-tracking measures (such as first fixation duration) did not show an effect until the third word after the pronoun, suggesting that this effect may not be operating quickly at the pronoun.

(16) David en Linda reden allebei behoorlijk hard. Bij een druk kruispunt botsten zij met hun auto's stevig op elkaar. **Linda bood haar excuses aan David omdat hij** volgens de getuigen van het ongeluk alle schuld had.

'David and Linda were both driving pretty fast. At a busy intersection they crashed hard into each other. **Linda apologized to David because he** according to the witnesses was the one to blame.'

Van Berkum and colleagues (2007) conducted an ERP study using the same stimulus materials. They found a P600 effect appearing 400-700ms after the onset of a gender-marked pronoun that mismatched with the verb bias. Interestingly, the mismatch caused a P600 and not a N400 – similar to the earlier cited studies investigating gender stereotype mismatches, which also showed a P600 effect instead

of a N400 effect (see Hagoort & Brown, 1999; Osterhout et al, 1997; Van Berkum et al., 1999a,b).

As the literature of both behavioural and neural studies suggests, implicit causality is inferred and used when integrating an anaphor into the preceding linguistic context. However, my previous discussion makes clear that there is no consensus on when exactly implicit causality is inferred. Importantly, studies supporting the Immediate Focusing Account suggest that implicit causality can be inferred quickly and is used to highlight one of the characters over another; i.e., the NP1-biased verbs like *frighten* highlight NP1 and NP2-biased verbs like *fear* highlight NP2 in a discourse. However, these studies have observed the implicit causality effect only following the pronoun and thus have not been able to determine whether it is the anaphoric marker that prompts activation of implicit causality information or whether implicit causality information is activated even without such a prompt. By adopting a strict interpretation of the Immediate Focusing Account, implicit causality may be activated even prior to the pronoun. However, such activation could not be found in these studies due to the methodological challenges: Even though implicit causality would have become activated prior to the pronoun, it is not clear how this could have been measured from the reading times and ERP components.

Relative importance of implicit causality. Just as studies of gender stereotypes, the above studies of implicit causality have shown that semantic information has an important role when resolving an antecedent of an anaphor. The question of whether implicit causality has a priority role during such backward integration has remained unresolved. The early priority role of implicit causality in pronoun resolution is predicted by the Immediate Focusing Account because it assumes that one of the referents is already highly prominent in the mental model and thus has direct access to be interpreted as an antecedent when encountering a pronoun in discourse (Koornneef & Van Berkum, 2006; Van Berkum et al., 2007). However, as shown in Example (16), the material in these studies included a gender-marked pronoun that referred unambiguously to either a male or a female character. This character was either consistent or inconsistent with the implicit causality bias. It could be argued that in such cases, implicit causality is not strongly tested against other possible cues. Taking the same analogy as with the gender stereotypes, it could be argued that a better study would be one in which implicit causality is pitted against discourse/structural cues and by having an ambiguous pronoun that can equally refer to both characters, such as in (15). In addition, the context following the pronoun should also be kept ambiguous in order to give the processor ample time to come up with a decision without forcing it to do so rather quickly.

In sum, studies of implicit causality have proposed a large set of hypotheses that have remained unresolved. The questions about whether implicit causality is activated prior to the pronoun or only later during integration phase; whether implicit causality has a priority role during anaphor resolution or is sensitive to other potential discourse/structural cues; and what is the timing of the effects during pronoun resolution, are important issues subject to further testing with a more appropriate method than used in the earlier studies. These issues are raised in Study II.

1.3.3 Proto-role properties

A somewhat different aspect of verb semantics is described by *proto-roles* that are typically defined in relation to the thematic roles[1], or alternatively thematic relations of the main verb arguments such as agent and patient (Dowty, 1991; Schlesinger, 1995). Dowty (1991) proposed that these thematic roles could be classified based on the proto-roles the arguments exhibit. According to this view, agent and patient properties of subjects and objects are computed based on a number of independently occurring subcomponents of basic semantic properties such as sentience, volition and movement (see Table 1 for a more complete list). The number of these subcomponents for each argument is not fixed, and therefore, it is likely that some thematic roles exhibit more of these properties than others. The list of proto-agent and proto-patient properties as defined by Dowty (1991; see also Hopper & Thompson, 1980; Kako, 2006) are shown in Table 1.

Prototypically agent-like properties involve elements such as agent volitionally causing an event, moving in the event or being aware of the participation in a certain state or event (sentience). In addition, a prototypical agent also exists in the world irrespective of the described event. In contrast, prototypical patients exhibit properties such as being an incremental theme, being causally affected by the event or undergoing another kind of change of state, and do not actively participate in the agent's movement. A prototypical patient is also often described in relation to the occurred event and does not exist independently of the event (or perhaps not at all).[2]

[1]Thematic roles are often assumed to serve an argument-indexing function. The theta-criterion in Government and Binding Theory (Chomsky, 1981) stipulates that each NP of a predicate carries only one thematic role in a given context, i.e., each thematic role defines what type of role each argument plays in an event described in the linguistic context. Typical thematic roles are agent, patient and theme among others (e.g., Fillmore, 1968, 1977; Jackendoff, 1972, 1976, 1983, 1990). Because of their conceptual or semantic nature, it is argued that thematic roles should be situated at the semantic or conceptual level of language description, not at the syntactic level, even though they are not unequivocally separable from syntactic roles of the arguments (Jackendoff, 1972, 1983, 1990; Kako, 2006).

[2]In order to further validate the proto-role hypothesis, Kako (2006) used a questionnaire to test proto-role properties of transitive verbs, more precisely, the number and nature of proto-agent properties subjects exhibit and the number and nature of proto-patient properties objects exhibit. The results confirmed the overall hypothesis that subjects would be rated to have more prototypically agent-like than patient-like properties, and similarly, objects to have more patient-like than agent-like properties.

TABLE 1.

Some of the proto-role properties of agents and patients proposed by Dowty (1991).

Proto-Agent Properties	**Proto-Patient Properties**
Volitional involvement in the event or state	Undergoes change of state
Sentience (and/or perception)	Incremental theme
Causing an event or change of state in another participant	Causally affected by another participant
Movement (relative to another participant)	Stationary relative to movement of another participant
Exists independently of the event named by the verb	Does not exist independently of the event, or not at all

Dowty's proto-role account has its roots in the Transitivity Hypothesis proposed by Hopper and Thompson (1980). According to this hypothesis, the above listed features (among some others) contribute to the salience or prominence of the entities in the sentence, i.e., how visible these entities are for the comprehender. The verbs can be categorised in the following way: the more proto-role properties the verb arguments carry, the *higher in transitivity* the verb is; and the fewer proto-role properties the verb arguments have, the *lower in transitivity* the verb is. Thus, the more prototypical properties the verb arguments exhibit, the more foregrounded they are in the discourse, while fewer prototypical properties they exhibit, the more backgrounded they are in the discourse. The proto-role properties are summarised in Table 2.

The arguments of verbs *hit* and *see* exhibit a different amount of proto-agent and proto-patient properties. The arguments of highly transitive verbs like *hit* exhibit several proto-agent and proto-patient properties: Their subjects are volitionally involved in and aware of the action, voluntarily causing the event or causing a state change of object as well as performing a movement. Their objects are similarly prototypical: The object undergoes a change of state and is causally affected by the event and does not necessarily move during the action. However, when considering low transitive verbs, their subjects exhibit only a limited number of proto-agent properties. For example, in the case of *see*, the agent is volitionally involved in and aware of the action, but does not have any other proto-agent properties. Furthermore, the object of *see* does not exhibit any proto-patient properties.

TABLE 2.

Proto-role properties for high-transitive verb *hit* and low-transitive verb *see*.

		Hit	**See**
Proto-Agent properties of the subject	Volitional involvement	✓	✓
	Sentience/Perception	✓	✓
	Causing event/change of state	✓	
	Movement	✓	
Proto-Patient properties of the object	Undergo change of state	✓	
	Incremental theme		
	Causally affected	✓	
	Stationary	✓	

Proto-roles during online comprehension. Unlike the literature on gender stereotypes and implicit causality, there is a very limited literature on whether high and low transitivity affects online language comprehension. One study that illustrates the importance of argument role properties was conducted by Rose (2005), who studied argument role properties of verbs and showed that they affect adult online processing. Instead of manipulating proto-roles directly, he took advantage of the assumption that theme is higher in the thematic hierarchy than semantic goals, as illustrated in (17) (Larson, 1988; Speas, 1990). The higher position in the hierarchy also entails a more foregrounded and salient/prominent role for themes than goals.

(17) agent > theme > goal

In a series of self-paced reading experiments, Rose (2005) showed that after both (18) and (19), participants were faster to read the pronoun (and the definite noun phrase) when it referred to the semantic theme (20a) than to the semantic goal (20b). Interestingly, the effect was found irrespective of the syntactic position, i.e., whether the theme preceded the goal or vice versa. However, in an off-line judgement task, in which participants were asked to rate the continuations (20a/b), he found the former more acceptable after (18) than (19).

(18) John sprayed some paint on a wall.

(19) John sprayed a wall with some paint.

(20a) It dribbled down and made a mess.

(20b) It was big and needed two coats.

Although this study manipulated the relative order of the antecedents (theme and goal) of the following pronoun instead of the proto-role properties of verbs (by using different types of verbs), it showed that semantic information might have a significant role in pronoun resolution. However, it should be noted that using the pronoun *it* to refer to the word *paint*, is more problematic than using *it* to refer to the word *wall*, because paint is not a countable noun whereas wall is, and thus it is more plausible to refer to wall with *it*.

However, this study further suggests that verb transitivity, including the proto-role properties of verbs, may be exploited in order to learn more about the importance of semantics in online inference generation. In addition, it is also of interest to study the effects of proto-role properties among children in order to resolve the developmental aspect about the acquisition of these properties (Ibbotson & Tomasello, 2009; Kako, 2006). In Chapter 1.5, I will return to the issue of why proto-roles may be preferred over gender stereotypes or implicit causality in the studies of the importance of semantics during online child language comprehension.

1.4 Discourse/structural factors in backward inferencing

In order to answer the questions introduced above about the relative importance of semantics compared to other cues, competing discourse/structural factors were introduced in the current studies. Several different discourse and structural factors have been found to be important during online inferencing: For example, factors such as order-of-mention, grammatical role and topicality have shown to contribute to the salience/prominence of the entities in discourse, and thus to inference generation (e.g., Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Bestgen & Vonk, 2000; Crawley, Stevenson, & Kleinman, 1990; Frederiksen, 1981; Gernsbacher, 1990; Gernsbacher & Hargreaves, 1988; Järvikivi, Van Gompel, Hyönä, & Bertram, 2005; Kaiser & Trueswell, 2008, in press; Vonk, Hustinx, & Simons, 1992). In what follows, I shortly highlight the main findings of studies investigating the effects of these factors on online inferencing.

1.4.1 Order-of-mention

In her Structure Building Theory, Gernsbacher (1990) argued that people construct a mental representation of sentences by mapping new information onto the first mentioned entity. Therefore, *order-of-mention* plays a fundamental role in organising and highlighting information during language comprehension. Because of its privileged status, the first-mentioned entity is also assumed to be the preferred antecedent for a subsequently presented pronoun (see also MacWhinney, 1977). This assumption has been confirmed experimentally, as reviewed next.

For example, using a probe recognition task in English, Gernsbacher and Hargreaves (1989) showed that when a probe word (e.g., *Lisa* or *Tina*) was presented after sentences such as (21) and (22), the responses to the probe words were faster when they referred to NP1 (Tina) than NP2 (Lisa) – irrespective of the grammatical role of the first-mentioned entity (see also Carreiras, Gernsbacher, & Villa, 1995; Gernsbacher, Hargreaves, & Beeman, 1989; MacWhinney, 1977; McDonald & MacWhinney, 1995). Thus, the study supports the first-mentioned character's privileged status in discourse.

(21) Tina beat Lisa in the state tennis match.

(22) Tina was beaten by Lisa in the state tennis match.

The first-mention advantage has later been confirmed in pronoun resolution (Arnold et al., 2000; Järvikivi et al., 2005; Kaiser & Trueswell, 2008, in press). For example, by using the visual world eye-tracking method in English, Arnold et al. (2000) found that after hearing the ambiguous pronoun in (23), the probability of fixations to the first-mentioned character was higher than to the second-mentioned character.

(23) Donald is bringing some mail to Mickey, while a violent storm is beginning. He's carrying an umbrella.

(24) Donald is bringing some mail to Minnie, while a violent storm is beginning. He's/She's carrying an umbrella.

This close temporal link between eye movements toward this referent following the pronoun revealed that people rely on order-of-mention as a cue when resolving ambiguous pronouns that could also potentially have referred to the second-mentioned character. However, the pronoun's grammatical gender was unambiguous, as in (24), the probability of looking at Donald increased quickly following the pronoun *he* (the preference was supported by both order-of-mention and gender-marked pronoun), compared to the probability of looking at Minnie which increased following the pronoun *she* (conflicting cues of order-of-mention that referred to NP1 and gender-marked pronoun that referred to Minnie). These findings indicate that in pronoun resolution unambiguous grammatical gender marking of the pronoun was a stronger cue than order-of-mention.

However, Arnold, Brown-Schmidt, Trueswell and Fagnano (2005) failed to observe a first-mention preference with children aged 5. Using the same materials as with adults (Arnold et al., 2000), they showed that children only exhibited a weak first-mention preference, and instead, their comprehension of pronouns was strongly driven by unambiguous gender marking of the pronoun (for similar findings, see also Arnold, Brown-Schmidt, & Trueswell, 2007, Experiment 2). However, contrary to Arnold et al. (2005, 2007), Song and Fischer (2005, 2007) found in a preferential

looking task a clear first-mention preference with children as young as 2 to 3 years old. They presented children with three-sentence long stories in English. In those sentences the antecedent was mentioned two to three times prior to the pronoun and always in the same sentence position. Thus, it could be argued that this repetition may have ensured the effect of first-mention preference.

In most above studies order-of-mention was confounded with grammatical role. For example, the first-mentioned character was also the sentence subject in the examples (23-24). Therefore, it is not possible to determine whether the effect was purely driven by order-of-mention or whether grammatical role might also have contributed to the findings (see e.g., Järvikivi et al., 2005).

1.4.2 Grammatical role

As indicated in the previous section, it has been suggested that the grammatical role of antecedents, that is, whether it is a *grammatical subject* or a *grammatical object*, may also contribute to discourse salience/prominence (e.g., Crawley et al., 1990; Frederiksen, 1981; Gordon et al., 1993; Grosz & Sidner, 1986). As mentioned above, Gernsbacher and Hargreaves (1989) used passive structures in order to disentangle the effect of order-of-mention from that of grammatical role (Example 22). However, they failed to show any subject advantage, but instead showed a first-mention preference irrespective of the grammatical role (cf. Gordon et al., 1993).

Recently, Järvikivi et al. (2005) used another type of manipulation in order to distinguish between the first-mention and subjecthood preferences. Due to its relatively free word order, Finnish allows structures in which either the grammatical subject is mentioned first in the sentence (25, SVO word order) or the object is mentioned first in the sentence (26, OVS word order), thus making it possible to tease apart first-mention and subjecthood effects without passive structures: while NP1 is the grammatical subject in (25), NP1 is the grammatical object in (26).

(25) Tony Blair kätteli George Bushia valkoisessa talossa. [SVO]

Hän…

'Tony Blair (SUBJ) shook hands with George Bush (OBJ) in the White House. S/he…'

(26) George Bushia kätteli Tony Blair valkoisessa talossa. [OVS]

Hän…

'George Bush (OBJ) shook hands with Tony Blair (SUBJ) in the White House. S/he…'

Shortly after people heard the gender ambiguous pronoun *hän* ('s/he'), the probability of fixations to the subject was higher than the probability of fixations to the object. In addition, the results showed more looks to the first-mentioned than the

second-mentioned character. The lack of interaction between these factors suggests that both preferences are operating independently by adding activation to the referents in the discourse.

Earlier psycholinguistic studies have also shown that grammatical subjects are recalled more accurately in memory tasks (e.g., Clark & Card, 1969) and that they are highly likely to be mentioned again in the subsequent sentences (Givón, 1983). Taking together the earlier and more recent studies (Gordon et al., 1993; Järvikivi et al., 2005), it seems that grammatical role contributes to discourse salience/prominence. Therefore, both factors, order-of-mention and grammatical role, are important to take into account when studying the role of the discourse/structural factors in backward inferencing.

1.4.3 Topicality

In addition to order-of-mention and grammatical role, it has been suggested that the linguistic topic[3] contributes to structural salience of entities (e.g., Givón, 1983; Lambrecht, 1994). Lambrecht (1994: 118) defined the term *(sentence/clause) topic* pragmatically in the following way: "*The topic of a sentence is the thing which the proposition expressed by the sentence is ABOUT*". The topic is thus at the centre or focus of people's attention in discourse, and the sentence describes the proposition(s) about this topic in a given situation (see also e.g., Gundel, 1976). However, the above definition differs from the one proposed for example by Givón (1983), who allows the term topic to refer to any participant in the discourse, and does not make a principled distinction between topical and non-topical participants. Because of its central role in the situation described by the sentence, the topic is often referred to as *discourse topic* or *psychological or discourse focus*[4] (e.g., Garrod & Sanford, 1982; Grosz & Sidner, 1986; Gundel, 1999, 2008; Gundel, Hedberg, & Zacharski, 1993). These terms highlight the cognitive role of the participant constituting the topic: it is either currently under a focal attention and thus accessible as a referent for the following pronoun (Gundel, 1999, Gundel & Hedberg, 2008; Gundel, Hedberg, & Zacharski, 1993) or it holds a special role by having a strong representation in memory and is thus available as a referent for the pronoun (Foraker & McElree, 2007).

The grammatical subject in English is generally the topic, and thus the topic is often described in relation to the grammatical subject (Chafe, 1976; Lambrecht, 1994). However, being a grammatical subject or topic means two different things, and therefore, these two concepts are not necessarily related. For example, in Finnish,

[3] The term linguistic topic refers to the sentence and clause topic.

[4] Note that in linguistics the term *focus* is often used to refer to newly introduced information in the text and often occurs in sentence-final position rather than sentence-initial position in both English and Finnish (Chafe, 1976; Hakulinen, Karlsson, & Vikuna, 1980; see however Gussenhoven, 2007, for the role of prosody in focus assignment in English and Vainio & Järvikivi, 2006, for Finnish). Therefore, psychological or discourse focus do not refer to the same concept as linguistic focus.

which has a relatively free word order, the topic is more often related to the first-mentioned position than the grammatical subject (Vilkuna, 1989). For example, in both (27) and (28) the topic is *John*; however, while *John* is the grammatical subject in the former example, it is the grammatical object in the latter.

(27) John potkaisi Maryä laiturilla. [SVO]
'John (SUBJ) kicked Mary (OBJ) at the pier'

(28) Johnia potkaisi Mary laiturilla. [OVS]
'John (OBJ) kicked Mary (SUBJ) at the pier'

As can been seen from the above definitions of topic, it can be challenging to differentiate it from factors such as order-of-mention or grammatical role; instead, it often co-occurs with at least one of these factors. This makes it difficult to determine which of these factors is contributing to the salience of antecedents in pronoun resolution. Instead, it is likely that some of the effects that are assumed to be the result of order-of-mention or grammatical role are in fact influenced by topicality as well (e.g., Arnold, 2000, 2005, 2007; Gordon et al., 1993; Järvikivi et al., 2005; Kaiser & Trueswell 2008, in press; Song & Fisher, 2005, 2007).

In sum, factors such as order-of-mention, grammatical role and topic have been shown to contribute to discourse/structural salience or prominence and thus also to pronoun resolution by making the first-mentioned character, grammatical subject and the topicalised character the preferred antecedent for a subsequently presented pronoun. In addition, it seems rather difficult to clearly disentangle these properties from each other and thus to know which of them plays the key role in anaphor resolution. In order to take advantage both of these findings and of the fact that it is difficult to differentiate between the factors, in the current experiments these factors were mapped together to form a "structurally" highly salient referent that in turn can be pitted against the semantically most plausible referent in backward inferencing. Thus, it is possible to have two referents, one that is the structurally preferred and another that is the semantically preferred antecedent (or both structurally and semantically preferred/dispreferred) for the ambiguous pronoun. Because the above-mentioned discourse/structural factors are not differentiated from each other in the current studies, I refer to these factors with the combined term *discourse/structural* cues.

1.5 Research questions and hypotheses

The focus of the present set of studies was to investigate the importance of semantics during inferential processes: to investigate whether semantics is activated elaboratively during discourse comprehension and what its relative importance is in backward inferencing compared to discourse/structural cues. In the previous research, these questions have remained unresolved partly due to limitations of previously used

methods. Thus, introducing a novel method was crucial. I employed the visual world eye-tracking method that involves the recording of participants' eye movements to visual scenes while they are listening to spoken language (see a detailed description of the method in Chapter 2). This method allowed me to disentangle backward and forward inferencing and also to tap into the time course of semantic and discourse/structural effects. In the following, I introduce the research questions and the hypotheses derived from the earlier literature. Detailed descriptions of the studies are provided in Chapter 3.

Elaborative activation of semantics. Forward activation of semantic information was studied in two experiments investigating two different aspects of semantics, gender stereotypes and implicit causality. *Study I* tested the elaborative activation of generically presented gender stereotype information in Finnish. Mental Models approaches allow general world knowledge such as stereotypes to become activated incrementally during the course of processing and thus elaborative activation is predicted (e.g., Garnham, 2001; Johnson-Laird, 1983; Zwaan & Radvansky, 1998). However, such activation in not predicted by the Minimalist Account as it allows only information that contributes to local coherence or is readily available in short-term memory to become activated in a forward fashion (McKoon & Ratcliff, 1986, 1988, 1992). Measuring participants' eye movements toward male and female characters while they heard a gender stereotype role name such as *chimney sweep* (a male stereotype) or *make-up artist* (a female stereotype) was expected to reveal whether listeners activated gender stereotype information elaboratively: For example, following the word *chimney sweep*, it was expected that if stereotype information is activated, the probability of fixating a male character is higher than the probability of fixating a female character; and vice versa following a female stereotype name.

Another aspect of semantics that was used to study whether elaborative inferences are activated online was implicit causality. *Study II* was designed to constrast the predictions of the Clausal Integration Account and the Immediate Focusing Account. The former proposes that implicit causality is not activated elaboratively, but instead is activated later when implicit causality information in the main clause is integrated with the explicitly stated causal information in the subordinate clause (Garnham, 2001; Garnham et al., 1996; Stewart et al., 2000). The latter assumes that implicit causality can become activated even prior to the pronoun by highlighting one of the characters (Greene & McKoon, 1995; Koornneef & Van Berkum, 2006; MacDonald & MacWhinney, 1995; McKoon et al., 1993; Van Berkum et al., 2007). This question was studied by using Finnish as it allowed a good control over the implicit causality verbs: both NP1- and NP2-biased verbs were formed from the same stem (e.g., *pelottaa* 'frighten' and *pelätä* 'fear'). As in the gender stereotype study, the fixation probabilities to the two characters, referred to by NP1 and NP2, were calculated following the NP1-biased and NP2-biased verbs. If implicit causality information is activated prior to the pronoun, as suggested by the focusing account, fixation probabilities to the bias-consistent character should be higher than to the bias-inconsistent character. Such a preference should not be observed if implicit causality is activated only following an explicit prompt realised by a later occurring pronoun (or even later), as suggested by the integration account. In order to further test the question of whether an explicit trigger such as the presentation of causal conjunction

because is needed to activate implicit causality, fixation probabilities were also measured prior to and following the causal conjunction onset that followed later in the sentence context. Some earlier studies have proposed that the activation of implicit causality information does not occur without a causal connective (e.g., Koornneef, 2008; Stevenson et al., 1994; Stevenson, Knott, Oberlander, & McDonald, 2000; see however McDonald & MacWhinney, 1995). However, if such a prompt is not needed for activation of implicit causality information, fixation probabilities to the bias-consistent character should be higher than to the bias-inconsistent character even prior to the causal conjunction.

Importance of semantics in backward inferencing. Backward use of semantics was tested in two different conditions: it was pitted against discourse/structural prominence in *Studies II, III* and *IV*, while in *Study IV* the relative information was presented after an anaphoric expression in order to see whether an additional inference would be made to revise the earlier resolved anaphoric expression.

Study II tested the importance of implicit causality information compared to structural cues when processing the gender ambiguous third person pronoun *hän* 's/he' in Finnish. It was hypothesised that even though implicit causality may already become activated prior to the pronoun (see above) it does not automatically lead it to have a privileged role in pronoun resolution. Instead, it may be possible that implicit causality is competing with discourse/structural cues and that its importance is weighed against the strength of the other cues. Note however that this assumption is not in line with the strict interpretation of the Immediate Focusing Account as it assumes a priority role for the referent that is highlighted by the implicit causality information (Greene & McKoon, 1995; Koornneef & Van Berkum, 2006; Van Berkum et al., 2007). In other words, the Immediate Focusing Account assumes that implicit causality should be the preferred cue in pronoun resolution regardless of the strength of other competing cues. A structurally salient character was introduced prior to the pronoun: the character was the combined first-mentioned character, grammatical subject and a shifted topic. Having the structurally salient character not only as a topic but also as a shifted topic, we increased its salience in the discourse (e.g., Bestgen & Vonk, 2000; Vonk et al., 1992). This cue was contrasted with implicit causality information in order to study which one of these cues was more pronounced in pronoun resolution. In order to do so, the fixation probabilities toward the structurally preferred antecedent were compared to the probabilities toward the semantically preferred antecedent during the course of processing following the pronoun: If listeners prefer structural cues, the probability of fixating the structurally preferred antecedent should be high, irrespective of whether the antecedent is semantically preferred or not. Similarly, if listeners prefer semantic cues, the probability of fixating the semantically preferred antecedent should be high, irrespective of whether the antecedent is structurally preferred or not. In addition, if implicit causality has a privileged role as suggested by the Immediate Focusing Account, the effect should be found very shortly following the pronoun and even prior to the discourse/structural effects.

In *Study III* another approach was adopted. As discussed in the Introduction, structural factors have been found to play an important role also in children's pronoun resolution. However, the question of whether children are also sensitive to semantic

cues during pronoun resolution has remained unstudied. In order to shed light on this question and also to contribute to the discussion of when semantic and discourse/structural cues are acquired during language development, semantic cues were contrasted with structural cues in a study with three-year-old children. Although it would have been interesting to study effects of gender stereotype information or implicit causality similarly to the adult studies, investigating these properties would have been problematic in child language studies. It would have been hard to find nouns for which children have gender stereotypes. Similarly, implicit causality verbs have been found to be rather infrequent in child language or in child-directed speech; thus, they do not form an ideal basis for testing semantic effects on children's pronoun resolution (Serratrice & Kidd, in preparation). A corpus investigation of child-directed speech in English revealed that transitive verbs are frequently used and would thus be better suited for testing semantic effects in children's pronoun resolution than gender stereotypes or implicit causality.

Study III investigated the effects of verb transitivity and proto-roles on pronoun resolution. As in the studies mentioned above, two potential antecedents for an ambiguous pronoun were introduced and their semantic and discourse/structural preferences were manipulated orthogonally. Based on the earlier literature, it was expected that when children hear the pronoun *he* after a high transitivity verb (e.g., *hit*), both the subject and object would be considered highly prominent candidates for an antecedent because they both exhibit several proto-role properties (Dowty, 1991; Hopper & Thompson, 1980; Kako, 2006). In contrast, following a low transitivity verb, for which neither the subject nor the object exhibit many proto-role properties, both the subject and object are less prominent candidates for an antecedent. However, with these verbs it was hypothesised that the subject was more accessible as an antecedent for the pronoun than the object. This is because the objects of low-transitive verbs (e.g., *see*) do not exhibit any of the prototypically object-like properties. The probability of fixations toward the target characters following the pronoun were expected to reveal whether 3-year-old children are sensitive to semantic cues. If they are not, only a first-mention subject preference was expected (based on earlier findings of Song & Fischer, 2005, 2007).

In *Study IV*, the focus was again on gender stereotype information. In this experiment it was of interest to test whether two different types of backward inferences, anaphoric and non-anaphoric inferences are affected by semantic information to the same extent. It has been suggested that because anaphors explicitly trigger a matching process with their possible antecedents, the semantically most appropriate antecedents may be preferred (Greene et al., 1992). Thus, it was hypothesised that semantic information may play a strong role in anaphor resolution. However, whether semantic information also plays such a strong role during non-anaphoric inferencing is not clear. Earlier studies suggest that salient discourse entities are easier to retrieve than less salient entities (e.g., McKoon et al., 1996; O'Brien, Albrecht, Hakala, & Rizzella, 1995; O'Brien, Plewes, & Albrecht, 1990), and thus it may be that when an active matching process is not prompted by an anaphor (non-anaphoric inferencing), discourse/structural salience might be an equally strong or even stronger cue than semantic information. In *Study IV* the use of gender stereotype information in comparison to the use of discourse/structural cues

were investigated when participants processed ambiguous gender stereotypical NPs that were presented either with or without an anaphoric marker (*moottoripyöränsä* 'his/her motorcycle' vs. *moottoripyörän* 'motorcycle'). Measuring participants' eye movements toward stereotypically consistent vs. inconsistent and structurally preferred vs. dispreferred characters during both anaphoric and non-anaphoric inferences, it was possible to test the strength of semantic cues: Based on the above assumptions, it was expected that the use of semantic and discourse/structural cues could differ in the anaphoric and non-anaphoric conditions.

In addition, the question of whether people draw additional inferences (i.e. after making an anaphoric inference) based on semantic information was motivated by a prediction derived from the Minimalist Account. The account suggests that people draw only a minimal number of inferences in discourse, and thus, if one inference is already successfully drawn, there should be no need for additional inferences (McKoon & Ratcliff, 1986, 1988, 1992). Therefore, it was hypothesised in *Study I* that an ambiguous anaphoric expression would initially be resolved by using the highly salient discourse/structural cues (a combination of order-of-mention, grammatical role and topicality). According to the minimalist account, this information should be enough to build a coherent interpretation and the subsequently presented gender stereotype information should not be needed to make additional inferences. In order to test this hypothesis, a gender stereotype noun was presented following an anaphoric verb phrase. Importantly, the anaphoric verb phrase may be resolved by using the discourse/structural cues and thus there would not be a need for additional inferences. However, if gender stereotype words invite additional inferences, it was expected that participants would start fixating the stereotypically consistent character irrespective of whether that was structurally preferred or not.

In sum, four studies tested the importance of semantics during forward and backward inferencing in discourse. Three different aspects of semantics, gender stereotypes, implicit causality and proto-roles, were used in order to investigate effects of different aspects of semantics, thus making it possible to draw more general conclusions about the use of semantics in discourse processing than would have been possible by studying only one aspect. In all studies the importance of semantics was also tested against discourse/structural cues in order to determine the relative strength of the semantic cues. Finally, Study III further expanded the scope of the studies to child language comprehension by investigating an aspect of semantics that 3-year-old children are known to be familiar with.

2 METHODOLOGY

In this chapter I review previous methods that have been used to study inference generation during online processing and point out challenges involved in using these methods when trying to disentangle forward and backward inferencing. I begin by introducing the visual world eye-movement method and explain how linking visual attention and spoken language can be helpful in understanding inferential processes. Moreover, I explain how the visual world method was employed in studies of inferences in this thesis and what kinds of spoken and visual stimuli were used and how the experiments were conducted. Finally, I explain the methods used in the pre-tests for selecting stereotypically male- and female-biased role names (Study I) and nouns (Study IV), subject- and object-biasing implicit causality verbs (Study II), and high and low transitive verbs (Study III).

2.1 Earlier methods to study online inferencing

As noted in the Introduction, many different behavioural and neurocognitive methods have been used in order to understand inference generation during discourse processing. These methods have provided a great amount of important information on the kinds of inferences people draw during online language processing and how different structural and semantic factors affect these processes. However, many of the earlier used methods have turned out to be problematic for distinguishing between forward and backward inferencing during natural discourse processing.

Measuring reading times. Two different reading time measures, self-paced reading and eye movement recording during reading, have been employed in studies of online inferencing. In self-paced reading, participants are asked to press a button in order to proceed to the next sentence (sentence-by-sentence self-paced reading) or to the next word (word-by-word self-paced reading), while in the eye movement recording participants are allowed to read the text on the computer screen without the additional requirements of pressing a button. Both of these methods rely on the assumption that prolonged reading times reflect comprehension difficulties: For example, when people are asked to read two-sentence stories such as (29) or (30), reading times for the personal pronouns have been found to be prolonged in (29) when the gender of the pronoun mismatches with the female gender stereotype information of its antecedent *secretary,* compared to reading times in (30) in which the pronoun matches with the gender stereotype information of its antecedent (e.g., Kennison & Trofe, 2003).

(29) The secretary distributed an urgent memo. **He** made it clear that work would continue as normal.

(30) The secretary distributed an urgent memo. **She** made it clear that work would continue as normal.

Both of the reading methods are dependent on an experimental manipulation in which a part of the discourse (e.g., a pronoun) mismatches with the previous context (such as in 29). When such mismatch is not present, as is the case with the gender-ambiguous Finnish pronoun *hän* 's/he' in the context *John kätteli Marya. Hän...* ('John shook hands with Mary. S/he…'), the pronoun can refer to either of the characters, and it is difficult to make predictions about how reading times would be affected when people relate the pronoun to the first mentioned male character or the second mentioned female character; i.e., the reaction times do not reveal whether the pronoun is resolved as referring to John or Mary.

More importantly, measuring reading times at or after the pronoun in (29) and (30) fails to be informative with regard to whether the gender stereotype information associated with *secretary* is activated when first encountering the word in the discourse or only later when the pronoun needs to be integrated into the preceding context. Thus, reading times for the inconsistent pronoun cannot differentiate between elaborative stereotype activation and later integration. Instead, the prolonged reading times at the pronoun may be due to the earlier activated gender stereotype information mismatching with the pronoun (secretary – he) or due to an integration effect prompted by the pronoun: the pronoun *he* makes people to look for the male antecedent and the failure to find one causes the processing cost.

Of reading time methods, eye movement recording during reading has an advantage in comparison to the self-paced reading paradigm, as reading times can be divided into more fine-grained measures, such as first fixation duration, first-pass fixation time and regression path duration, which reflect different stages in the processing time course. In addition, as mentioned earlier, eye movement recordings allow participants to read text quite naturally without requiring button presses in order to advance through the text. Because of these advantages, eye movement measurements make it possible to study how early during processing different structural and semantic factors exert an effect on bridging inferences.

Neurocognitive methods. In neurocognitive studies of online inferencing similar challenges may be faced. A popular method for studying online inferencing is the recording of event related potentials (ERPs). For ERP measurements, small voltage fluctuations of electric pulses are measured from the scalp produced in response to events of interest. These voltage differences are averaged across many trials to produce an event-related potential. The advantage of this method is that electric pulses appear on the scalp very quickly after linguistic information is processed in the brain; thus, the method provides a very good timing accuracy. Even though the ERP method has turned out to be very helpful for studying online language processing, the basic experimental design needed to make interesting comparisons relies on a similar mismatch between a pronoun and its antecedent, as described above in relation to reading time studies. Thus, similar to reading times measures, ERP studies have not been able to differentiate between forward activation of semantic information and the use of semantic information in backward inferencing.

In ERP studies, two components, the N400, a negative peak in the waveform around 300-500 ms after target onset, and the P600, a positive peak around 500-700 ms post-target have been associated with semantic and syntactic integration

processes, respectively (see Kutas & Federmeier, 2007, for an overview of ERP components). However, it is relatively unclear what component would represent inferential integration (e.g., Van Berkum et al., 2007). One component that has been found when comparing the processing of an ambiguous and unambiguous anaphor is referred to as the "Nref", which is the negative peak around 300 ms after the target onset and is observed for ambiguous referents in the anterior sites of the brain (Van Berkum, Brown, & Hagoort, 1999a). However, there is yet no consensus on whether this component is reference-specific or even specific to language processing. More importantly, the previous evidence does not explain whether Nref is a neural correlate for just simply noticing that the pronoun is ambiguous, whether it reflects an active resolution of the antecedent, or keeping the possible referents active in memory until they become disambiguated (see Van Berkum et al., 2007, for further discussions). Therefore, even though the ERP paradigm provides a helpful tool for answering questions concerning online inferential processing, more studies are needed to specify the components that reflect different aspects of predictive and integrative processes during inferencing.

Probe recognition task. Many earlier studies of inferencing have employed probe recognition tasks in which participants are reading (or listening to) linguistic stimuli and are occasionally interrupted and asked to judge whether the probed word was presented in the previous linguistic context, or to decide whether the target word is a word or not, or whether it was mentioned in the preceding linguistic context. By measuring reaction times to probes, it is possible to make inferences about the degree of activation of the target word at a certain point during discourse processing. Even though this method has been found to be indicative of the kind of information people use to draw inferences during online language comprehension, it has also been argued that the probe recognition task interrupts the normal language comprehension process and places a relatively large memory load on the comprehender, as well as invites readers to concentrate on surface features of the text at the expense of meaning processing (for a discussion, see e.g., Arnold et al., 2000). Moreover, it has been suggested that the method may also invite processing strategies different from normal (uninterrupted) language processing (Gordon, Hendrick, & Foster, 2000).

Lexical decision method. In addition to the above-mentioned online methods, the lexical decision paradigm with priming has also been used to study activation of semantic information. In this task, a speeded response is made to target words presented following a prime word in isolation of any discourse context. For example, it has been used to study whether gender stereotypes are activated elaboratively (e.g., Banaji & Hardin, 1996; Oakhill et al., 2005). Even though these studies suggest that people do activate gender stereotype information elaboratively, the results are not directly generalisable to language comprehension in context.

As discussed above, several methods have been successfully employed to study online inferences. Like any method, the aforementioned reading time, ERP, probe recognition and lexical decision methods have limitations when applied to the study of inference generation during online discourse processing – especially elaborative activation of semantic features. Therefore, it is of importance to develop methods that can be used to study online inferencing, and especially, to disentangle elaborative activation from integrative processing. The visual world eye-tracking method

(Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995) is a relatively new method that has the potential to serve this purpose.

2.2 The current approach: The visual world method

The visual world eye movement method has its basis in the seminal study of Cooper (1974), who measured participants' eye movements to pictures while they were listening to spoken stories referring to those pictures. He found that the eyes tend to move in a (nearly) automatic fashion to the objects of the viewed scene when they were referred to in spoken language. The method has more recently been employed by Tanenhaus et al. (1995) by using a modern eye-tracking recording technique. In their study of syntactic ambiguity resolution, Tanenhaus et al. (1995) confirmed that there is a close link between spoken language comprehension and eye movement behaviour. Moreover, they demonstrated that by measuring people's eye movements to the visual environment it is possible to make inferences about online language comprehension. In their studies, participants were asked to move around objects based on spoken instructions such as (31), which is temporarily ambiguous at the time when the phrase *on the towel* is uttered. The phrase can be interpreted as the goal where the apple should be moved to or as a modifier indicating where the apple that should be moved is currently located.

(31) Put the apple on the towel… (in the box)

When the visually depicted referents included only one apple, participants initially adopted the goal interpretation when they heard *on the towel*, as indicated by an increased probability of fixating an empty towel. However, when the visually depicted referents included two apples, one of which was on a towel, participants initially adopted the modifier interpretation, as indicated by an increased probability of fixating the apple that was on the towel. The study showed that eye movements to visually depicted objects may be informative in resolving what kind of interpretation is initially adopted when people comprehend ambiguous structures in spoken language (Eberhard, Spivey, Sedivy, & Tanenhaus, 1995; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; see also Trueswell, Sekerina, Hill, & Logrip, 1999, for similar findings with children).

Later studies have further confirmed the close link between spoken language comprehension and eye-movement behaviour. For example, studies on word recognition have employed simple instructions such as *Click on the TARGET WORD. Now put it below…* These studies have shown that listeners rapidly direct their eyes towards the intended picture or their phonological or visual competitor (such as a beetle after hearing *beagle* or a rope after hearing *snake*, respectively) generally within 300 ms from the spoken word onset (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, & Tanenhaus, 2001; Dahan & Tanenhaus, 2005; Sedivy, Tanenhaus, Chambers, & Carlson, 1999). Considering that it takes approximately 200 ms to plan and execute a saccade (e.g., Matin, Shao, & Boff,

1993), this demonstrates that eye movements are tightly time-locked to the processing of linguistic material.

Since the Tanenhaus et al. (1995) study the visual world eye movement method has captured a lot of attention and has successfully been employed to investigate a variety of research topics related to language comprehension, production and dialogue (see e.g., Tanenhaus, 2007, for an overview). The following two subsections discuss how visual attention is linked to spoken language comprehension and inferencing as well as introduce different types of tasks used with this method. Finally, I explain how this method was employed in the current studies of inference processing during spoken language comprehension.

2.2.1 Linking visual attention and spoken language

It is widely assumed that attention has an important role in eye movement guidance during inspection of visual environments. Fixations are used as an overt measurement of information people are attending to at a given moment. Thus, eye movements to objects available in the visual environment are expected provide a window to attentional processes (e.g., Findlay, 2004; Irwin, 2004). In psycholinguistic studies the allocation of attention to certain objects at any given moment is measured in order to make interpretations of the cognitive processes underlying spoken language comprehension. *Linking hypotheses* of the coordination of visual attention and spoken language explain the dynamic interaction between these two modalities.

Recently, two types of hypotheses for the linking of visual attention and spoken language processing have been introduced and implemented in computational models (Altmann & Kamide, 2007, 2009; Altmann & Mirkovic, 2009; Crocker, Knoeferle, & Mayberry, in press, Knoeferle & Crocker, 2006, 2007; Mayberry, Crocker, & Knoeferle, 2009). Crocker et al. (in press) introduced a network called CIANet that models language comprehension in coordination with the visual environment. The model has as its basis the *Coordinated Interplay Account* (Knoeferle & Crocker, 2006, 2007), a model that includes three separate processes for establishing interpretation of spoken language in coordination in the visual context: (i) searching for a referent, (ii) establishing a referent, and (iii) use of visual scene to confirm or alter the linguistic interpretation. In the first processing phase, the spoken language input guides attention to the scene, as a result of which people search for objects and establish reference to these objects and events and anticipate potential or likely up-coming referents. In the last phase after people have attended to the (anticipated) referents, the scene information quickly influences the comprehension process by confirming or altering the currently built interpretation. The model also states that comprehension and attention allocation in the scene are closely time-locked so that the use of scene information for language comprehension may even take priority over other types of information, such as general world knowledge of typical events (Knoeferle & Crocker, 2006). Even though the processing phase (i) and (ii) can be conceptually distinguished, they do not necessarily occur serially, but instead in reality the processes may overlap and occur in parallel.

A slightly different approach is the computational model by Altmann and Mirkovic (2009). According to their view, the anticipated linguistic meaning and scene information are not distinguishable from each other; instead, they are seen to interact with each other through coindexation and subsequent updating of the linguistic meaning and scene information. In visual world studies, the visual world has been shown to people from the onset of the spoken stimuli or even prior to it. Thus, it is possible that identification and representation of the visual information precede spoken language. Later, linguistic input activates features of scene objects in the hearer's mind, including affordances (information of how the objects interact with other objects in the real world), resulting in early eye movements to visually affordable (and anticipated) objects (Altmann, & Kamide, 2007, 2009). It is also highlighted that the nature of the visual world is not visual per se, but rather a mental representation formed by coordinating visual perception with spoken language and which is updated on the basis of the unfolding spoken language (see also Allopenna et al., 1998; Tanenhaus et al., 2000).

Evaluation of the proposed linking hypotheses is especially important to explain the results of the studies in which visual information has been found to affect, constrain or alter interpretation processes such as in the seminal study by Tanenhaus et al (1995) and several later studies of language comprehension (e.g., Altmann & Kamide, 1999, 2007, 2009; Kamide et al., 2003; Knoeferle & Crocker 2006, 2007; Knoeferle et al., 2005). However, the visual world eye movement method can also be used as a method to investigate how people process linguistic information rather than study how visual information affects language comprehension. For example, in their pronoun resolution study, Järvikivi et al. (2005) presented two potential antecedents (*George Bush* and *Tony Blair*) of the ambiguous pronoun (*he*) in the visual stimuli, and importantly, the visually depicted pictures did not provide any cues on how to resolve the pronoun (see the examples 29-30 in Chapter 1.4.2, above). Instead, increased fixations to the characters were interpreted as reflecting how participants resolved the pronoun. Thus, the study took advantage of earlier findings showing that spoken language stimuli make people attend to pictures that are referred to in the linguistic context, but the visual stimuli did not create a bias that affected ongoing interpretation processes.

The current experiments in the thesis were not planned to test different types of linking hypotheses. Instead, the visual stimuli were chosen such that they did not give any cues on what types of inferences should be made and when; instead, the probability of fixating one referent in the visual scene more than another provides a measure of what type of inferences are drawn during online comprehension. For example, the following patterns were of interest: do people fixate a stereotypically consistent character after hearing a stereotype word such as *a secretary* and do people fixate structurally or semantically preferred antecedents following an anaphor.

2.2.2 Active and passive tasks

As indicated above, the visual world method can be used in different ways. In addition to the potential role of the visual scene in comprehension processes, different

tasks are introduced. The tasks can be divided into two groups based on whether the task requires an additional motoric movement or not. In active tasks, participants are asked to perform a motor action based on spoken instructions such as *put the apple in the box* or *click on the bacon*. These experiments involve two tasks: Comprehension of the spoken language as well as planning and performing a motor action based on those instructions. Even though eye movements in these experiments have been found to be indicative of underlying language comprehension (people tend to fixate the visual referents of the spoken input), it is not always easy to interpret what proportions of the fixations are controlled by these language comprehension mechanisms and what proportions by the action planning mechanisms (e.g., Boland, 2005). For example, Hayhoe & Ballard (2005) conducted an experiment using tasks that do not include any language comprehension, in which participants were asked to either search for an object or reach for an object. The probability of fixations was higher in the latter condition than in the former.

A different type of visual world eye movement study, which uses passive tasks, mimics what comprehenders spontaneously do. In these tasks, people are not asked to perform any motoric movements; instead, they are only instructed to listen to the stories for comprehension. Participants may be explicitly instructed to attend to the pictures while listening to the stories (e.g., Järvikivi et al., 2005; Knoeferle & Crocker, 2007), or they are not given any direct instructions to relate the depicted pictures to the spoken stimulus (e.g., Altmann & Kamide, 1999, Experiment 2; Knoeferle & Crocker, 2006). This line of visual world experiments has employed different secondary tasks in order to make sure people follow the stories for comprehension. Examples of such tasks are a picture verification task (e.g., Altmann & Kamide, 1999, Experiment 1; Arnold, et al., 2000), comprehension questions (Arnold et al., 2007) or a continuation task in which people are asked to continue the story by using the words and pictures occurring in the task (Järvikivi et al., 2005). Interestingly, participants' eye movement behaviour does not seem to be dependent on different instructions used in the passive versions of the visual world studies. For example, Altmann and Kamide (1999) showed that when people were given a picture verification task (Experiment 1) the pattern of eye movements were the same as when they were not specifically instructed to relate the pictures to the text (Experiment 2). In other words, people related the spoken language to the pictures even without specific instructions. This effect is also shown by Altmann (2004), who demonstrated similar eye movement behaviour even in the *Blank screen paradigm* in which the visual scene was presented only prior to the spoken stimulus and thus absent during the spoken language comprehension.

The current studies employed the passive version of the visual world eye movement method in order to avoid potential additional effects of motoric tasks on eye movements. The methodological details are explained in the following chapter.

2.3 Visual world method as employed in the thesis

The inference studies in this thesis adapted the passive version of the visual world eye movement method. In Studies I, II and IV participants were asked to carefully

listen to the stories and at the same time follow with their gaze the pictures depicted on the screen, in the same order they appeared in the story. In addition, they were occasionally asked to continue the story by using the persons and places mentioned therein. The continuation responses were not recorded or analysed because they are not of theoretical interest; the task only ensured that participants listened to the stories for comprehension. Study III differed slightly in its instructions: Children were instructed to see who would do something silly and no continuation task or any other additional tasks were performed. The experiments took between 20 minutes (Study III) and 40 minutes (Study IV) to complete.

2.3.1 Spoken and visual stimuli

The spoken stimuli in the studies consisted of stories containing between two and four sentences. The stories were spoken by native speakers of Finnish (Studies I, II and IV) or a native speaker of British English (Study III). Recordings were carried out in the acoustic laboratories at the Center for Cognitive Neuroscience, University of Turku and at the School of Psychology, University of Dundee. The phonetic software Praat (Boersma, 2001) was used to record, edit and analyse the sound files. The stories were spoken as complete stories with a natural speech rate of 4-7 syllables per second and with natural sentence intonations avoiding unnatural prosodic features. The onsets of the critical target words for data analyses were determined on a trial-by-trial basis determining the exact acoustic onsets for each target item in each study. Measuring was done using phonetic details of both the acoustic waveform and the visual spectrogram.

In order to increase the phonetic control of potential prosodic variability over the materials, the stories used for Studies II and III were cross-spliced. In Study II one of the experimental conditions' recording was selected as the base story (e.g., Condition 1, see Example 39, Chapter 3.2) and other conditions were created by replacing the pronoun with another type of pronoun (e.g., *hän* 'he' with *hänet* 'him' in order to create Condition 2). Further conditions were created by replacing the verb in the base story (the verb *pelkäsi* 'feared' with *pelotti* 'frightened' in order to create Condition 3) and by splicing a complete subclause into Condition 3 in order to create another condition (i.e., by splicing the subclause from Condition 2 into Condition 3). A similar procedure was used in Study III (see example stories in Figure 8, Chapter 3.3).

In the studies of this thesis, the visual stimuli were neutral with respect to the linguistic stimuli; they did not bias towards a particular interepretation or inference (cf. Altmann & Kamide, 1999, 2007, 2009; Kamide et al., 2003; Knoeferle & Crocker, 2006, 2007; Knoeferle et al., 2005; Tanenhaus et al., 1995). In other words, the visual stimuli were designed not to give any obvious cues for what kinds of inferences to make in discourse and when. The visual stimuli comprised three or four pictures of the same size: photographs collected from the Internet in the adult studies (Study I, II and IV, see Figure 1) and colour drawings in the child study (Study III). The pictures were assigned to positions on the screen and their location was counterbalanced between items in order to avoid strategic effects that could have arisen if participants learnt the location in which a certain character (e.g. first-

mentioned character or grammatical subject) always appeared; using this procedure it was also possible to avoid a situation in which some screen locations would potentially receive generally more fixations than other locations. The studies of adult language comprehension always displayed four pictures that included two (competing) target characters for investigating inferences and two other objects depicting locations mentioned in the linguistic context. The visual stimuli in the child study displayed three pictures: two competing antecedents for the pronoun and a location. An example of the visual stimuli of Study I is provided in Figure 1 (see example of the stimuli of the child study in Figure 6 (Chapter 3.3). In the experimental procedure, the pictures appeared on the screen 50 ms before the sound file onset and remained there until the sound file ended.

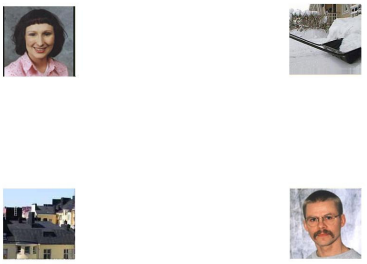

Figure 1. An example of the type of visual stimuli used in Study I.

2.3.2 Eye movement recording and coding for analyses

Eye movements in all adult studies were recorded using an EyeLink II head-mounted eyetracker manufactured by SR Research Ltd, Mississauga, Ontario, Canada. The eyetracker is a head-mounted near infrared video-based tracking system combined with hyperacuity image processing. The eyetracker samples pupil location at 500 Hz, thus providing a very fine-grained measure of eye locations on the screen. The spatial accuracy of the system is better than 0.5 degrees of arc. In all studies viewing was always binocular, but only one eye was recorded. The tracker was calibrated using a 9-point calibration grid; calibration was validated before each experiment and also during the experiment whenever necessary. Each trial was preceded by a fixation point that allowed the eye-tracking software to make a drift correction if necessary. Software for EyeLink developed in Saarbrücken was used to programme the experiments and prepare the data for analyses (see e.g., Knoeferle et al., 2005, for more information of the software). The visual scenes were accompanied with colour coded bitmap templates that corresponded to the picture areas on the screen. The software matched the colour coding with the x-y coordinates of the recoded fixations, so that it was possible to identify which of the pictures or a background participants were fixating at any given moment during the trials.

Children's eye movements in Study III were videoed using a Panasonic NV-GS55 mini DV digital video camera mounted on the centre top of a 50 x 30 cm coloured screen. In this experiment, the experimenter was present next to the child in order to help him/her to concentrate on the task. The eye movement data were coded by a research assistant blind to the hypotheses of the experiment using the INTERACT© video coding software.

The fixation probabilities to the target characters were used as the dependent measure. As the main inferential statistic analyses, repeated measures analyses of variance (ANOVA) or hierarchical log-linear modeling were used. Detailed descriptions of the inferential statistical analyses are reported in the original publications (see the Appendix).

2.4 Methods used in the pretests

A variety of pretests were carried out with the experimental materials. For the two studies investigating gender stereotype biases of occupations/role names such as *chimney sweep* and *make-up artist* (Study I) or nouns such as *motorcycle* and *apron* (Study IV) were pretested. In total, 124 occupations and 100 nouns were pretested using a 7-point Likert scale with 1 indicating an extremely masculine and 7 an extremely feminine occupation/noun. For both studies, the 18 most masculine and 18 most feminine occupations/nouns were selected for the experimental stimuli. The stimuli selected for Studies I and IV are listed in Tables 3 and 4.

TABLE 3
List of stereotypically male-biased and stereotypically female-biased role names selected for Study I.

Male stereotype occupations	**Female stereotype occupations**
asentaja 'mechanic'	kassanhoitaja 'cashier'
hautausurakoitsija 'funeral director'	keittäjä 'cooker'
kalastaja 'fisher(man)'	kirjastovirkailija 'librarian'
kivenhakkaaja 'stonecutter'	kodinhoitaja 'home aid'
korjaaja 'repairman'	kosmetologi 'beautician'
lentäjä 'pilot'	kotiavustaja 'home help'
maanviljelijä 'farmer'	lähihoitaja 'hospital attendant'
mekaanikko 'mechanic'	matkaopas 'travel guide'
metsästäjä 'hunter'	meikkaaja 'make-up artist'
muurari 'mason'	ompelija 'dressmaker'
nuohooja 'chimney sweep'	parturikampaaja 'hairdresser'
palotarjastaja 'fire inspector'	sairaala-apulainen 'hospital attendant'
seppä 'smith'	sairaanhoitaja 'nurse'
sotilas 'soldier'	sihteeri 'secretary'
teknikko 'technician'	siivooja 'cleaning lady'/'cleaner'
vahtimestari 'doorman'/'janitor'	sisustussuunnittelija 'interior designer'
yksityisetsivä 'detective'	tanssija 'dancer'
ylikonstaapeli 'police sergeant'	tarjoilija 'waiter'/'waitress'

TABLE 4
List of stereotypically male-biased and stereotypically female-biased role names selected for Study IV.

Male stereotype nouns	**Female stereotype nouns**
alkometri 'breathalyser'	balettitossu 'ballet shoe'
haalari 'overalls'	esiliina 'apron'
jääkiekkomaila 'hockey stick'	hiusponnari 'hair clip'
kitara 'guitar'	huulirasva 'lip balm'
moottoripyörä 'motorcycle'	jumppapallo 'gymnastic ball'
moottorisaha 'chain saw'	käsikoru 'bracelet'
nahkaliivi 'leather waistcoat'	käsirasva 'hand cream'
naulapyssy 'nail gun'	korurasia 'jewel case'
nyrkkeilyhanska 'boxing glove'	korvakoru 'earring'
peräkärry 'trailer'	kynsiviila 'nail file'
pilkkijakkara 'ice fishing seat'	mankeli 'mangle'
putkikassi 'sport bag'	meikkipussi 'toilet bag'
ruohonleikkuri 'lawn mower'	neula 'needle'
salibändymaila 'floorball stick'	nilkkaketju 'ankle bracelet'
sirkkeli 'circular saw'	päiväkirja 'diary'
taskukello 'pocket watch'	siivousämpäri 'cleaning bucket'
traktori 'tractor'	silkkihuivi 'silk scarf'
vasara 'hammer'	sormuskello 'finger watch'

For Study II subject- and object-biasing implicit causality verbs were selected by using a sentence completion task that was also used in the earlier studies (Garvey & Caramazza, 1974). In total, 90 verbs were embedded in a sentence frame *NP1 Verb NP2, because…* Based on whether participants referred to the subject or object in their continuations, the 16 most subject and 16 most object biasing verbs were selected and used to create the linguistic materials for the experiment. Table 5 lists the used verbs.

TABLE 5

Subject- and object-biased implicit causality verbs selected for Study II.

Subject-biasing verbs	**Object-biasing verbs**
epäilyttää 'make somebody suspicious'	epäillä 'suspect'
hämmästyttää 'amaze'	hämmästellä 'be amazed'/'wonder'
hätkähdyttää 'faze'	hätkähtää 'startle'
ihastuttaa 'delight'	ihailla 'admire'
ihmetyttää 'puzzle'	ihmetellä 'marvel'
inhottaa 'disgust'	inhota 'detest'
kauhistuttaa 'horrify'	kauhistella 'be horrified'
kavahduttaa 'startle'	kavahtaa 'be startled'
kummastuttaa 'puzzle'	kummastella 'marvel'
oudoksuttaa 'to be found strange'	oudoksua 'find strange'
pelottaa 'frighten'	pelätä 'fear'
säälittää 'make somebody pity'	sääliä 'pity'
surettaa 'distress'	surra 'mourn'
vihastuttaa 'anger'	vihata 'hate'
äimistyttää 'amaze'	äimistellä 'be amazed'/'wonder'
ällistyttää 'amaze'	ällistellä 'be amazed'/'wonder'

In order to assess the proto-role properties of the verbs for Study III, 30 reversible transitive verbs that occurred frequently in child directed speech in the Manchester Corpus on the CHILDES database (MacWhinney, 2000; Theakston, Lieven, Pine, & Rowland, 2001) were first selected. Transitivity of those verbs was tested by creating a questionnaire using some of the proto-role questions employed by Kako (2006, Experiment 1). Participants saw the verb embedded in nine questions presented in Table 6 and were asked to rate on a 7-point Likert scale the agent- and patient-like properties of the subject and object.

The 10 verbs that had most agent- and patient-like properties were selected as the high transitive verbs (*feed, pinch, phone, cuddle, squash, kiss, squeeze, kick, bang* and *hit*) and the 10 verbs having the least agent- and patient-like properties represented the low transitive verbs (*frighten, scare, pull, tickle, push, poke, caught, chase, leave* and *upset*).

TABLE 6

An example of the transitivity rating questions for the verb *chase*.

Given the sentence: ***Jane chased Kate***
(1) How likely is it that Jane chose to be involved in chasing?
(2) How likely is it that Jane was aware of being involved in chasing?
(3) How likely is it that Jane caused a change in Kate?
(4) How likely is it that Jane caused Kate to do something?
(5) How likely is it that Jane moved?
(6) How likely is it that Jane existed before chasing took place?
(7) How likely is it that Kate was changed in some way as a result of chasing?
(8) How likely is it that Kate was created as a result of chasing?
(9) How likely is it that Kate was stationary?

3 OVERVIEW OF THE ORIGINAL STUDIES

This thesis consists of four separate studies, of which three have been accepted for publication in peer-reviewed journals and one will be submitted for publication. Studies I, II and IV were carried out at the University of Turku, Finland and Study III at the Max Planck Child Study Centre at the University of Manchester, United Kingdom.

The studies examined the use of semantic information and structural factors in inference making. Studies I (Pyykkönen, Hyönä, & Van Gompel, 2009) and II (Pyykkönen & Järvikivi, 2009) tested both the elaborative activation of semantics as well as the use of semantic information in reference processing when making bridging inferences: Study I tested the elaborative activation of gender stereotypes during online processing and the use of gender stereotypes to revise coherence relationships, while Study II investigated the elaborative activation of implicit causality in pronoun resolution and how implicit causality competes with structural linguistic factors. Study III (Pyykkönen, Matthews, & Järvikivi, 2009) investigated the role of one important type of semantic information, proto-roles, in comparison to structural prominence, in pronoun resolution among children. Finally, Study IV (Pyykkönen, Van Gompel, & Hyönä, 2009) examined the importance of gender stereotype information and to discourse/structural salience in making two types of bridging inferences (anaphoric and non-anaphoric).

All studies exploited the visual world eye-movement paradigm as the main methodology. A total of 232 participants took part in the studies: 98 undergraduate students in the eye movement experiments in Turku and 19 children in Manchester; in addition, 115 undergraduates from Turku and Manchester participated in the pretests on gender stereotypicality (Studies I and IV), implicit causality biases of verbs (Study II) and verb transitivity (Study III).

3.1 Activation of gender stereotypes (Study I)

Previous psycholinguistic studies have shown that gender stereotypes affect language comprehension in a backward fashion when resolving pronouns referring back to stereotyped words (e.g., secretary… s/he) (e.g., Duffy & Keir, 2004; Irmen, 2007; Kennison & Trofe, 2003; Kreiner et al., 2008; Osterhout et al., 1997; Sturt, 2003a,b). It has also been suggested that gender stereotypes are activated elaboratively in forward fashion during text comprehension (Garnham et al., 2002; Oakhill et al., 2005; Sanford, 1995; Reynolds et al., 2006). However, these previous studies have not been able to clearly differentiate between elaborative activation and backward use (in the form of bridging inferences) of gender stereotype information.

Mental Models Theory and the Minimalist Account described in Chapters 1.2.1 and 1.2.2 make opposite predictions regarding the above questions. Elaborative activation of gender stereotypes is predicted by Mental Models, which suggests that people build a mental representation of the situation in the world described by the text by using both explicitly stated information in the text and implicit information using

world knowledge such as stereotypes (e.g., Garnham, 2001; Johnson-Laird, 1983; Zwaan & Radvansky, 1998). Thus, when people encounter the word *surgeon*, they infer that this person is male even if this inference is not necessary to establish coherence in the text (Garnham, 2001; Garnham et al., 2002; Sanford, 1985). However, this type of activation is not necessarily predicted by the Minimalist Account, which suggests that inferences that do not help to build local coherence in the text or are not readily available for the comprehender are not routinely made (McKoon & Ratcliff, 1986, 1988, 1992).

In addition, the Minimalist Account and Mental Models Theory make different predictions about the number of inferences created during online processing. The Minimalist Account suggests that comprehenders make a minimum number of inferences, only those necessary for establishing textual coherence (McKoon & Ratcliff, 1986, 1988, 1992). Therefore, it suggests that after people have made a bridging inference in order to integrate the currently processed information with the previous context, there is no need for additional inferences to revise the coherence relationship. However, Mental Models suggests that people update and modify the situation model in an incremental manner during discourse processing and therefore, such additional inferences do occur during online language processing (Garnham, 2001; Garnham et al., 2002; Johnson-Laird, 1983; Zwaan & Radvansky, 1998).

In Study I, the following questions were addressed: (1) Do listeners elaboratively activate gender stereotype information in role names during online spoken language comprehension? (2) Is gender stereotype information used to make additional inferences to revise an already established coherence relation? Short, three-sentence long stories in Finnish were presented auditorily to the participants. For example, when people were listening to Story (32) consisting of a stereotypically male role name *chimney sweep*, they saw pictures of a male and female character, a picture of an instrument mentioned in Sentence 2 (snow shovel), and a picture of the noun mentioned last in Sentence 2 (roof), see Figure 3 (Chapter 2.3.1).

(32) Kuvissa näet Sinikan, 35-vuotiaan naisen Jyväskylästä ja Mikon, 40-vuotiaan miehen Tampereelta. Pihatöiden lomassa Sinikka arvioi Mikon kanssa nuohoojan kohtaamia vaaratilanteita liukkailla katoilla. Kouluttauduttaan nuohoojaksi hän oli oppinut monia keinoja hoitaa työnsä turvallisesti.

‘On the screen you see Sinikka, a 35 year-old woman from Jyväskylä and Mikko, a 40 year-old man from Tampere. While doing year work Sinikka evaluated with Mikko the dangerous situations a chimney sweep gets into on slippery roofs. After having graduated to become a chimney sweep, s/he had learned many ways to work safely’

In the second sentence, a gender stereotype noun such as chimney sweep was presented generically, i.e., it referred to any chimney sweep rather than to a specific person; moreover, it co-referred with neither the female nor the male character

mentioned in the story. Therefore, an increased probability to fixate the male character after hearing *chimney sweep* would suggest that participants elaboratively activate the stereotype that chimney sweeps are usually male. Indeed, eye movements following the stereotype noun onset in the second sentence showed more fixations to the stereotypically consistent character (male in the above example) than to the stereotypically inconsistent character, providing evidence for elaborative activation of gender stereotypes, as suggested by Mental Models Theory (Garnham, 2001; Garnham et al., 2002; Johnson-Laird, 1983; Zwaan & Radvansky, 1998). The stereotype effect is depicted in Figure 2.

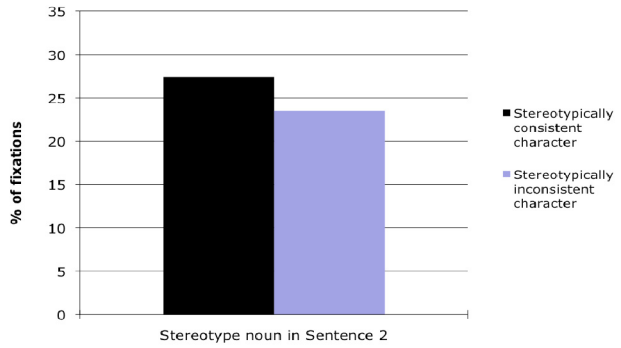

Figure 2. Elaborative activation of gender stereotypes in the second sentence of Study I.

The third sentence tested whether people use gender stereotype information to revise an already established coherence relation (see Figure 3). The sentence started with an anaphoric verb (e.g., *kouluttauduttuaan* 'after having graduated he/she') followed by the same stereotype noun as in Sentence 2. The verb form contained a gender-ambiguous third person anaphoric suffix –AAN. The suffix indicates that the verb is related to one of the two characters mentioned in the story, but because Finnish lacks gender marking it can refer to either of them. In order to make one of the characters the preferred antecedent of the anaphoric verb, one of the two characters was made more salient by mentioning him/her before the other as the subject of the second sentence (e.g., Gernsbacher & Hargreaves, 1989; Järvikivi et al., 2005; Kaiser & Trueswell, in press). Thus, comprehenders were assumed to interpret the anaphoric suffix as referring to the salient character, which should result in more looks to this character relative to the less salient character. This was confirmed by participants' eye movements: They fixated the salient character more often than the less salient character when hearing the anaphoric verb. Interestingly however, during the following gender stereotype noun, people started to fixate the stereotypically consistent character more frequently than the inconsistent character – irrespective of whether this character was linguistically salient or not. This result confirmed that after comprehenders made a bridging inference using saliency information at the anaphoric verb, they subsequently revised it at the stereotype noun using gender stereotype

information. This finding does not fit well with the Minimalist Account, which suggests that people make only the minimum number of inferences in order to resolve local incoherence (McKoon & Ratcliff, 1986, 1988, 1992), but the finding is in line with Mental Models Theory, which assumes that comprehenders make inferences continuously during discourse processing (Garnham, 2001; Garnham et al., 2002; Johnson-Laird, 1983; Reynolds et al., 2006; Zwaan & Radvansky, 1998).

Figure 3. Stereotype effect in the third sentence of Study I.

Figure 3 depicts the stereotype effect found in the third sentence. The result showed that people used both salience and gender stereotype information very quickly. The timing of the salience effect observed during the anaphoric verb implies that people did not wait until they heard the word-final anaphoric suffix before starting to build a link between the most salient character in the preceding context and the verb phrase. Similarly, the gender-stereotype effect at the stereotype noun was observed before the noun's acoustic offset. These effects provide evidence for highly incremental language processing; people update their interpretation of the text at each word in order to achieve maximal coherence. I have named this the Principle of Maximal Coherence. According to this principle, people try to maximise coherence in the text by establishing and updating coherence relations as soon as they can by using a variety of coherence cues. This process explains why both salience and stereotype information had such rapid effects in Sentence 3.

In sum, Study I showed that listeners activate gender stereotypes during online discourse processing even if this does not result in greater discourse coherence, as evidenced by the stereotype activation in Sentence 2. In addition, listeners also use gender stereotypes to update previously established coherence relations by making more inferences than would be necessary for coherence building. Study I also demonstrates that the visual-world eye-tracking method is a very useful method for studying both elaborative activation of semantic features as well as backward inferences during discourse comprehension and for disentangling these two processes from each other.

3.2 Activation and persistence of implicit causality (Study II)

It has been debated at what point the causality information implicit in interpersonal verbs such as *frighten* and *fear* becomes activated in sentences such as *John frightened Bill, because he…* and *John feared Bill, because he.* Two opposing accounts have been suggested: According to the Immediate Focusing Account, implicit causality information is activated shortly after encountering an implicit causality verb, and it is used to focus one of the referents at the expense of the other, thus producing an effect on reference resolution immediately when a pronoun is encountered (Greene & McKoon, 1995; Koornneef & Van Berkum, 2006; MacDonald & MacWhinney, 1995; McKoon et al., 1993; Van Berkum, et al., 2007). In contrast, much later activation of implicit causality information is suggested by the Clausal Integration Account, which states that implicit causality information has an effect on the comprehension process only during the clausal integration phase when implicit causality information in the main clause is integrated with the explicitly-stated causal information of the subordinate clause (Garnham, 2001; Garnham et al., 1996; Stewart et al., 2000).

These two accounts make different predictions about the activation of implicit causality information in discourse. The immediate focusing account predicts that a focusing effect due to activation of implicit causality information should already be measurable before the pronoun. In addition, it proposes that this focusing might lead to a prioritised use of this cue compared to structural cues when resolving a pronoun antecedent. In contrast, the clausal integration account suggests that implicit causality should not affect the relative activation of the potential antecedents until people have enough information to determine whether the information in the subordinate clause is in agreement or disagreement with the implicit causality bias expressed in the main clause.

As previous studies have failed to clearly differentiate between the elaborative activation of verb semantics versus the later use of semantics during integration, Study II was carried out to examine whether (1) implicit causality information is activated on the basis of verb information prior to encountering the pronoun and the causal connective *because*; or whether (2) its activation occurs only after the pronoun, that is, when this information is needed to resolve the referential relationship between the ambiguous pronoun and its potential antecedents. The previous literature suggests that activating implicit causality prior to the pronoun makes it possible to use it as a focusing cue immediately at the pronoun in reference resolution. However, it has also been suggested that the activation of implicit causality information does not occur without causal connectives (e.g., Ehrlich, 1980; Stevenson et al., 1994, 2000). Thus, Study II also investigated as part of question (1) whether activation of implicit causality information occurs prior to the conjunction *koska* 'because' (see 33 below) or only after it but still prior to the pronoun.

Participants listened to stories such as (33) in Finnish while their eye movements to the picture of a guitarist and a butler (in addition to pictures of two locations mentioned in the story) were measured. In order to answer question (1), the second sentence contained a subject biasing stimulus-experience verb or object biasing experiencer-stimulus verb created using the same verb stem.

(33) Kitaristi oli valmistautumassa illan esitykseen. Hovimestari **pelkäsi/pelotti** kitaristia ravintolasalissa, **koska** koko päivän **hän (hänet)** kummallista kyllä oli….

'The guitarist was preparing for the night's performance. The butler **feared/frightened** the guitarist in the dining room, **because** for the whole day **s/he (or him/her[5])** curiously enough had…'

Eye movements to the target characters revealed that shortly following the verb people were more likely to fixate the bias-consistent character than the bias-inconsistent character, as shown in Figure 4. Following the verb onset, the probability of fixating the subject character that was just mentioned was high following both subject- and object-biased verbs. However, shortly following the verb (900 ms from its onset) the probability of fixating the subject after subject-biased verbs remained higher than the probability of fixating the object. In addition, the fixations toward the object were similarly affected by the verb bias: the probability of fixating the object following object-biased verbs was higher than following subject-biased verbs. This finding supports the Immediate Focusing Account (Greene & McKoon, 1995; Koornneef & Van Berkum, 2006; McKoon et al., 1993; Van Berkum et al., 2007) but provides evidence against the Clausal Integration Account (Garnham, 2001; Garnham et al., 1996; Stewart et al., 2000). Interestingly, the activation of implicit causality information occurred even before the causal connective *because*. This result is inconsistent with previous results suggesting that activation of implicit causality does not occur without such a connective (e.g., Ehrlich, 1980; Stevenson et al., 1994, 2000; cf. McDonald & MacWhinney, 1995).

Although implicit causality was activated following the verb and before the causal connective, this in itself does not necessarily imply that it is used when interpreting an ambiguous pronoun later in the sentence. In order to test the persistence of implicit causality and its use during pronoun resolution, the gender-ambiguous pronoun *hän* 's/he' or *hänet* 'him/her' was presented in the second clause of the target sentence. We investigated whether implicit causality has an effect only after structural information is used or whether it is used at the same time or even earlier than structural factors: Previous research has shown that ambiguous pronouns are preferentially interpreted as coreferential with the subject and first-mentioned antecedent of the preceding clause (Crawley et al., 1990; Frederiksen, 1981; Gernsbacher & Hargreaves, 1988; Järvikivi et al., 2005) as well as with the antecedent that has the same grammatical role as the pronoun (e.g., Chambers & Smyth, 1998; Grober, Beardsley, & Caramazza, 1978; Sheldon, 1974; Smyth, 1994). Therefore, both the subject and the object form of the gender ambiguous 3rd person pronoun, *hän* 's/he' or *hänet* 'her/him' were used. In addition, in order to make the structural factors competitive with implicit causality information in pronoun resolution, a topic shift was introduced in the beginning of the second sentence: The first-mentioned subject and topic of the first sentence (e.g., *guitarist*, Example (33))

[5] The object pronoun *hänet* 'him/her' is grammatically correct in this position in Finnish.

was mentioned in the second sentence as a grammatical object, while a new entity was introduced as the first-mentioned subject and the linguistic topic (e.g. *butler,* Example (33)) and thus making it easily available as an antecedent, as suggested by previous literature (e.g., Bestgen & Vonk, 2000; Vonk et al., 1992).

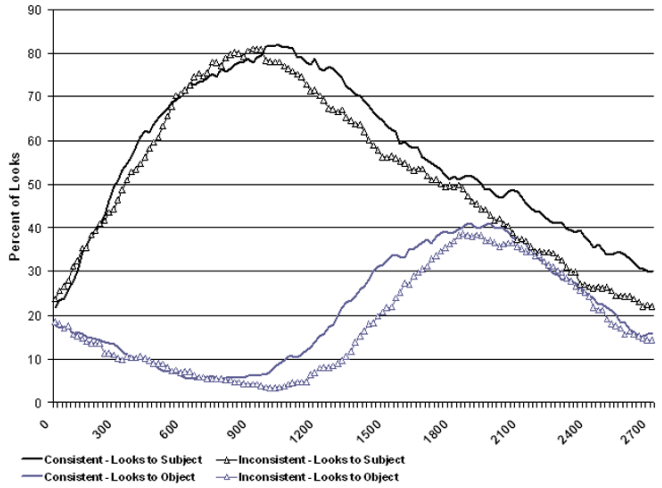

Figure 4. Implicit causality effect from the acoustic onset of the implicit causality verb in Study II.

Eye movements to the depicted characters following the pronoun onset revealed that even though implicit causality became activated prior to the pronoun and highlights one of the referents over another, this focusing may not function as a priority cue guiding pronoun resolution - at least when the structural cues are strong enough (see Figure 5). Although an implicit causality effect was found during the first time segment (0-300 ms) following the pronoun onset, it is unlikely that this reflects any effect related to pronoun resolution as it takes around 200 ms to plan and execute a saccade (Matin et al., 1993). After this time segment, the main effect of implicit causality was absent until 900 ms after the pronoun onset. However, an interaction between implicit causality information and grammatical role were found already 600 ms after the pronoun onset. The interaction showed an implicit causality effect for the first-mentioned subject but not for the object (see also McDonald & MacWhinney, 1995; Stewart & Gosselin, 2000; cf. Long & DeLay, 2000). It is likely that the privileged status of the first-mentioned entity (Gernsbacher & Hargreaves, 1988) and the topic shift (e.g., Bestgen & Vonk, 2000; Vonk et al., 1992) may have made it easier to detect the implicit causality effect with subject- than object-biasing verbs (see also Stewart & Gosselin, 2000). The difference in the information status of subjects versus objects caused by the topic shift may also have highlighted the main effect of grammatical role that was observed in all time windows, as evidenced by more looks toward subject than object entities. However, it is worth noting that the

effect of grammatical role did not differentiate between subjecthood, first-mention priority and topicality.

Finally, the results following the pronoun onset did not show an interaction between implicit causality and pronoun type, indicating that implicit causality information had the same effect on pronoun resolution regardless of the pronoun type (*hän* 's/he' vs. *hänet* 'him/her'). On the other hand, the results showed an interaction between grammatical role of the antecedent and pronoun type, which is a replication of the effect of grammatical role parallelism (Chambers & Smyth, 1998; Smyth 1994). Interestingly, this effect was observed after the implicit causality effect, which suggests that the use of implicit causality information is not delayed relative to the influence of structural parallelism.

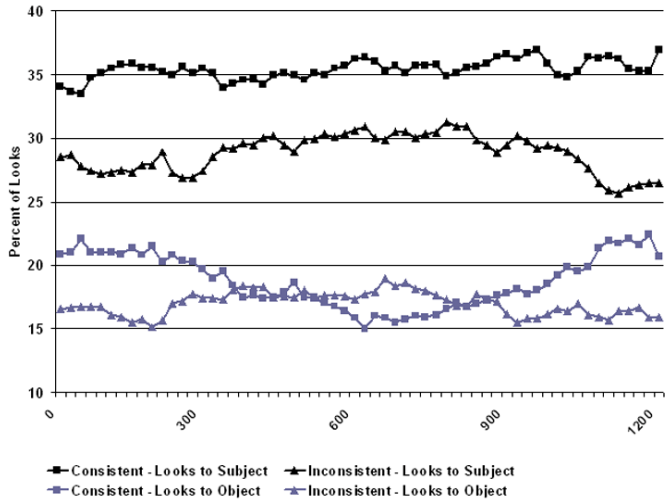

Figure 5. Interaction between implicit causality and pronoun type following the pronoun onset in Study II.

In sum, Study II showed that implicit causality information is inferred and activated as soon as people hear an implicit causality verb in discourse and even before an explicit trigger of causality, such as the causal connective *because* is presented. Despite this early activation, implicit causality information may not have a privileged status in pronoun resolution, but may instead interact with, or be temporally masked by, other structural and discourse cues (such as a topic shift) used during the processing of pronouns. In addition, Study II provided further evidence that the visual world eye-tracking paradigm is a suitable method for investigating both forward and backward inferencing during online discourse processing.

3.3 Proto-roles in pronoun resolution (Study III)[6]

Similarly to adults, there is evidence that young children are sensitive to structural factors when assigning antecedents to pronouns. However, it is unclear whether young children (aged three) are able to take advantage of semantic prominence due to the verb during referential processing. Therefore, I extended my research to children's online language processing as well.

Earlier studies of child language have failed to agree on the kinds of cues children use to draw inferences (Arnold et al., 2005, 2007; Song & Fischer, 2005, 2007). Previous evidence from pronoun resolution has revealed that children from age two onwards use cues like subjecthood and order-of-mention in assigning antecedents to ambiguous anaphors (Song & Fisher, 2005, 2007; but see Arnold et al., 2005, 2007). Importantly, these studies have not shown what kind of role semantics has during pronoun resolution. In Study III it was investigated whether children aged three are already sensitive to semantic cues during pronoun resolution.

One aspect of semantics, namely transitivity of verbs (Dowty, 1991; Hopper & Thompson, 1980; Kako, 2006) was examined in Study III. It has been argued that one of the most basic semantic properties of verbs is transitivity (Hopper & Thompson, 1980). The degree of transitivity of a given verb can be defined by how many prototypically agent-like or patient-like features, or proto-role properties, its arguments (subject and object) exhibit (Dowty, 1991; Kako, 2006). Hopper and Thompson (1980) have demonstrated that the lower the degree of transitivity of the verb is, the more backgrounded its arguments are in discourse; and, vice versa, the higher the transitivity of the verb, the more foregrounded, and therefore more accessible its arguments are.

Children were presented with four-sentence stories in English, as illustrated in Figure 6. The critical pronoun occurred in the initial position in the third sentence (*he did something very silly*), which was identical across the different conditions of each story. Prior to the experiment the children were instructed by saying "*Some of them might do something silly now. Let's see who does something silly*". In addition, the first sentence was identical in the two conditions except for the order of the nouns phrases, which was counterbalanced by splicing the same verb into each story.

[6] The discussion here refers to the accepted, final version of the article. Note that the appendix includes the original submitted version of the manuscript due to the copyright restrictions.

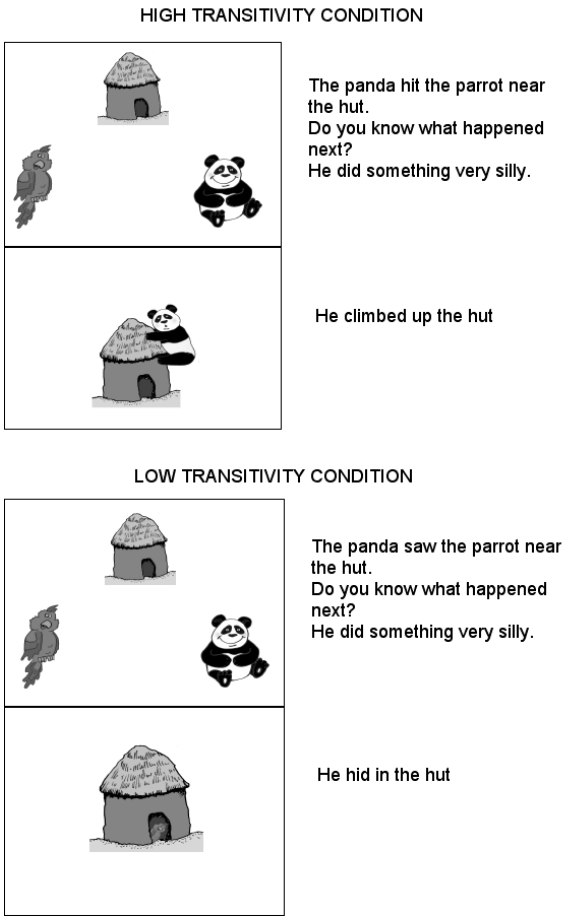

Figure 6. Sequences of pictorial and linguistic simuli used in the high and low transitivity conditions in Study III.

Based on the theories of proto-roles presented in the Introduction, it was hypothesised that when children hear the pronoun *he* after a high transitivity sentence, it is likely that both subject and object (the panda and the parrot in the story in Figure 8) are highly accessible as antecedents for the pronoun as they exhibit several prototypically agent- and patient-like features. In contrast, following the low transitivity sentences, neither the subject nor object are semantically highly prominent, and are thus less accessible as antecedents as they do not exhibit many prototypically agent- or patient-like features. In other words, both subject and object are more foregrounded following the high transitive verbs than following the low transitive verbs. Moreover, unlike in the high transitive condition, as the objects following the low-transitive verbs did not exhibit many, if any at all, of the prototypically object-like properties, it was hypothesised that following the low-transitive verbs, the subject (the panda in the story in Figure 6) would be more foregrounded and therefore more accessible as an antecedent for the pronoun than the

object (the parrot in the story in Figure 6). Thus, the difference between subjects and objects was expected to be larger in the low transitive than in the high transitive condition. Moreover, it was hypothesised that if children are sensitive to structural cues, they might show a general preference for the subject and first-mentioned character irrespective of the verb type.[7]

Eye movements to the target characters confirmed the above hypotheses: With high transitivity verbs there were more looks toward both characters following the pronoun onset than in the low transitive condition. Low transitivity verbs stimulated less looking behaviour overall to the target characters than high transitivity verbs, and even more importantly, particularly towards the object of the low transitivity verb, as indicated by the interaction between structural and semantic prominence (see Figure 7). This interaction was predicted based on the transitivity hierarchy (Hopper & Thompson, 1980; Rose, 2005).

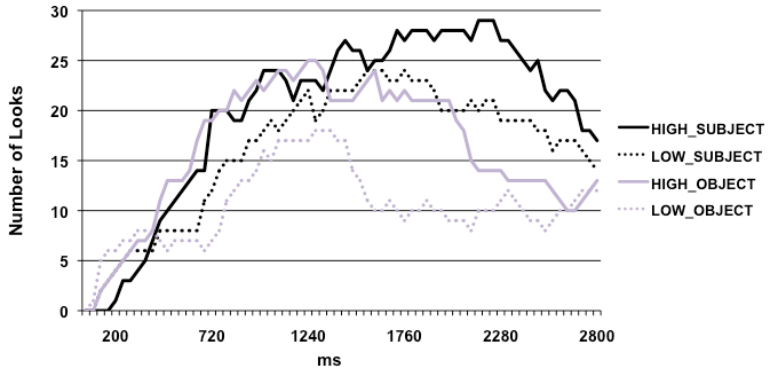

Figure 7. Time course depicting both the main effect of transitivity and the interaction between transitivity and grammatical role as number of looks to subjects and objects in high and low transitivity conditions following the pronoun onset in Study III.

In sum, Study III showed that children as young as three are able to take advantage of semantic prominence inherent in transitive verbs. Thus, Study III complements Study II by adding one more dimension to the list of semantic prominence factors that adults and children take into account during pronoun resolution.

3.4 Gender stereotypes in anaphoric and non-anaphoric inferences (Study IV)

Even though anaphoric and non-anaphoric bridging inferences often serve the same purpose of establishing a coherent interpretation of discourse, it has been

[7] Even though the last sentence, in which the referent was disambiguated, was not of interest, the disambiguated character was counterbalanced so that the first-mentioned subject was referred to equally as often as the second-mentioned object both in high and low transitivity conditions.

suggested that drawing anaphoric inferences differs from drawing non-anaphoric inferences (Clark & Haviland, 1977; Garrod, Freudenthal, & Boyle, 1994; Garrod et al., 1990; Greene et al., 1992; Fincher-Kiefer, 1995). Because anaphors explicitly signal the need for building a coherence relation, the process is assumed to be a matching process between an explicit anaphoric marker and the most plausible antecedent to create the semantically most appropriate text representation (Garrod et al., 1990; 1994). Therefore, it is hypothesised that semantic factors may play a more important role when resolving anaphoric expressions compared to making non-anaphoric inferences.

In Study IV gender stereotype information was pitted against salience cues in order to resolve the relative importance of the semantic cues during anaphoric and non-anaphoric inferences. Earlier studies suggest that both salience and gender stereotypes provide strong cues for anaphoric and non-anaphoric inferences (Almor, 1999; Gordon et al., 1993, Gordon & Scearce, 1995; Duffy & Keir, 2004; Kennison & Trofe, 2003; Kreiner et al., 2008; Osterhout et al., 1997; Sturt, 2003a,b). However, earlier studies have not tested the effect of gender stereotypes on anaphoric and non-anaphoric inferences. Moreover, the effect of gender stereotypes has not been compared to the effects of other strong cues such as discourse/structural salience. Salience has been found to be important during anaphor resolution, but it has been hypothesised that it is important during non-anaphoric inferencing, given that salient discourse entities are easier to retrieve than less salient ones (e.g., McKoon et al., 1996; McKoon, Ratcliff, Ward, & Sproat, 1993; O'Brien et al., 1990, 1995).

Study IV tested the relative importance of semantics for making anaphoric and non-anaphoric inferences by comparing semantic effects to salience effects. Participants heard stories like (34). In these stories the last sentence always started with a gender stereotype noun like *motorcycle*. This word was presented either in the anaphoric or non-anaphoric form: The word either contained the anaphoric suffix -(n)SA, which is a gender-ambiguous possessive anaphoric suffix signalling that the motorcycle is related to one of the previously mentioned characters, or contained no anaphoric suffix. Without the suffix, people had to draw a non-anaphoric bridging inference to relate the motorcycle to either of the characters.

(34) Ruudulla ovat Äänekoskelta kotoisin oleva Sirpa ja Kauniaisista kotoisin oleva Sami. Syksyisellä iltanuotiolla Sirpa kommentoi Samin kanssa sitä, millaisia onnettomuuksia viime kesänä oli tapahtunut läheisellä metsätiellä. **Moottoripyörän(sä)** kolaroiminen oli aiheuttanut hänelle suuret taloudelliset vahingot.

'On the screen you will see Sirpa (female), who comes from Äänekoski and Sami (male), who comes from Kauniainen. In the fall around a camp fire, Sirpa talked with Sami about what kind of accidents happened last summer on the nearby forest road. Crashing **(his/her) motorcycle** caused him/her great financial damage.'

The probability of fixating the female and male characters following the target word (*motorcycle*) showed that gender stereotype information was inferred quickly: the probability of fixating the stereotypically consistent character increased shortly following the stereotype word. In addition, the effect of discourse/structural salience was also found: people tended to look back to the character that was the most salient in the previous context.

Importantly however, the results also showed that the stereotype effect continued to be strong during later time windows when people were resolving anaphoric expressions, whereas in the non-anaphoric condition salience eventually outperformed gender sterotype information. These effects can be clearly seen from Figures 8 and 9, of which the former depicts the stereotype effect following the target word and the latter the structural salience effect following the target word, separately for the anaphoric and non-anaphoric conditions.

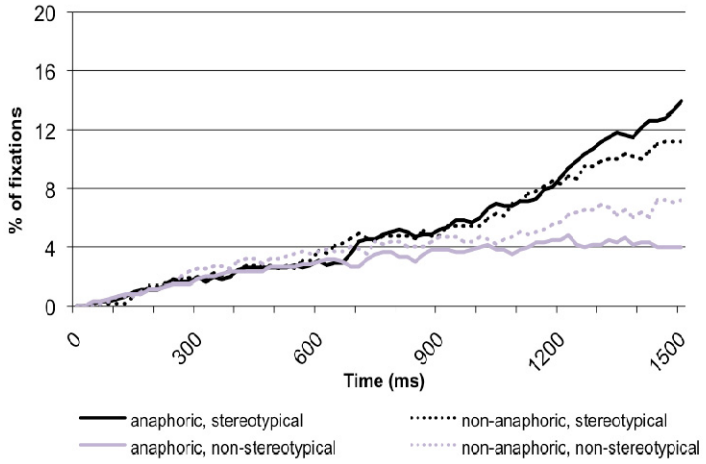

Figure 8. Time course of the semantic effects starting from the target onset when making anaphoric and non-anaphoric inferences in Study IV.

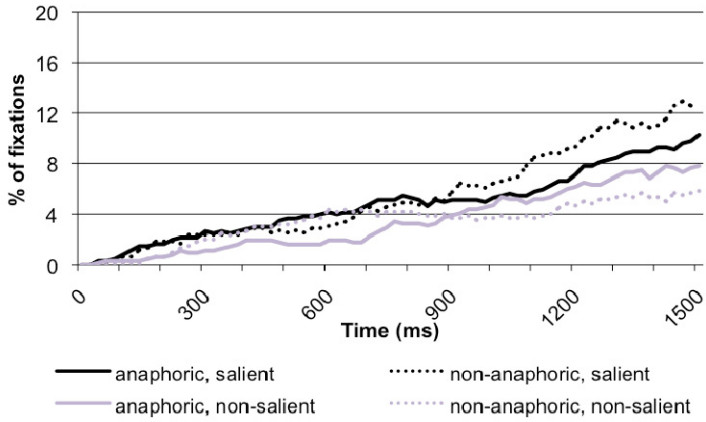

Figure 9. Time course of the discourse/structural salience effects starting from the target onset when making anaphoric and non-anaphoric inferences in Study IV.

These results lend support to the view that anaphoric and non-anaphoric inferences differ during online discourse processing (Clark & Haviland, 1977; Garrod et al., 1990, 1994; Greene et al., 1992; Fincher-Kiefer, 1995). Importantly, Study IV showed that semantic information plays a more important role during anaphoric inferencing than it does during non-anaphoric inferencing, whereas discourse/structural cues are more prominently used during non-anaphoric than anaphoric inferencing. This implies that the relative importance of semantic cues in comparison to discourse/structural cues is dependent on whether the inference is explicitly triggered by an anaphoric marker or not: When such a trigger is present, semantic cues outweighs salience, whereas when an explicit trigger is absent, salience outweighs semantics.

4 GENERAL DISCUSSION

Section 4.1 highlights the relevance of the findings of the thesis with respect to the importance of semantics during inferencing, particularly elaborative activation of gender stereotypes, the role of implicit causality information, and interactions between semantics and other linguistic cues during backward inferencing. These findings suggest modifications for the current theories of language comprehension. Theoretical modifications are suggested in order to account for forward inferences in online processing as well as to account for the role of semantics in relation to other (discourse and structural) cues.

4.1 Main findings and their theoretical implications

The studies showed that people activate different types of semantic information elaboratively during online language processing (Studies I and II). In addition, the studies showed that semantics plays an important role in anaphor resolution (Studies II, III and IV), but also that anaphor resolution is sensitive to other discourse/structural cues (Study II). Furthermore, Study IV demonstrated a more pertinent role of semantics during anaphoric bridging inferences than non-anaphoric inferences. Finally, Study I demonstrated that semantic information is used quickly to revise earlier established coherence relationships. In the following, I will discuss these findings and their theoretical implications in detail.

4.1.1 Elaborative activation of semantics

Studies I and II showed that people activate semantic information elaboratively during online language comprehension. This finding was observed with respect to two types of semantic information: gender stereotypes and implicit causality.

Activation of gender stereotypes. Consider the following example, which was also presented in the Introduction (Example 1):

> (35) The chimney sweep was climing onto the roof. He was worried whether it could be slippery.

This example highlights an aspect of discourse that has been studied in psycholinguistics already for several decades, i.e., whether gender stereotypes are activated when encountering a stereotype word such as *chimney sweep* or only later when a subsequent pronoun referring to *chimney sweep* needs to be integrated with the antecedent in the preceding context. Earlier studies have failed to resolve whether gender stereotype information is activated when the word is first encountered in the discourse or only later at the point when the pronoun referring to the stereotype word

is assigned an antecedent (Duffy & Keir, 2004; Garnham et al., 2002; Reynolds et al., 2006, Sanford, 1985). Study I showed that when people heard a generic gender-stereotype noun such as *chimney sweep* in a discourse, their visual attention was directed more toward a male than a female character depicted on the screen – even though it was impossible to infer that this particular character was a chimney sweep (see example 32, Chapter 3.1). This finding is taken as evidence for elaborative activation of gender stereotype information during discourse processing.

Several earlier studies of gender stereotypes have also proposed that gender stereotypes are activated elaboratively in discourse (Carreiras et al., 1996; Garnham, 2001; Garnham et al., 1996, 2002; Reynolds et al., 2006; Sanford, 1985). However, as noted earlier, due to methodological limitations, these studies have failed to present unequivocal evidence for such activation during online comprehension. Study I was the first to show that people do indeed activate them in a forward fashion and that their activation is not dependent on an explicit prompt such as a pronoun. The results from Study I also allowed us to extend the results from studies that showed that elaborative activation of gender stereotypes occurs when stereotype words are presented in isolation (Banaji & Hardin, 1996; Cacciari & Padovani, 2007; Oakhill et al., 2005) by showing that such elaborative activation also occurs during natural discourse processing.

The elaborative activation of gender stereotypes is in line with Mental Models theories (e.g., Garnham, 2001; Garnham et al., 2002; Johnson-Laird, 1983; Zwaan & Radvansky, 1998), which suggest that during online language comprehension people not only activate literally stated information but also general world knowledge such as stereotypes, and that they may do so elaboratively when encountering gender-stereotype words in discourse. However, elaborative activation of gender stereotypes is not predicted by the Minimalist Account (McKoon & Ratcliff, 1986, 1988, 1992), which suggests that inferences that do not help building local coherence in the text are not routinely made. As mentioned in the Introduction, according to the Minimalist Account only those inferences for which the relevant information is 'readily available' are made (McKoon & Ratcliff, 1992). Even though it is not completely clear what exactly 'readily available' means and how readily available gender stereotypes are in the comprehender's mind, it can be argued that our findings are not consistent with this hypothesis: McKoon and Ratcliff (1992) restricted readily available information to include the explicitly stated words in the current or immediately preceding context and the propositions derived from them that are held in short-term memory (Kintsch & Van Dijk, 1978). As gender stereotypes cannot be derived from text propositions, it can be argued that they do not belong to the category of 'readily available' in the Minimalist Account and thus should not have been activated elaboratively during online comprehension.

In sum, earlier theories have made contrasting predictions about whether gender stereotypes are activated elaboratively during discourse comprehension. The use of the visual world eye-tracking method turned out to be a fruitful method to disentangle elaborative activation from later integration and to demonstrate that gender stereotypes are indeed activated elaboratively; a hypothesis that had not been resolved with previously used methods.

Activation of implicit causality. Just as the visual world eye-tracking paradigm was used to examine gender stereotypes, it was also used in Study II to demonstrate elaborative activation of implicit causality information related to interpersonal verbs (e.g., *frighten*/*fear*). Shortly following the verb, implicit causality biases of the verbs affected people's eye movements: Following the NP1-biased verbs such as *frighten* the probability of fixating the NP1 (the subject) was higher than after NP2-biased verbs. Similarly, the fixations toward the object were affected by the verb bias: Following the NP2-biased verbs such as *fear* the probability of fixating the NP2 (the object) was higher than after NP1-biased verbs.

This finding supports the hypothesis derived from the Immediate Focusing Account (Greene & McKoon, 1995; Koornneef & Van Berkum, 2006; MacDonald & MacWhinney, 1995; McKoon et al., 1993; Van Berkum et al., 2007) that implicit causality is activated prior to the pronoun. This account predicts early effects of implicit causality bias during pronoun resolution: it is argued that the effect of implicit causality at the pronoun is possible because implicit causality information is quickly inferred and used to highlight one entity over another. This highlighted entity is therefore immediately interpreted as the antecedent for the following pronoun. Although the focusing account does not clearly specify how early the activation of the bias-consistent discourse referent is, it suggests that the activation may occur even prior to the pronoun. Such early activation has not been shown in earlier studies - again, partly due to methodological limitations. However, Study II showed that implicit causality is indeed activated before the pronoun is heard.

Interestingly, implicit causality became activated even prior to the causal conjunction *because*, which suggests that such activation is not dependent on any explicit prompts from the discourse, not even a causal conjunction as proposed by Stevenson et al., (1994, 2000). Instead, elaborative activation of implicit causality is due to the verb itself and does not require any additional prompts (see also McDonald & MacWhinney, 1995, for a similar hypothesis).

Similar to the activation of gender stereotypes, elaborative activation of implicit causality can also be explained with Mental Models Theory. This theory does not strictly limit the type of semantic factors that can be activated elaboratively but instead allows both propositional meanings inherent in the discourse as well as meanings derived from the mental representations that people have created on the basis of events described in the text (Garnham, 2001; Johnson-Laird, 1983; Zwaan & Radvansky, 1998) to be used in discourse comprehension. Although the origin of the implicit causality bias is not clear, it is possible that such a bias is learned via prior exposure: The probability of encountering a continuation referring to NP1 or NP2 following verbs such as *frighten* and *fear* might have created such a bias. Regardless of the ultimate basis for such a bias, these models assume activation of this bias as soon as the verb is encountered in the discourse. Recent mental simulation accounts suggest that encountering such a verb leads to a mental simulation of the event (Zwaan & Madden, 2004; Zwaan & Taylor, 2006; see also Fischer & Zwaan, 2008 for a recent discussion of simulation accounts).

4.1.2 Semantics in backward inferencing

The importance of semantics in backward inferencing and anaphor resolution was also examined in the present thesis. All experiments investigated some aspects of this question. The studies consistently showed that semantics plays an important role in backward inferencing: Study II showed an implicit causality effect during pronoun resolution, Study III showed an effect of proto-roles during children's pronoun resolution, Study IV showed that semantics has a more prominent role in anaphor resolution than during non-anaphoric backward inferencing, and finally, Study I showed that gender stereotype information is used to revise an earlier drawn inference.

Implicit causality and protoroles in pronoun resolution. Study II showed that implicit causality is used as a cue in pronoun resolution and its timing showed that it was used prior to the clausal integration phase. Therefore, these findings are difficult to explain in the framework of the Clausal Integration Account, which states that implicit causality is only activated and used in pronoun resolution in the phase when the explicitly stated causal information of the subclause is integrated with the causality implied in the main clause (Garnham, 2001; Garnham et al., 1996; Stewart et al., 2000). Instead, the finding seems to fit better with the Immediate Focusing Account, which assumes that implicit causality is activated early and used early during pronoun resolution (Greene & McKoon, 1995; Koornneef & Van Berkum, 2006; MacDonald & MacWhinney, 1995; McKoon et al., 1993; Van Berkum et al., 2007).

As noted earlier, the focusing account also assumes that because implicit causality is activated early and used to highlight one of the characters over another, this highlighted character has a privileged role to be interpreted as the antecedent. This assumption also implies that when the pronoun is ambiguous, the character that has become prominent by the focusing function of implicit causality would immediately become interpreted as an antecedent (Koornneef & Van Berkum, 2006; Van Berkum et al., 2007). However, the findings of Study II did not support this assumption. Instead, when a topic shift was introduced in the discourse prior to the pronoun, it made people more likely to shift their attention to the first-mentioned subject after hearing the pronoun irrespective of the implicit causality bias of the verb. This preference temporally masked the effect of implicit causality information on pronoun resolution. Earlier studies have shown that a topic shift makes the shifted topic highly focused and prominent in the discourse and at the same time also suppresses information from the preceding sentence, leading to a reduced activation of the non-topical character (Bestgen & Vonk, 2000; Vonk et al., 1992). It seems that a topic shift masks the effect of implicit causality during the early phases of pronoun resolution. Importantly, this masking is temporary; an implicit causality effect was found again 900 ms following the pronoun onset, thus suggesting that implicit causality contributed to pronoun resolution well before the clause-final integration process proposed by the Clausal Integration Account (Garnham, 2001; Garnham et al., 1996; Stewart et al., 2000).

Study III established an effect of semantics in children's pronoun resolution as well. This study included verbs that differed in the degree to which their subject and

object exhibited proto-role properties of agents and patients, respectively. In the case of high transitive verbs, the subject carried several prototypically agent-like properties while the object carried patient-like properties. In contrast, in the case of low-transitive verbs, subjects exhibited only a few prototypically agent-like properties and patient hardly any patient-like properties. These linguistic properties affected children's eye movement behaviour when they processed an ambiguous pronoun: There was a higher proportion of fixations toward both the agent and patient in the high transitivity condition compared to the low transitivity condition. Importantly also, while in the low transitivity condition the subject was fixated more often than the object, such a preference was not so strong in the high transitive condition. Instead, the results indicated that in the high transitive condition both the subject and object were considered (competing) antecedents for the pronoun.

These findings were expected based on the Proto-Role Hypothesis stating that the more prototypical agent- and patient-like properties subject and object exhibit for a given verb, the more foregrounded they are in discourse and thus the more preferred they are as antecedents for a pronoun (Dowty, 1991; Hopper & Thompson, 1980; Kako, 2006). Interestingly, the transitivity effect arose prior to the effect of discourse/structural cues suggesting that children were faster to use semantic cues than structural cues when interpreting ambiguous pronouns in discourse. This leads to the question why structural cues were delayed relative to semantic cues in this study but not in the study of implicit causality. On first sight, these studies seemed to provide contradictory findings regarding the time course of the use of semantics as a cue. However, I believe that these findings may be due to the different strengths of the discourse/structural cues: In the study of implicit causality the shifted topic may have served as a stronger cue than the first-mentioned subject (and topic) in the study of proto-roles. In support of this assumption, Järvikivi, Van Gompel, and Hyönä (submitted) showed that implicit causality preceded structural cues when the first-mentioned character was not the topic. Furthermore, the interaction between the grammatical role and pronoun type (subject vs. object pronoun) in Study II showed that the object antecedent was fixated more often than the subject pronoun following an object pronoun, while the subject was fixated only marginally more often than the object following a subject pronoun. Interestingly, this observed effect of structural parallelism (see also Chambers & Smyth, 1998; Smyth, 1994) occurred later than the main effect of implicit causality, suggesting that semantic cues are not delayed compared to the effect of structural parallelism.

Taking all of these findings together, it seems that semantics has an important role in backward inferencing. However, the strength of semantic cues is dependent on the relative strength of other, here discourse/structural, cues. When the structural cues consist of a combination of cues that have been independently shown to be important in backward inferencing, semantic effects may be masked, whereas when the discourse/structural cues consist of a combination of fewer prominent cues, semantics can have an earlier effect than structural factors in pronoun resolution (see 4.1.3 for mapping of the prominence of discourse/structural cues).

These findings are difficult to explain for the structurally motivated theories introduced in Chapter 1. Of these, Structure Building Theory as well as Centering Theory and Discourse Prominence Theory assume that pronouns are resolved by

using either order-of-mention or grammatical role information, or a combination of these cues (Gernsbacher, 1988; 1990; Gernsbacher & Hargreaves, 1989; Gordon et al., 1993; Gordon & Hendrick, 1997a,b, 1998). In addition, the Two-Stage Model proposed for example by Garrod and Sanford (1994) does allow semantic cues to contribute to pronoun resolution but only in the second stage of processing after structural cues are first identified and used (see also Garrod & Terras, 2000). Even though this approach explains the late use of implicit causality found in pronoun resolution in Study II, it is unclear how this approach would explain the finding that implicit causality was inferred even prior to the pronoun. In addition, this approach faces difficulties in explaining the effect of proto-roles that preceded that of structural cues in Study III, as well as other recent findings showing earlier effects of implicit causality than structural cues in sentences where the structural cues are not as prominent as in Study II (Järvikivi et al., submitted; Koornneef & Van Berkum, 2006; Van Berkum et al., 2007).

Gender stereotypes during anaphoric and non-anaphoric inferences. In the previous literature on inferences it has been suggested that because anaphors are explicit markers for coherence building, drawing anaphoric inferences might differ from non-anaphoric inferences (Clark & Haviland, 1977; Garrod, Freudenthal, & Boyle, 1994; Garrod et al., 1990; Greene et al., 1992; Fincher-Kiefer, 1995). Study IV confirmed this hypothesis by showing that the relative importance of semantics compared to discourse/structural cues was different for these two types of processes. While gender stereotypes played a prominent role in anaphoric inferencing, during non-anaphoric bridging inferences were more strongly affected by discourse/structural factors.

As was noted earlier, gender stereotypes have been shown to be important cues in pronoun resolution (Duffy & Keir, 2004; Kennison & Trofe, 2003; Kreiner et al., 2008; Osterhout et al., 1997; Sturt, 2003a,b). Study IV of the present thesis replicated this finding. However, even more importantly, it showed that this cue was more prominent during anaphoric than non-anaphoric inferencing. In addition, the study showed that discourse/structural salience cues were more prominent during non-anaphoric than anaphoric inferencing. The study suggests that an explicit marker, for example, an anaphoric expression, prompts an active search for a semantically most plausible antecedent, whereas when such a marker is absent, people rely more on discourse/structural cues. Although the previous studies had not established clear prominence for discourse/structural cues in anaphoric inferencing, it was expected based on an assumption that it is easier to make a bridging inference between two discourse elements when the earlier element is easier to retrieve given that structurally salient discourse entities are easier to retrieve than less salient entities (e.g., McKoon et al, 1993, 1996; O'Brien et al., 1990; 1995)**.**

Study IV adds an important aspect to the relative importance of semantics when compared to discourse/structural cues. In this study the relative importance of semantics was shown to be dependent on the type of inference drawn. Altogether, these findings suggest that the relative importance of semantics compared to discourse/structural cues is dependent on the strength of the structural cues and also on the type of inference, and thus, it is reasonable to capture both aspects in the models of inference generation.

Gender stereotypes and additional backward inferences. One hypothesis concerning the importance of semantics drawn from the Minimalist Account was tested in Study I. According to the Minimalist Account, only a minimal number of inferences are drawn in discourse in order to achieve a coherent interpretation (McKoon & Ratcliff, 1986, 1988, 1992). Based on this assumption, it was hypothesised that if an anaphoric expression is ambiguous, a discourse/structural cue would be enough to satisfy the comprehender and a coherent interpretation is established by using this cue. Discourse/structural cues were again combined in order to introduce a structurally highly preferred antecedent for a verb phrase anaphor. The results showed that this cue was used to resolve the anaphoric expression. However, when listeners later encountered a gender stereotype role name, the visual attention shifted toward the stereotypically consistent character – irrespective of whether that character was the structurally preferred character or not. This finding suggests that comprehenders may not limit the number of inferences drawn in discourse. Instead, information such as gender stereotypes invites new, additional, inferences in the discourse. This finding goes against the Minimalist Account of inference generation in discourse comprehension (McKoon & Ratcliff, 1986, 1988, 1992). However, the finding can be explained in Mental Models theories, which allows updating of discourse interpretation every time comprehenders encounter new information in the discourse (Garnham, 2001; Johnson-Laird, 1983; Zwaan & Radvansky, 1998).

Taking all of the findings demonstrating the importance of semantics during forward and backward inferencing together, it seems that the Mental Model theories (Garnham, 2001; Johnson-Laird, 1983; Zwaan & Radvansky, 1998) provide the most suitable theoretical framework to explain inference generation in discourse as well as the use and importance of semantics in these processes. However, as noted in the Introduction, this approach is the most flexible of the existing theories, as it allows several types of information to be used as cues in making forward inferences. Although this approach seems generally well suited to explaining inference generation in discourse, the relative importance of semantics compared to discourse/structural cues should be specified more explicitly in these models. I will return to this issue in Section 4.2, where I will propose alternative extensions to these models. Before that, I briefly discuss the discourse/structural findings obtained in the present studies.

4.1.3 Discourse/Structural cues in backward inferencing

Although the use of discourse or structural cues such as order-of-mention, grammatical role and topicality was not the main topic of the thesis, the studies also provided converging evidence for the use of these cues during adult and child pronoun resolution as well as during anaphoric and non-anaphoric inferences.

Study II showed an interaction between structural cues and implicit causality in pronoun resolution. The interaction indicated that the effect of implicit causality was more pronounced for the subject than object antecedent (for similar findings, see e.g., McDonald & MacWhinney, 1995; Stewart & Gosselin, 2000; cf., Long & DeLey, 2000, for an opposite observation). Stewart and Gosselin (2000) noted that it may

well be that the first-mentioned preference exerts an additive effect so that it becomes easier to detect implicit causality effects with the first-mentioned character than the second mentioned character: The NP1 preference for NP1-biased verbs might thus be a combination of implicit causality and order-of-mention cues. As noted in Study II, the first-mentioned subject was also a shifted topic, making that particular character even more salient and thus a preferred antecedent for the pronoun (Bestgen & Vonk, 2000; Vonk et al., 1992). Therefore, it is highly probable that the interaction between structural cues and implicit causality was partly observed because of such a strong discourse/structural cue. Nevertheless, the interaction showed that implicit causality does not operate independently but in collaboration with structural information.

The earlier findings about the use of discourse/structural cues in children's pronoun resolution have turned out to be controversial as well. While Song and Fischer (2005, 2007) established a first-mentioned subject preference even with 2-year old children, Arnold et al. (2005, 2007) failed to show the same with 4-year old children. In the former study, the fact that the first-mentioned subject was repeated at least twice before the pronoun might have added to its salience, and thus contributed to the observed effects. In Study III the first-mentioned subject (and topic) was mentioned only once, but it was nevertheless found to be a significant cue in pronoun resolution. This finding provides further evidence for the use of discourse/structural cues in pronoun resolution with children as young as 3 years old.

In addition, a comparison of the discourse/structural cues in Studies II and III indicates that different factors might contribute to the strength of structural cues: The structurally salient character in Study III was a combination of the three discourse/structural factors introduced in the Introduction: they were first-mentioned characters, grammatical subjects and also linguistic topics of the sentence. However, in Study II the structurally salient character was not only a combination of these three factors, but it was also a shifted topic. Thus, it could be argued that the structural cues were even stronger in Study II than in Study III. As discussed above, such shifting may not only affect the salience or prominence of the focused entity but also operates by suppressing information in the earlier discourse (Bestgen & Vonk, 2000; Vonk et al., 1992). This leads to the less-salient entity being even less salient than it would be without a topic shift. Therefore, the structurally preferred antecedent has become even more preferred by a topic shift. Such strengthening of the structural cues might be the reason why the implicit causality effect was masked during early phases of anaphor resolution.

Finally, Study IV showed an interesting interaction between discourse/structural factors and the type of expression. While gender stereotype information had a prominent role during anaphor resolution, discourse/structural cues were more prominent when drawing non-anaphoric inferences. This finding indicates that effects of structural cues are dependent on the type of referring expression itself (see also Almor, 1999; Kaiser & Trueswell, in press).

4.2 Extending the Mental Models accounts

As noted earlier, the Mental Model theories seem like a fruitful framework to explain inference generation and the role of semantics in discourse processing. In what follows, I sketch two potential models formed by extending the current models in order to explain the findings of the present thesis. These models are not intended to be exhaustive descriptions of inference generation in discourse, but instead are intended to form a basis for future studies needed to make comparisons between the proposed models and to suggest further modifications to them. The main difference between these models lies in the interactional nature of discourse/structural factors and semantic factors in forward inferencing: while Mental Model I does not assume such a link, Mental Model II does so.

4.2.1 Mental Model I

Mental Model I, depicted in Figure 10, describes the sequences of discourse processing and aims to explain what types of inferential processes are operating simultaneously during the course of processing. The highlighted colour represents a processing sequence that refers to processes that are occuring in the current context (Context *N*), whereas the preceding and the subsequent sequences are shadowed. The upper part of the figure illustrates the processes involved in backward inferencing, and the lower part of the figure illustrates the processes involved in forward inferencing. During both types of processes, semantic and discourse/structural cues are represented as separate boxes, and the arrows represent causal relations during the course of processing.

In any given sequence of processing (Context *N*) both backward and forward inferences are drawn simultaneously as indicated by the arrows. Even though Context *N* may refer to a unit consisting of more than one word, the current studies suggest that the processor updates the inferences incorporated in the mental model at each and every word. This is also supported by the finding of an additional inference in Study IV: gender stereotype information was used to revise the initial inference even though that would not have been absolutely necessary in order to build a coherent interpretation. Therefore, the model assumes an extreme version of processing incrementality.

Another reason why the model is highly incremental is because it assumes continuous drawing of forward inferences. The processor works in an optimal way by simultaneously drawing backward and forward inferences. As in the earlier Mental Model theories, this model does not restrict the type of forward inferences drawn: it assumes that both elaborative activation as well as detailed predictions of upcoming events are carried out during online discourse processing.

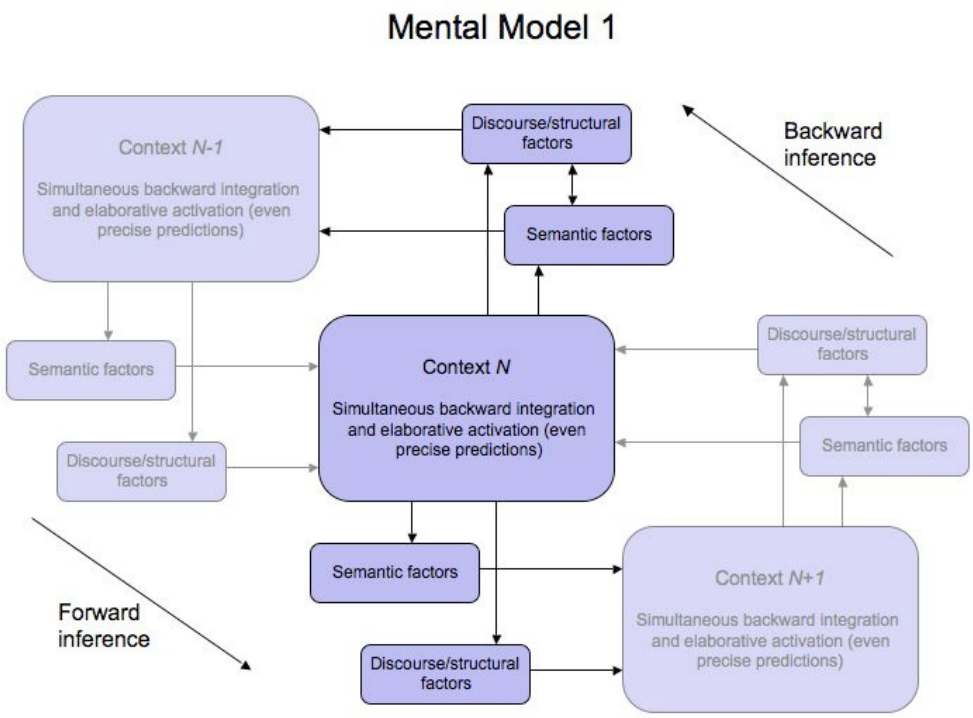

Figure 10. An illustration of the processing sequence postulated by Mental Model I.

An important feature of this model is the assumed interactive nature between discourse/structural and semantic factors in backward inferencing. As shown in Study II and III, semantic cues were weighed in relation to the discourse/structural cues. When the relative strength of the discourse/structural cues was high, as in Study II, the semantic effects were masked early in the integration phase, whereas when the relative strength of the discourse/structural cues was lower, as in Study III[8], semantic factors contributed to pronoun resolution earlier than discourse/structural cues.

However, the interactional nature between discourse/structural cues and semantic cues was not tested in forward inferencing, and thus this model does not assume interactions between these cues during forward inferencing. The underlying idea is that discourse/structural cues provide independent cues to draw forward inferences – either elaborative activation of information or predictions of the upcoming events (see also the Introduction). One example of structural predictions would be for example predictions made based on low level features, for example, the use of transitional probability between adjacent words to predict upcoming words in the discourse (e.g., McDonald & Shillcock, 2003). Similarly, it is assumed that the semantic content directly invites forward inferences without interacting with discourse/structural

[8] While the latter study included a first-mentioned subject that was also the linguistic topic, the former study added topic shift as a further manipulation.

factors. Based on the current studies it cannot be concluded whether this hypothesis of independent effects of semantics and discourse/structural cues (without interaction) is correct or whether these two types of cues interact with each other in forward inferencing. The latter possibility is incorporated in Mental Model 2.

It should be noted here that because these models are not intended as exhaustive explanations of inference generation, discourse and structural factors are mapped as one group of cues and just roughly separated from the semantic cues. Even though discourse/structural cues are combined together, the model does not assume that they behave as one unified component and that the effects of the individual structural factors (order-of-mention, grammatical role, topicality, etc.) cannot be determined independently. In other words, the various structural factors can operate separately and can have different independent weights (see e.g., Järvikivi et al., 2005, for evidence that grammatical role and order-of-mention operate independently in pronoun resolution). Therefore, further studies are needed in order to resolve whether different discourse/structural factors are always operating separately or in interaction with each other. The earlier literature suggests that topicality is inherently part of either subjecthood (as in English, e.g., Lambrecht, 1994) or the first-mentioned entity (as in Finnish, e.g., Vilkuna, 1989).

Similarly, all semantic factors are combined into one single category. However, the model does not assume that all semantic factors contribute to the same extent to inference generation. Interestingly, the present studies investigating three types of semantic information, gender stereotypes, implicit causality and proto-roles, seem to provide converging evidence for the use of semantics in both forward and backward inferences: gender stereotypes and implicit causality seemed to become activated elaboratively, while all three semantic features exhibited an important role in backward inferencing. In the present thesis the elaborative activation of proto-role properties was not tested; whether they are activated elaboratively is an interesting question to be studied in the future.

In addition, the model also includes a direct link both via semantic and discourse/structural factors from the current Context (Context *N*) to Context *N-1*. This was added to the model as it is not clear whether there is always an interaction between discourse/structural and semantic factors in backward inferencing or whether they could also operate independent of one another. It may be that interactions between discourse/structural and semantic cues are only found in highly competitive contexts, as was the case with ambiguous anaphor resolution in the current studies. This is another interesting question for future studies to resolve.

Finally, the model does not make explicit predictions about the timing of the effects, i.e., whether discourse/structural cues always precede semantic cues or vice versa. Instead, the relative importance of one type of cue over another is assumed to be resolved in the interactional link in backward inferencing. As Studies II and III showed, when more structural cues converge, they together have a very early effect and temporarily mask semantic cues (Study II) while when the structural cues are less prominent, semantic cues play a faster role in pronoun resolution (Study III).

4.2.2 Mental Model 2

Mental Model 2, depicted in Figure 11, is a modified version of Mental Model I with only one modification: Mental Model 2 assumes an interactional link between discourse/structural factors and semantic factors in forward inferencing (the red arrows in Figure 11). Mental Model 1 did not assume such a link, but instead assumed that discourse/structural and semantic cues independently affected inferences. This link was added in the second model as it may well be that discourse/structural cues and semantic cues first interact with each other before, based on their relative strengths, forward inferences are made. The assumption here is that not all possible inferences become activated in a forward fashion, but only those that are confirmed by the interactional link. For example, semantic inferences might receive more activation if they are also supported by discourse/structural factors and vice versa. However, when the discourse/structural factors are pointing to another type of inference than semantic factors, it may be that only either of these inferences becomes activated (based on the weights of these cues) or neither of them becomes activated. Again, the current studies were not designed to test this hypothesis, but because this link plays a key role in selecting between Model 1 and Model 2, studies should be planned to test whether this link exists or not.

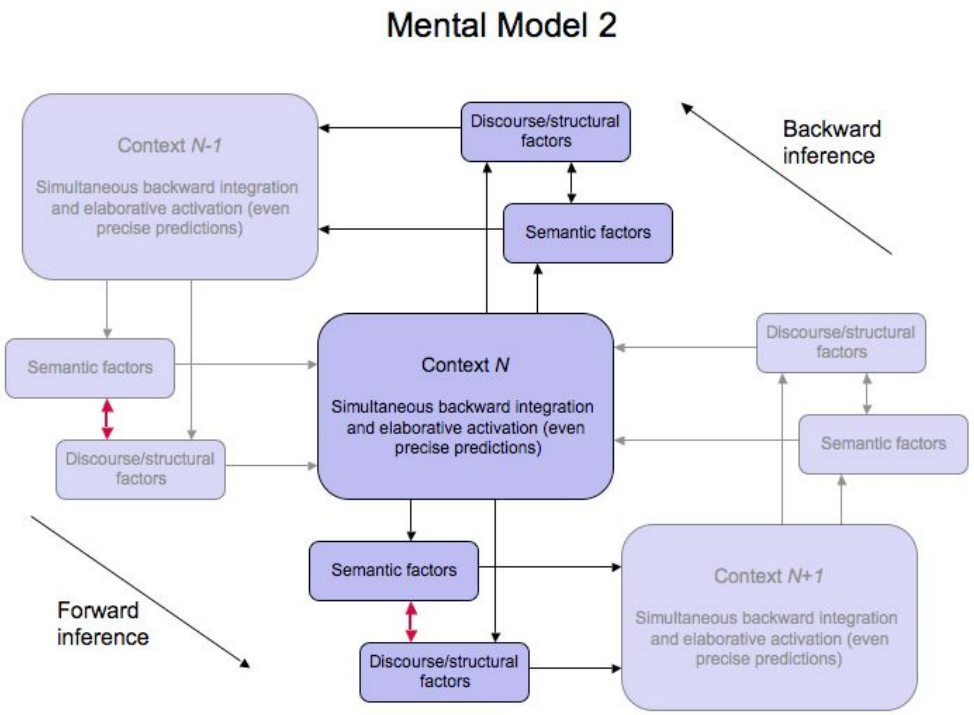

Figure 11. An illustration of the processing sequence postulated by Mental Model 2.

One feature that should also be discussed here is the relative importance of semantic cues compared to discourse/structural cues. However, as noted earlier, the strength of the semantic cues might be dependent on the strength of the discourse/structural cues (as was shown in studies II and III). The current studies did not find any obvious difference in the strength of different semantic cues; instead, all semantic cues which were studied played an important role, but their relative strength was not tested by pitting them against each other. The current studies did not show that semantic cues are subordinate to discourse/structural cues. Interestingly, Study IV showed that different prompts for backward inferences (anaphoric vs. non-anaphoric) led to differing strength of the semantic and structural cues. Therefore, future modelling will benefit from studies testing different types of prompts and the relative importance of discourse/structural cues as well as different aspects of semantics in making inferences following different prompts.

In addition, both of the above models are descriptions of inferential processes that occur isolated from other sensory input. As current research on situated language comprehension suggests, other cues, such as the visual environment simultaneously present with the language input, may play an important role in language comprehension (e.g., Crocker et al., in press; Altmann & Mirkovic, 2009). How this environment contributes to inference generation is considered in the following section.

Before that, it should be noted that the same models can be used to make hypotheses about children's inference generation. One interesting aspect of testing the same model with both adults and children is to study the developmental path of when different inferential processes are acquired during language development. Study III showed that (at least) one dimension of semantic information is available and used as early as at age 3. Further studies are needed to replicate these findings and investigate how other aspects of semantics affect inferencing in children. Importantly however, the current study implies that semantic effects are not subordinate to discourse/structural factors; instead, semantic effects are weighed in relation to structural factors. As noted earlier, when structural factors were a combination of first-mentioned entity, grammatical subject and linguistic topic, semantics was operating earlier in pronoun resolution (Study III). In contrast, in adult Study II, where there was also a topic shift, the structural effects masked the semantic effect during early phases of pronoun resolution. These findings suggest that in backward inferencing the strengths of individual semantic and discourse/structural cues are weighted in relation to the interaction between semantic and structural factors. Moreover, the study with children showed that such weighting may be functional even in young children, although further studies are needed to confirm this assumption.

4.3 The Situated Mental Model of inferences – future directions

As is typical for scientific research, the current thesis has led to many open questions to be tackled in future research. Some of those have already been specified above when putting forth the two alternative models of inference generation. One

question that can be asked when using the visual world method is what role the visual environment plays in inference generation. As argued in Chapter 2, in the current studies the visual information was carefully selected not to give any obvious cues on what kind of inferences to draw and what kinds of cues to use in the inferential processes. However, the active role of visual environment should not be ignored in theories of inference generation.

Earlier studies suggest that visual scenes play an important or even major role when predicting upcoming arguments in situated language comprehension (Altmann & Kamide, 1999, 2007, 2009; Kamide et al., 2003; Knoeferle & Crocker, 2006, 2007; Knoeferle et al., 2005). Therefore, the earlier view of language comprehension as an amodal process has been replaced by assumptions that language comprehension occurs as an active interaction between cues and constraints from the surrounding sensory environment and that the visual environment has an important role in incremental language comprehension (e.g., Barsalou, 1999; Zwaan & Madden, 2004).

For example, Altmann and Kamide (1999) showed participants visual scenes depicting pictures of a boy, a cake, a tractor and other objects. While participants were viewing the scene they heard sentences such as (36) and (37) in English.

(36) The boy will eat the cake.

(37) The boy will move the cake.

The eye movements to the visual scene showed that participants made saccades toward the cake after the verb *eat* (36) even before the cake was mentioned in the spoken stimuli. However, no such preference was present after the verb *move* in (37). The results indicated that verbs such as *eat* set constraints for the potential upcoming arguments, i.e., it must be something edible, and when the visual scene provides information that fulfils those constraints, people are able to predict the upcoming arguments. However, it is also worth noting that such predictions of the upcoming words can be made without the co-presence of visual stimuli if the linguistic context is sufficiently constraining (e.g., DeLong, Urbach, & Kutas, 2005; Kutas & Federmeier, 2000; Otten & Van Berkum, 2008; Van Berkum, Brown, Zwiterlood, Kooijman, & Hagoort, 2005).

Interestingly, Knoeferle and Crocker (2006) also found that a visual scene may temporarily even override general world knowledge in language comprehension. While participants were listening to German sentences such as (38) they saw visual scenes depicting three characters, a pilot in the middle, a detective on one side of the pilot and a magician on the other side. The scene also showed actions and rather unusually depicted a magician spying on a detective (as typically the detective is spying on people).

(38) Den Piloten bespitzelt gleich der Detecktiv.

'The pilot (OBJ) spies-on soon the Detective (SUBJ)'

When participants heard the word *bespitzelt* ('spy-on'), they quickly started to fixate the magician, as it was the one that was performing the action of the screen. However, in the experiment in which only characters, and not actions, were depicted on the scene, participants started to fixate the detective after hearing the verb *bespitzelt*. These results showed that when the visual scene does not provide cues on who-is-doing-what-to-whom, people rely on their general world knowledge; however, when visual scene does provide cues, those cues can be (temporarily) even stronger than cues arising from the general world knowledge. These studies suggest that visually depicted objects and actions affect on-line language comprehension.

Based on Mental Model 1 introduced earlier, I propose a model of situated inference generation; Situated Mental Model 1 is depicted in Figure 12. The important questions here are what kinds of inferences the visual environment can prompt people to draw and how different cues interact with each other during inference generation. Situated Mental Model 1 includes a separate subcomponent for processing visual information (as one type of sensory input). The model assumes that this component interacts with discourse/structural and semantic factors during backward inferencing. However, only one interactional link is assumed for forward activation: an interaction between visual information and semantic factors. This link is proposed here based on the above findings that visual information is used to constrain semantically derived predictions during language comprehension: For example, in the Altmann and Kamide (1999) study, the visual information was used to constraint the scope of predictions for the verb *eat*. However, similarly to Mental Model 1, this model lacks an interactional link between discourse/structural and semantic factors. In addition, an interactional link between discourse/structural factors and visual information is not assumed in this model.

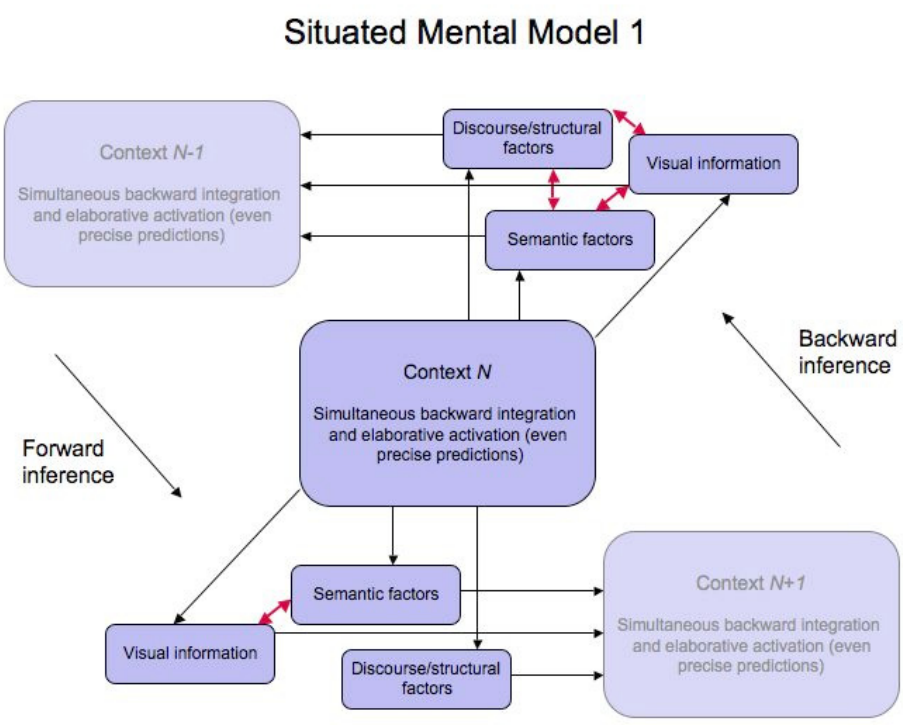

Figure 12. An illustration of the processing sequence postulated by Situated Mental Model 1 (a restricted interaction model).[9]

Similarly to Mental Model 2, a version of the Situated Mental Model can be formed in which more interactional links are assumed in forward inferencing. Situated Mental Model 2 is depicted in Figure 13 as an example of a fully interactive model. This model assumes that there are interactional links between visual information, discourse/structural factors and semantic factors in both backward and forward inferencing. Thus, it assumes that forward inference generation is not only a result of separate processes pertaining to visual, discourse/structural and semantic cues or to an interactional link between visual information and semantics, but rather interactions are also assumed to occur between discourse/structural factors and visual information. Thus, inferences are drawn based on the relative strength of these individual cues in combination.

[9] This illustration differs from the mental models presented above in Chapters 4.2.1 and 4.2.2, in that (in the interest of visual clarity) only the nodes from the currently operating sequence of the processor are drawn here.

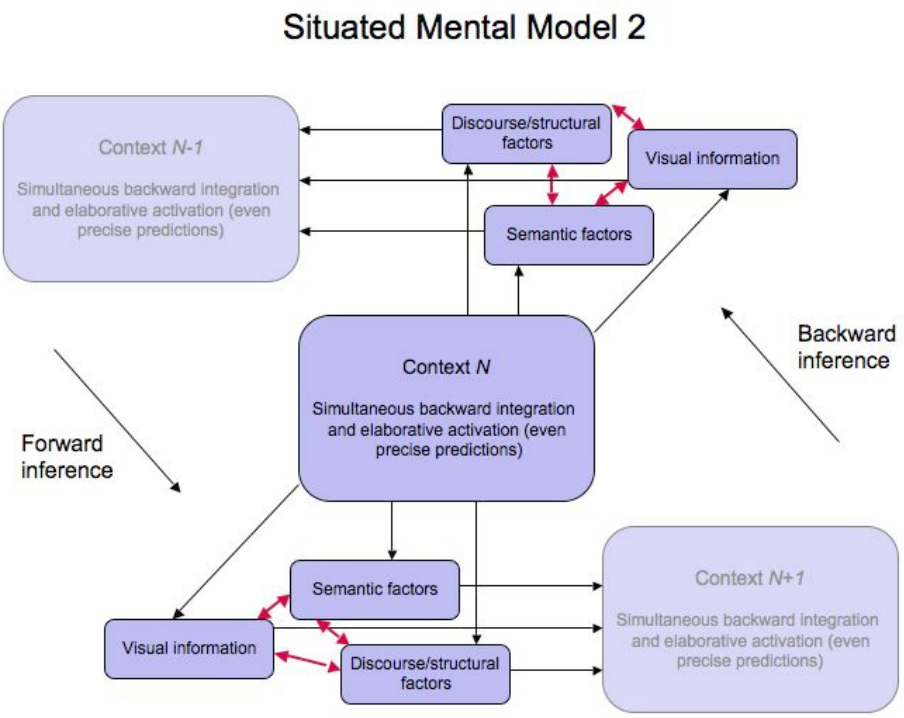

Figure 13. An illustration of the processing sequence postulated by Situated Mental Model 2 (a fully interactive model).

In order to select between the suggested models, a set of studies should be carried out in the future. Disentangling the role of visual environment in inference generation and interactions between different types of information (visual and linguistic) are questions that I expect will attract a lot of attention in future studies. Moreover, it is likely that the proposed models are able to only roughly capture inferential processes and that significant modifications are needed. For example, it may be that the proposed interactional links operate only between certain types of visual information and certain types of semantic information, or between certain types of visual information and certain types of discourse/structural factors.

5 CONCLUSIONS

The present thesis investigated the importance of semantics in generating inferences during discourse processing: whether semantics is activated elaboratively during discourse comprehension and what its relative importance is in backward inferencing compared to discourse/structural cues. The earlier literature has provided several hypotheses concerning my research questions; however, the question has remained largely unresolved partly due to limitations in the previously used methods. By using the visual world eye-tracking method I showed that semantics plays an important role in both backward and forward inferencing: Gender stereotypes and implicit causality information are activated elaboratively during online discourse comprehension. Moreover, gender stereotypes, implicit causality and proto-role properties of verbs are all used in backward inferencing. Importantly, the studies demonstrated that the importance of semantic cues is modulated by the presence of discourse/structural cues. When the structural cues consist of a combination of cues that have been independently shown to be important in backward inferencing, semantic effects may be masked, whereas when the structural cues consist of a combination of fewer prominent cues, semantics can have an earlier effect than structural factors in pronoun resolution. In addition, the type of inference matters, too: During anaphoric inferences semantics plays a prominent role, while discourse/structural salience attains more prominence during non-anaphoric inferences. Finally, semantics exhibits a strong role in triggering new inferences to revise inferences made earlier even in the case that an additional inference is not needed to establish coherence in discourse.

Although the findings are mainly in line with the Mental Model theories, not all of the above aspects are implemented in these models. Therefore, I proposed two extended models, Mental Model 1 and 2, which incorporate the current findings as well as findings in the earlier literature. These models allow both forward and backward inferencing to occur at any given moment during the course of processing and also allow semantic and discourse/structural cues to contribute to both of these processes. The difference between these two models is that while the former does not assume interaction between semantic and discourse/structural factors in forward inferencing, the latter does so. Further research is needed in order to select between these models. In addition, the current thesis also raised questions of the role of the visual environment during inference generation; thus, I sketched two alternative situated mental models as a basis for future research. They differ in the amount of interaction that is assumed to take place between semantic, discourse/structural and visual information.

Finally, in the present thesis the visual world eye-tracking method proved to be an ideal tool for studying both forward and backward inferences, and importantly, to disentangle forward activation from backward integration. In addition, the thesis showed the method can be successfully used in both adult and child language studies of inferencing. This method is already gained a lot of popularity in language comprehension research, and I assume it will become a very attractive method to be applied in the future studies of inference generation.

REFERENCES

Allbritton, D. (2004). Strategic production of predictive inferences during comprehension. *Discourse Processes, 38,* 309-322.

Allopenna, P.D., Magnuson, J.S., & Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419-439.

Almor, A. (1999). Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review, 106,* 748-765.

Altmann, G.T.M. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition, 93,* B79-B87.

Altmann, G.T.M., & Kamide, Y. (1999). Incremental interpretation of verbs: Restricting the domain of subsequent reference. *Cognition, 73,* 247-264.

Altmann, G.T.M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57,* 502-518.

Altmann, G.T.M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition, 111,* 55-71.

Altmann, G.T.M., & Mirkovic, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33,* 583-609.

Arnold, J. E., Brown-Schmidt, S., & Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes, 22,* 527-565.

Arnold, J. E., Brown-Schmidt, S., Trueswell, J., & Fagano, M. (2005). Children's use of gender and order of mention during pronoun comprehension. In J. Trueswell & M.K. Tanenhaus (Eds.), *Approaches to studying world-situated language use* (pp. 261-281). Cambridge, MA: MIT Press.

Arnold, J. E. Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The immediate use of gender information: Eyetracking evidence of the time-course of pronoun resolution. *Cognition, 76,* B13-B26.

Au, T.K. (1986). A verb is worth of a thousand words: The causes and consequencies of interpersonal events implicit in language. *Journal of Memory and Language, 25,* 104-122.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science, 7,* 136-141.

Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577-660.

Bestgen, Y., & Vonk, W. (2000). Temporal adverbials as segmentation markers in discourse comprehension. *Journal of Memory and Language, 42,* 74-87.

Bloom, C.P., Fletcher, C.R., Van den Broek, P., Reitz, I., & Shapiro, B.P. (1990), An on-line assessment of causal reasoning during comprehension. *Memory & Cognition, 18,* 65-71.

Boersma, P. (2001). PRAAT, a system for doing phonetics by computer. *Glot International, 5,* 341-345.

Boland, J.E. (2005). Visual arguments. *Cognition, 95,* 237-274.

Brown, R. & Fish, D. (1983). The psychological causality implicit in language. *Cognition, 14,* 237-273.

Cacciari, C., & Padovani, R. (2007). Further evidence of gender stereotype priming in language: Semantic facilitation and inhibition in Italian role nouns. *Applied Psycholinguistics, 28,* 277-293.

Calvo, M. G., & Castillo, M.D. (1996). Predictive inferences occur on-line, but with delay: Convergence of naming and reading times. *Discourse Processes, 22,* 57-78.

Calvo, M. G., & Castillo, M.D. (1998). Predictive inferences take time to develop. *Psychological Research, 61,* 249-260.

Calvo, M.G., Meseguer, E., & Carreiras, M. (2001). Inferences about predictable events: Eye movements during reading. *Psychological Research, 65,* 158-169.

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *Quarterly Journal of Experimental Psychology, 49,* 639-663.

Carreiras, M., Gernsbacher, M.A., & Villa, V. (1995). The advantage of first mentioned in Spanish. *Psychonomic Bulletin & Review, 2,* 124-129.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C. Li (Ed.), *Subject and topic* (pp. 25-56). New York: Academic Press.

Chambers, C.G., & Smyth, R. (1998). Structural parallelism and discourse coherence: A test of centering theory. *Journal of Memory and Language, 39,* 593-608.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Clark, H.H. (1977). Bridging. In P.N. Johnson-Laird & P.C. Wason (Eds.), *Thinking: Readings in cognitive science* (pp. 411-420). Cambridge: Cambridge University Press.

Clark, H.H. & Card, S.K. (1969). Role of semantics in remembering comparative sentences. *Journal of Experimental Psychology, 82,* 545-553.

Clark, H.H., & Haviland, S. (1977). Comprehension and the given-new contract. In R.O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1-40). Hillsdale, NJ: Erlbaum.

Cooper, R.M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6,* 84-107.

Corbett, A.T., & Chang, F.R. (1983). Pronoun disambiguation: Accessing potential antecedents. *Memory & Cognition, 11,* 283-294.

Crawley, R., Stevenson, R., & Kleinman, D. (1990). The use of heuristic strategies in the intepretation of pronouns. *Journal of Psycholinguistic Research, 19,* 245-264.

Crinean, M., & Garnham, A. (2006). Implicit causality, implicit consequentiality and semantic roles. *Language and Cognitive Processes, 21,* 636-648.

Crocker, M.W., Knoeferle, P., & Mayberry, M. (in press). Situated sentence comprehension: The coordinated interplay account and a neurobehavioral model. *Brain & Language*.

Dahan, D., Magnuson, J.S., & Tanenhaus, M.K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42,* 317-367.

Dahan, D., & Tanenhaus, M.K. (2005). Looking at the rope when looking for a snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review, 12*, 453-459.

DeLong, K.A., Urbach, T.P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8,* 1117-1121.

Dowty, D. (1991). Thematic roles and argument selection. *Language, 67,* 547-619.

Duffy, S.A. (1986). Role of expectations in sentence integration. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 12,* 208-219.

Duffy, S.A., & Keir, J.A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition, 32,* 551-559.

Eberhard, K.M., Spivey, M.J., Sedivy, J.C., & Tanenhaus, M.K. (1995). Eye-movements as a window into spoken language comprehension in natural contexts. *Journal of Psycholinguistics Research, 24,* 409-436.

Ehrlich, K. (1980). Comprehension of pronouns. *The Quarterly Journal of Experimental Psychology, 32,* 247-255.

Fillmore, C. (1968). The case for case. In E. Bach, & R. Harms (Eds.), *Universals in Linguistic Theory* (pp. 1-90). New York: Holt, Rhinehart & Winston.

Fillmore, C. (1977). Scenes-and-frames semantics. In A. Zampolli (Ed.), *Fundamental studies in computer science* (pp. 55-81). Amsterdam: North Holland Publishing.

Fincher-Kiefer, R. (1993). The role of predictive inferences in situation model construction. *Discourse Processes, 16,* 99-124.

Fincher-Kiefer, R. (1995). Relative inhibition following the encoding of bridging and predictive inferences. *Journal of Experimental Psychology: Leaning, Memory, & Cognition, 21,* 981-995.

Fincher-Kiefer, R. (1996). Encoding difference between bridging and predictive inferences. *Discourse Processes, 22,* 225-246.

Findlay, J.M. (2004). Eye scanning and visual search. In J.M. Henderson and F. Ferreira (Eds.), *The interfaces of language, vision and action: Eye movements and the visual world* (pp. 135-160). New York: Psychology Press.

Frederiksen, J. (1981). Understanding anaphora: Rules used by readers in assigning pronominal referents. *Discourse Processes, 4,* 323-347.

Fischer, M.H., & Zwaan, R.A. (2008). Embodied language – A review of the role of the motor system in language comprehension. *Quarterly Journal of Cognitive Psychology, 61,* 825-850.

Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Availability versus accessibility. *Journal of Memory and Language, 56,* 357-383.

Garnham, A. (1993). Is logicist cognitive science possible? *Mind and Language, 8,* 49-71.

Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Hove: Psychology Press.

Garnham, A., Oakhill, J.V., & Cruttenden, H. (1992). The role of implicit causality and gender cue in the interpretation of pronouns. *Language and Cognitive Processes, 7,* 231-255.

Garnham, A., Oakhill, J.V., & Reynolds, D.J. (2002). Are inferences from stereotyped role names to characters' gender made elaboratively? *Memory & Cognition, 30,* 439-446.

Garnham, A., Traxler, M., Oakhill, J., & Gernsbacher, M.A. (1996). The locus of implicit causality effects in comprehension. *Journal of Memory and Language, 35,* 517-543.

Garrod, S.C., Freudenthal, D., & Boyle, E. (1994). The role of different types of anaphor in the on-line resolution of sentences in a discourse. *Journal of Memory & Language, 33,* 39-68.

Garrod, S.C., O'Brien, E.J., Morris, R.K., & Rayner, K. (1990). Elaborative inferences as an active or passive process. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 250-257.

Garrod, S.C., & Sanford, A.J. (1982). The mental representation of discourse in a focussed memory system: Implications for the interpretation of anaphoric noun phrases. *Journal of Semantics, 1,* 21-41.

Garrod, S.C., & Sanford, A.J. (1994). Resolving sentences in a discourse context: How discourse representation affects language comprehension. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 675-698). San Diego, CA: Academic Press.

Garrod, S.C., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory & Language, 42,* 526-544.

Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry, 5,* 517-543.

Garvey, C., Caramazza, A., & Yates, J. (1975). Factors influencing assignments of pronoun antecedents. *Cognition, 3,* 227-243.

Gernsbacher, M.A. (1989). Mechanisms that improve referential access. *Cognition, 32,* 99-156.

Gernsbacher, M.A. (1990). *Language comprehension as structure building.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Gernsbacher, M.A., & Hargreaves, D.J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language, 27,* 699-717.

Gernsbacher, M.A., Hargreaves, D.J., & Beeman, M. (1989). Building and accessing clausal representations: The advantage of first mention vs. the advantage of clause recency. *Journal of Memory and Language, 28,* 735-755.

Givón, T. (1983). *Topic continuity in discourse: Quantified Cross-Language Studies,* TSL 3. Amsterdam: John Benjamins Publishing Company.

Gordon, P.C., Grosz, B.J., & Gilliom, L.A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science, 17,* 311-347.

Gordon, P.C., & Hendrick, R. (1997a). Intuitive knowledge of linguistic co-reference. *Cognition, 62,* 325-370.

Gordon, P.C., & Hendrick, R. (1997b). The representation and processing of coreference in discourse. *Cognitive Science, 22,* 389-424.

Gordon, P.C., & Hendrick, R. (1998). The representation and processing of coreference in discourse. *Cognitive Science, 22,* 389-424.

Gordon, P.C., Hendrick, R., & Foster, K.L. (2000). Language comprehension and probe-list memory. *Cognitive Science, 17,* 311-347.

Gordon, P.C., & Scearce, K.A. (1995). Pronominalization and discourse coherence, discourse structure and pronoun interpretation. *Memory & Cognition, 23,* 313-323.

Graesser, A.C., Haberlandt, K., Koizumi, D. (1987). How is reading time influenced by knowledge-based inferences and world knowledge? In B.K. Britton & S.M. Glynn (Eds.), *Executive control processes in reading* (pp. 217-251). Hillsdale, NJ: Erlbaum.

Greene, S.B., & McKoon G. (1995). Telling something we can't know: Experimental approaches to verbs exhibiting implicit causality. *Psychological Science, 6,* 262-270.

Grober, E., Beardsley, W., & Caramazza, A. (1978). Parallel function strategy in pronoun assignment. *Cognition, 6,* 117-133.

Grosz, B.J., Joshi, A., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics, 21,* 203-225.

Grosz, B.J., & Sidner, C.L. (1986). Attention, intention, and the structure of discourse. *Computational Linguistics, 12,* 175-204.

Guerry, M., Gimenes, M., Caplan, D., & Rigalleau, F. (2006). How long does it take to find a cause? An online investigation of implicit causality in sentence production. *The Quarterly Journal Of Experimental Psychology, 59,* 1535-1555.

Gundel, J.K. (1999). On different kinds of focus. In P. Bosch and R. van der Sandt (Eds.). *Focus: Linguistic, cognitive and computational perspectives* (pp. 293-305). Cambridge: Cambridge University Press.

Gundel, J.K., & Hedberg, N. (2008). Introduction. In J.K. Gundel and N. Hedberg (Eds.), *Reference. Interdisciplinary perspectives* (pp. 3-10). New York: Oxford University Press.

Gundel, J.K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language, 69,* 274-307.

Gussenhoven, C. (2007). Types of focus in English. In C. Lee, M. Gordon and D. Büring (eds). *Topic and focus: Cross-linguistic perspectives on meaning and intonation* (pp. 83-100). Heidelberg, New York, London: Springer.

Hagoort, P., & Brown, C.M. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of Psycholinguistic Research, 28,* 715-728.

Hakulinen, A., Karlsson, F., & Vilkuna, M. (1981). *Suomen tekstilauseiden piirteitä: Kvantitatiivinen tutkimus*. Department of General Linguistics Publication. Helsinki: University of Helsinki.

Haviland, S.E., & Clark, H.H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13,* 512-521.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences, 9,* 188-194.

Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language, 56,* 251-299.

Ibbotson, P., & Tomasello, M. (2009). Prototype constructions in early language acquisition. *Language and Cognition, 1,* 59-85.

Irmen, L. (2007). What's in a (role) name? Formal and conceptual aspects of comprehending personal noun. *Journal of Psycholinguistic Research, 36,* 431-456.

Irwin, D.E. (2004). Fixation location and fixation duration as indices of cognitive processing. In J.M. Henderson and F. Ferreira (Eds.), *The interfaces of language, vision and action: Eye movements and the visual world* (pp. 105-134). New York: Psychology Press.

Jackendoff, R. (1972). *Semantic interpretation in generative grammar.* Cambridge: MIT Press.

Jackendoff, R. (1976). Toward an explanatory semantic representation. *Linguistic Inquiry, 7,* 89-150.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge: MIT Press.

Jackendoff, R. (1990). *Semantic structures*. Cambridge: MIT Press.

Johnson-Laird, P.N. (1983). *Mental models: Toward a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.

Johnson-Laird, P.N. (1994). Mental models and probabilistic thinking. *Cognition, 16,* 1-61.

Järvikivi, J., Van Gompel, R.P.G., & Hyönä, J. (submitted). The interplay of implicit causality, structural heuristics, and type of pronoun in anaphor resolution.

Järvikivi, J., Van Gompel, R.P.G., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. *Psychological Science, 16,* 260-264.

Kaiser, E., & Trueswell, J.C. (2008). Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes, 23,* 709-748.

Kaiser, E., & Trueswell, J.C. (in press). Investigating the interpretation of pronouns and demonstratives in Finnish: Going beyond salience. In E.Gibson & N. Pearlmutter (Eds.) *The processing and acquisition of reference.* Cambridge: MIT Press.

Kako, E. (2006). Thematic role properties of subjects and objects. *Cognition, 101,* 1-42.

Kamide, Y., Altmann, G.T.M., & Haywood, S. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49,* 133-156.

Keefe D.E., & McDaniel, M.A. (1993). The time course and durablility of predictive inferences. *Journal of Memory and Language, 32,* 446-463.

Kennison, S.M., & Trofe, J.L. (2003). Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research, 32,* 355–378.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95,* 163-182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* New York: Cambridge University Press.

Kintsch, W. & Van Dijk, T.A. (1978). Toward a model of discourse comprehension and production. *Psychological review, 85,* 363-394.

Klin, C.M., Murray, J.D., Levine, W.H., & Guzmán, A.E. (1999). Forward inferences: From activation to long-term memory. *Discourse Processes, 27,* 241-260.

Knoeferle, P., & Crocker, M.W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science, 30,* 481-529.

Knoeferle, P., & Crocker, M.W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye-movements. *Journal of Memory and Language, 57,* 519-542.

Knoeferle, P., Crocker, M.W., Scheepers, C., & Pickering, M.J. (2005). The influcence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition, 95,* 95-127.

Koornneef, A.W., & Van Berkum, J.J.A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language, 54,* 445-465.

Kreiner, H., Sturt, P., & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language, 58,* 239-261.

Kutas, M., & Federmeier, K.D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4,* 463-470.

Kutas, & Federmeier, (2007). Event-related brain potential (ERP) studies of sentence processing. In M.G. Gaskell (Ed.), *Oxford handbook of psycholinguistics* (pp. 385-406). Oxford: Oxford University Press.

Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus and the mental representation of discourse referent.* Cambridge: Cambridge University Press.

Larson, R. (1988). On the double object construction. *Linguistic Inquiry, 19,* 335-392.

Long, D.L., & De Ley, L. (2000). Implicit causality and discourse focus: The integreation of text and reader characteristics in pronoun resolution. *Journal of Memory and Language, 42,* 545-570.

MacWhinney, B. (1977). Starting points. *Language, 53,* 152-168.

Magliano, J.P., Bagget, W.B., Johnson, B.K., & Graesser, A.C. (1993). The time course of generating causal antecedent and causal consequence inferences. *Discourse Processes, 9,* 57-90.

Matin, E., Shao, K.C., & Boff, K.R. (1993). Information-processing time with and without saccades. *Perception & Psychophysics, 53,* 372-380.

Mayberry, M., Crocker, M.W., & Knoeferle, P. (2009). Learning to attend: A connectionist model of situated language comprehension. *Cognitive Science, 33,* 449-496.

McDaniel, M.A., Schmalhofer, F., & Keefe, D.E. (2001). What is minimal about predictive inferences? *Psychonomic Bulletin & Review, 8,* 840-846.

McDonald, J. L., & MacWhinney, B. (1995). The time-course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language, 34,* 543-566.

McDonald, S.A., & Shillcock, R.C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science, 14,* 648-652.

McKoon, G., Gerrig, R.J., & Greene, S.B. (1996). Pronoun resolution without pronouns: Some consequences of memory-based text processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22,* 919-932.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 12,* 82-91.

McKoon, G., & Ratcliff, R. (1988). Contextually relevant aspects of meaning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14,* 331-343.

McKoon, G., & Ratcliff, R. (1989). Semantic association and elaborative inference. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15,* 326-338.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99,* 440–466.

McKoon, G., Ratcliff, R., Ward, G., & Sproat, R. (1993). Syntactic prominence effects on discourse processes. *Journal of Memory and Language, 32,* 593-607.

McRae, K., Spivey-Knowlton, M.J., & Tanenhaus, M.K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language, 38,* 283-312.

Miller, G.A., & Johnson-Laird, P.N. (1976). *Language and perception.* Cambridge: Cambridge University Press.

Millis, K., Morgan, D., & Graesser, A.C. (1990). The influence of knowledge-based inferences on the reading time of expository test. In A.C. Graesser and G.H. Bower (Eds.), *The psychology of learning and motivation* (pp. 197-212). New York: Academic Press.

Murray, J.D., Klin, C.M., & Myers, J.L. (1993). Forward inferences in narrative text. *Journal of Memory and Language, 32,* 464-473.

Myers, J.L., Shinjo, M., & Duffy, S.A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language, 26,* 453-465.

Oakhill, J. V., Garnham, A., & Reynolds, D. J. (2005). Immediate activation of stereotypical gender information. *Memory & Cognition, 33,* 972-983.

O'Brien, E.G., Albrecht, J.E., Hakala, C.M., Rizzella, M.L. (1995). Activation and suppression of antecedents during reinstatement. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21,* 626-634.

O'Brien, E.J., Plewes, P.S., & Albrecht, J.E. (1990). Antecedent retrieval processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16,* 241-249.

Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violation of gender stereotypes. *Memory & Cognition, 25,* 273–285.

Otten, M., & Van Berkum, J.J.A. (2008). Discourse-based word anticipation during language processing: Prediction or priming. *Discourse Processes, 45,* 464-496.

Pickering, M. J., & Majid, A. (2007). What are implicit causality and consequentiality? *Language and Cognitive Processes, 22,* 780-788.

Potts, G.R., Keenan, J.M., & Golding, J.M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language, 27,* 399-415.

Pyykkönen, P., Hyönä, J., & Van Gompel, R.P.G. (2009). Activating gender stereotypes during online spoken language processing: Evidence from visual world eye-tracking. *Experimental Psychology*. Article in press.

Pyykkönen, P., & Järvikivi, J. (2009). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*. Article in press.

Pyykkönen, P., Matthews, D., & Järvikivi, J. (2009). Three-year-olds are sensitive to semantic prominence during online language comprehension: A visual world study of pronoun resolution. *Language and Cognitive Processes*. Article in press.

Pyykkönen, P., Van Gompel, R.P.G., & Hyönä (2009). The importance of semantics and salience during anaphoric and non-anaphoric inferencing. *Manuscript.*

Reynolds, D.J., Garnham, A., & Oakhill, J. (2006). Evidence of immediate activation of gender information from a social role name. *The Quarterly Journal of Experimental Psychology, 59,* 886-903.

Rose, R.L. (2005). *The relative contribution of syntactic and semantic prominence to the salience of discourse entities*. Unpublished PhD Thesis. Northwestern University, Illinois.

Rudolph, U., & Fösterling, F. (1997). The psychological causality implicit in verbs: A review. *Psychological Bulletin, 121,* 192-218.

Sanford, A.J. (1985). *Cognition and cognitive psychology*. London: Weidenfeld & Nicolson.

Schlesinger, I. (1995). *Cognitive space and linguistic case.* Cambridge: Cambridge University Press.

Sedivy, J.C., Tanenhaus, M.K., Chambers, C.G., & Carlson, G.N. (1999). Achieving incremantal semantic interpretation through contextual representation. *Cognition, 71,* 109-147.

Serratrice, L. & Kidd, E. (in preparation). Implicit causality and order of mention in pronoun resolution: Evidence from eye-tracking in adults and children.

Sheldon, A.L. (1974). The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior, 13,* 272-281.

Singer, M., & Ferreira, F. (1983). Inferring consequences in story comprehension. *Journal of Verbal Leaning & Verbal Behavior, 22,* 437-448.

Smyth, R. (1994). Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research, 23,* 197-229.

Song, H., & Fisher, C. (2005). Who's "she"? Discourse structure influences preschoolers' pronoun interpretation. *Journal of Memory and Language, 52,* 29-57.

Song, H., & Fisher, C. (2007). Discourse prominence effects on 2.5-year-old children's interpretation of pronouns. *Lingua, 117,* 1959-1987.

Speas, M. (1990). *Phrase structure in natural language.* Dordrecht: Kluwer Academic Publishers.

Spivey, M.J., Tanenhaus, M.K., Eberhard, K.M., & Sedivy, J.C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiquity resolution. *Cognitive Psychology, 45,* 445-481.

Stevenson, R.J., Crawley, R.A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes, 9,* 519-548.

Stevenson, R.J., Knott, A., Oberlander, J., & McDonald, S. (2000). Interpreting pronouns and connectives: Interactions among focusing, thematic roles and coherence relations. *Language and Cognitive Processes, 15,* 225-262.

Stewart, A.J., & Gosselin, F. (2000). A simple categorisation model of anaphor resolution. In L.R. Gleitman & A.K. Joshi (Eds.), *Proceedings of the twenty-second annual conference of the cognitive science society* (pp. 930-935). Hillsdale, NJ: Erlbaum.

Stewart, A.J., Pickering, M.J., & Sanford, A.J. (2000). The time course of the influence of implicit causality information: Focusing versus integration accounts. *Journal of Memory and Language, 42,* 423-443.

Sturt, P. (2003a). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language, 48,* 542-562.

Sturt, P. (2003b). A new look at the syntax-discourse interface: The use of binging principles in sentence processing. *Journal of Psycholinguistic Research, 32,* 125-139.

Tanenhaus, M.K. (2007). Spoken language comprehension: Insights from eye movements. In M.G. Gaskell (Ed.), *Oxford handbook of psycholinguistics* (pp. 309-326). Oxford: Oxford University Press.

Tanenhaus, M.K., Eberhard, K.M., & Sedivy, J.C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Science, 45,* 447-481.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632-1634.

Trueswell, J.C., Sekerina, I., Hill, N., & Logrip, M. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73,* 89-134.

Trueswell, J.C., Tanenhaus, M.K., & Garnsey, S. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33,* 285-318.

Vainio, M., & Järvikivi, J. (2006). Tonal features, intensity, and word order in the perception of prominence. *Journal of Phonetics, 34,* 319-342.

Van Berkum, J.J.A., Brown, C.M., & Hagoort, P. (1999a). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41,* 147-182.

Van Berkum, J.J.A., Brown, C.M., & Hagoort, P. (1999b). When does gender constrain parsing? Evidence from ERPs. *Journal of Psycholinguistic Research, 28,* 555-566.

Van Berkum J.J.A., Koornneef, A.W., Otten, M., & Nieuwland, M.S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research, 1146,* 158-171.

Van Berkum, J.J.A., Brown, C.M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 3,* 443-467.

Vilkuna, M. (1989). *Free word order in Finnish: Its syntax and discourse functions.* Helsinki: Suomalaisen kirjallisuuden seura.

Von Eckardt, B., & Potter, M.C. (1985). Clauses and the semantic representation of words. *Memory & Cognition, 13,* 371-376.

Vonk, W., Hustinx, L.G.M.M., & Simons, W.H.G. (1992). The use of referential expressions in structuring discourse. *Language and Cognitive Processes, 7,* 301-333.

Wasow, T. (1986). Reflections of anaphora. In B. Lust (Ed.), *Studies in the acquisition of anaphora* (pp.107-122). Dordrecht: D. Reidel.

Whitney, P., Richie, B.G., & Crane, R.S. (1992). The effect of foregrounding on readers' use of predictive inferences. *Memory & Cognition, 20,* 424-432.

Zwaan, R.A., & Madden, C.J. (2004). Updating situation models. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30,* 282-288.

Zwaan, R.A., & Taylor, L.J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General, 135,* 1-11.

Zwaan, R.A., & Radvansky, G.A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123,* 162-185.